The Optimal Extent of Discovery *

Frances Z. Xu Lee† Dan Bernhardt‡

October 5, 2015

Abstract

We characterize how the process of publicly-gathering information via discovery affects strategic interactions between litigants. It allows privately-informed defendants to signal through the timing of settlement offers, with weaker ones attempting to settle pre-discovery. Discovery reduces the probability of trial. Properly designed limited discovery reduces expected litigation costs. Stronger defendants gain more (lose less) from a given amount of discovery. We find that the court should grant more discovery when defendants are believed to be stronger and should grant discovery on more efficient sources of information, leaving less efficient ones to trial.

*We thank Odilon Camara, Kathy Spier, Wing Suen, Satoru Takahashi, the audience at 2013 Annual Meeting of ALEA at Vanderbilt University, Midwest Theory Conference Spring 2013 at Michigan State University, and the seminar at National University of Singapore in 2012, and two anonymous referees.

†francesxu312@gmail.com

‡University of Illinois at Urbana Champaign, danber@illinois.edu

1 Introduction

Legal discovery has been part of the civil litigation process in the United States since the adoption of the 1938 Federal Rules of Civil Procedure. The motivation was that "full access to the evidence would end trial by ambush and surprise. Open discovery would promote settlements; with both sides obliged to turn over all their important cards, secrets would disappear and realistic negotiations would occur" (Rosenberg 1989). The rules of procedure provide judges substantial discretion in exercising judicial control over discovery. For example, Rule 26(b)1 says that "For good cause, the court may order discovery of any matter relevant to the subject matter involved...", while Rule 26(b)2 advises judges to evaluate the benefits and costs "the court must limit the extent of discovery ... if it determines that ...the expense of the proposed discovery outweighs its likely benefit..."

The law and economics literature typically compresses the civil litigation *process*—the information gathering, litigation costs and legal outcome—into a single trial stage, possibly combined with a pre-trial settlement offer stage. In practice, litigants have many opportunities to try to settle throughout the litigation process, and a large portion of litigation expenses—gathering and reviewing case materials, preparing motions, hiring experts, pre-trial depositions, etc.—is incurred in the process of discovery, rather than in the trial itself.[1]

Despite discovery's prominent role, there has been little analysis of how its design affects decision-making by litigants or the litigation costs that they incur. In this article, we endogenize the timing of settlement offers relative to the discovery, analyzing the costs and benefits of greater discovery when the asymmetrically-informed litigants are *strategic* in their settlement decisions, and settlement amounts and timing can signal information. Given the ability of a court to control the extent of discovery, we characterize the optimal level of discovery, and show how it depends

[1] Glaser (1968) reported that in antitrust cases discovery represented 65% of plaintiffs' costs and 63% of defendants' costs, and in patent cases discovery accounted for 21% of plaintiffs' costs and 54% of defendants' costs. Discovery constitutes about 43% of the litigation time for heavy discovery cases, and 25-31% for the average case.

on both the likely underlying strength of the case and the cost structure of discovery. We demonstrate that shifting a *limited* amount of information gathering from the trial to discovery can benefit litigants. This is not simply because earlier revelation of information can guide future decisions, which is the rationale underlying bifurcated trials.[2] Rather, it is because discovery allows the timing of settlement offers to signal a defendant's private information. Weak defendants strategically make settlement offers prior to discovery, whereas strong ones wait until after discovery to do so. We show that the court should grant more discovery when a defendant is expected to be stronger, and that the court should focus discovery on sources of information that are less costly to obtain, leaving more costly ones for a trial.

At the outset of many litigations, a plaintiff and defendant are often very asymmetrically informed about their prospects in a trial. We focus on settings in which the defendant alone has private information. Our model of discovery builds on Reinganum and Wilde (1986). In their model, the informed party, the plaintiff,[3] makes a pre-trial, take-it-or-leave-it settlement offer that reflects its private information. If the settlement offer is rejected, they proceed to a costly trial that determines whether the defendant is liable. They show that equilibrium offers are separating, and, to induce the informed party to truthfully reveal its private information via its settlement offer amount, the uninformed party must be more likely to reject less generous offers.

In Reinganum and Wilde (1986), exogenous information revelation only occurs at trial. With public discovery, both the extent of information revelation and its costs are spread out over time. In our model, discovery is characterized by (a) the probability that a defendant's private information is uncovered publicly, and (b) the costs that the defendant and plaintiff incur in that discovery. More extensive discovery that is more likely to uncover a defendant's information incurs a greater share of the total litigation costs. Before discovery, a defendant can make a settlement offer; if the

[2]Civil cases can be bifurcated into separate liability and damages proceedings. It can avoid costly litigation over the amount of damage if the defendant is not found liable. See Landes (1993) and White (2006) for economics analysis of bifurcated trials and their unintended consequences.

[3]We consider an informed defendant, rather than an informed plaintiff, but this difference is cosmetic.

plaintiff rejects it, the parties go on to discovery; and post-discovery, offers can again be made that reflect both the information revealed by earlier settlement offers and by any information uncovered in discovery. After discovery, absent a settlement, a trial determines whether a defendant is liable.

We first take the design of discovery—the probabilities that discovery uncovers a defendant's information, and the cost of discovery—as given, and solve for how it affects the timing of when different defendant types make their first offers, and the sizes of these offers. When the extent of discovery is not too high, defendant types partition themselves. The weakest defendants—those facing plaintiffs who are likely to win—make their first offers prior to discovery. The sizes of settlement offers by defendant types who make offers at the same time fully separate their types, eliminating all information asymmetries. As a result, defendant types that make pre-discovery offers either settle immediately, or they settle just after discovery on terms that reflect the information revealed by their offers. Stronger defendant types wait until after discovery to make offers.

Thus, the process of publicly gathering information in discovery also allows defendants to signal via the timing of their initial settlement offers. Moreover, discovery sometimes uncovers a defendant's private information, obviating the need to engage in costly signaling via settlement offers that incur an equilibrium risk of rejection and thus further litigation costs. Of course, the process of discovery is itself costly, so that it is not clear how more extensive discovery affects expected litigation costs. Indeed, perfectly exhaustive discovery that uncovers all information is akin to shifting the timing of a trial forward, delivering exactly the same expected litigation costs as a legal system with no discovery process, where litigants go directly to trial if they fail to settle.

This leads us to derive how discovery affects the payoffs of the plaintiff and different defendant types. Greater discovery reduces the likelihood that a pre-discovery settlement offer is rejected, but it also raises the discovery costs incurred when offers are rejected. When information costs are proportional to the extent of discovery, these two effects exactly cancel out for weak defendants that make their first offers prior to discovery: expected total litigation costs incurred with weak defendants do not vary with the extent of discovery. We then highlight the role of a judge in directing discovery: if discovery can selectively targets lower marginal cost sources of information, limited discovery reduces expected litigation costs for weak defendants; but if discovery targets higher marginal cost sources of information, expected litigation costs are higher for weak defendants.

In contrast, as long as discovery is not too inefficiently targeted toward higher marginal cost sources of information, it always facilitates signaling by strong defendant types who wait until after discovery to make their first offers, reducing their expected litigation costs. To signal via settlement offers, better types must face higher rejection probabilities in order to deter mimicking by worse types. This rejection probability must rise *concavely* for better types, reflecting how less generous offers lose more by being rejected. For strong defendants who wait until after discovery to settle, relative to no discovery, successful discovery eliminates the need to signal, but unsuccessful discovery makes subsequent signaling harder because remaining trial costs are less. However, the inherent concavity in the rejection probability in signaling by settlement offers means that discovery reduces the expected rejection probability and thus expected litigation costs.

When information costs are proportional to the extent of discovery, sufficiently limited discovery *always* reduces expected total litigation costs, and the informed defendant extracts all of the gains. We characterize the extent of discovery that minimizes expected litigation costs, which, in turn, maximizes a defendant's ex-ante payoffs. With greater discovery, more defendant types make offers pre-discovery to avoid incurring discovery costs. We establish that among those defendants who wait until after discovery to make offers, the stronger is a defendant, the more it gains from a given increase in discovery. Thus, greater discovery is optimal when the distribution of defendant types is better/stronger in the conditional first order stochastic dominance sense.

The value of limited discovery is reinforced when the limited discovery is divided into more rounds: it facilitates signaling, reducing the expected litigation costs incurred. Intuitively, "shrinking the distance" from the worst type to be separated from in a given round reduces the distortion required in settlement offers. Interestingly, further dividing discovery, allowing more rounds of settlement, induces more defendant types to "play tough": with more rounds, more defendant types wait until after *all* discovery ends, making their first offers just before a trial if discovery fails to reveal their private information.

We assume that successful discovery does not affect subsequent trial costs. If successful discovery, in fact, reduces trial costs, it facilitates separation of weak defendants by raising the costs of mimicking a stronger type's offer. Defendants may now be hurt by discovery, despite its social benefit, because successful discovery reduces the trial costs that defendants use to threaten plaintiffs in post-discovery settlements. Indeed, plaintiffs now want excessive discovery, providing a rationale for why a court should deny some of a plaintiff's discovery requests.

We conclude by analyzing discovery in a *screening* setting where an uninformed plaintiff makes settlement offers to an informed defendant. Because plaintiffs now hold the bargaining power, discovery harms defendants by reducing their information advantage. Discovery reduces total litigation costs with strong defendants because it can prevent a trial. However, it raises total litigation costs with intermediate defendant types that would settle immediately absent discovery, but now wait until after discovery to settle. The social value of discovery hinges on the relative likelihoods of these two groups: with enough intermediate types, zero discovery may be optimal; with enough strong types, the optimal level of discovery exceeds its level in a signaling setting. Even so, the optimal extent of discovery remains sharply limited; and plaintiffs always seek socially excessive discovery.

Broader interpretations of "discovery". Discovery is, in essence, a costly way to publicly reveal information that is indicative of what will happen in a trial, when the parties can settle before or after the information revelation. Our analysis applies to other settings with this structure.

Summary Jury Trial. Summary jury trials were first proposed by federal district judge Thomas Lambros. See Lambros (1985). Under this procedure, attorneys present summaries of their cases to a mock jury arranged by the court. Summary trials are designed to show litigants how a trial jury might evaluate a case. The jury's "verdict" is generally non-binding.[4] If the parties do not settle after a summary jury trial, they go to a real trial with a new jury. A summary trial corresponds to "discovery", where a judge can decide on how abbreviated the summary jury trial is, i.e., on the level of "discovery". Each attorney has a specific amount of time (e.g., an hour) in which to present a case summary.[5] Lambros (1993) reports that "over 82 percent of the summary jury trial cases were resolved more quickly than the average of comparable cases that were not assigned to summary jury trial. On average, assigning a case to summary jury trial reduced the time a case remained pending by 337 days, or about 11 months". Like discovery, summary jury trials shift information acquisition forward at a cost. Unlike discovery, summary jury trials may impose costs on third parties—the time of judges and juries. We discuss how this affects results in Section 8.

Court-Annexed Arbitration. Court-Annexed Arbitration is similar to a summary jury trial except that the mock trial is in front of a panel of attorneys/arbitrators, not a jury. Studies find that the design promotes settlement pre and post arbitration. Lind and Shapard (1983) report in a Federal Judicial Center analysis that "court-annexed arbitration can serve as an effective deadline for case preparation, substituting for trial not as a forum for case resolution but as a stimulus to settlement."

Tax Audits. The benefits of shifting some information revelation and information expenses forward extend to other contexts. For example, the IRS can pre-commit to the extent of an audit. The IRS publishes a transfer pricing audit road map on its website, detailing the steps and estimated time of each step. After an audit, a tax payer can try to settle with IRS by making an "offer in compromise". Our article suggests that it may make sense to allow for settlement offers prior to an audit. The IRS can send a warning letter about a future audit, giving a tax payer time to formulate an offer.[6]

[4] In contrast to the original version of the summary jury trials in the federal court program in the Northern District of Ohio, some later versions in New York, Nevada, South Carolina and California stipulate that the outcomes are binding and that parties cannot appeal the decisions. See Hannaford-Agor (2012).

[5] See Posner (1986) or Webber (1989) for more details. By 2011, six state courts had implemented some form of summary jury trial according to a study by National Center for State Courts reported in Hannaford-Agor (2012).

[6] The economics literature typically assumes that tax audits always reveal a tax payer's private information (e.g., Border and Sobel (1987), Reinganum and Wilde (1985) and Mookherjee and Png (1989)). In practice, audits may be

Settle-able disputes. Our finding that the optimal level of discovery is positive but limited extends to any dispute in which the parties can settle. It implies that it is optimal to commit to "pauses" in the investigation stage to allow the parties to consider settling and eliminating the remaining investigation. For example, if an employee complains that a supervisor didn't give an appropriate promotion, instead of committing to getting to the bottom of the matter, an organization can facilitate a mid-investigation meeting between the parties, allowing them to "settle", perhaps with an offer of a promotion or bonus, which, if accepted, would close the case. We argue that the parties should have a chance to settle both prior to the investigation, as well as at stages during the investigation. That is, breaking the investigation into stages facilitates settlement, reducing costs.

Discovery Literature. Hay (1994) assumes that greater discovery raises a plaintiff's chance of winning against a defendant, so that more discovery always hurts defendants. He studies how this affects a defendant's ex-ante incentive to take precautions. In Sobel (1989) both the plaintiff and defendant have private information, but discovery only reveals the defendant's private information. A key difference from our model is that Sobel assumes that a defendant can only make an offer prior to discovery, whereas post-discovery the plaintiff can make an offer. This exogenous structure precludes signaling via an offer's timing, which is key in our model. Sobel shows that discovery in this form of mandatory disclosure hurts the defendant. Shavell (1989) considers a setting where an offer can only be made by an uninformed defendant after discovery. As a result, mandatory discovery hurts the informed plaintiff.

We model discovery as an inevitable step in the litigation process unless the litigants settle and the extent of the discovery is pre-determined by the court. Schrag (1999) lets each litigant choose their discovery effort, where the return from discovery effort is higher when an opponent is weaker. Each litigant has incentive to play tough in the pre-trial settlement to discourage the opponent's discovery effort. Schrag shows that if the court exogenously limits the extent of discovery, it can raise the chances of a pre-discovery settlement. Schwartz and Wickelgren (2009) model discovery

less than fully revealing.

as a conscious choice by the uninformed party. In their screening model, an uninformed defendant makes settlement offers pre- and post-discovery. A defendant's low pre-discovery offer keeps the threat of discovery credible in case its offer is rejected. Farmer and Pecorino (2005) incorporate mandatory disclosure as a conscious choice of the uninformed party in the Reinganum and Wilde (1986)'s signaling model. They allow the informed party to voluntarily disclose information, but do not allow for pre-discovery settlement. They find that discovery is not used if the informed party makes take-it-or-leave-it offers.

Cooter and Rubinfeld (1994) study discovery in a model where the settlement outcome solves a Nash bargaining problem—there is no signaling or screening via explicit settlement offers. Discovery changes the distributions from which the litigating parties' subjective beliefs are drawn. They show that if discovery narrows the gap between the means of the two parties' distributions or reduces the variances of the distributions, then trials become less likely. However, discovery can also increase the gap in the means of the two parties' distributions by uncovering information that makes at least one party more pessimistic about trial outcomes, raising the likelihood of a trial.

An implicit premise in our model is that a defendant cannot costlessly disclose its private information—if a defendant had evidence that was known to encapsulate its private information that it could just hand over, then private information would unravel. The strongest type would want to immediately reveal the strength of its case, and then progressively weaker types would follow suit (Hay 1994). But, in many settings, a defendant may not be able to convey its private information.[7] For example, a defendant may know whether there is damaging evidence against it in a class of documents. A defendant that knows this evidence does not exist would like to convey that non-existence, but it has no way to directly do so: a defendant with damaging evidence can conceal it and mimic the non-existence of evidence. In this circumstance, discovery that lets an uninformed plaintiff examine a subset of documents that is not chosen by the defendant is one way

[7]So, too, private information may not unravel if the act of disclosing is itself costly (see Sobel 1989). Shavell (1989) offers other reasons.

to (stochastically) and credibly uncover the existence or non-existence of damaging evidence. The possible sanction that a court can impose for refusing to turn over documents or other evidence in discovery is key because it makes the absence of incriminating evidence informative. Alternatively, a defendant can *indirectly* convey its private information by signaling via its settlement offer.

Mediation has been studied as a way to facilitate information revelation in the litigation process. In Doornik (2014), mediation reduces information costs, not by producing private information directly as with discovery, but rather by reducing the downside of voluntary disclosure of verifiable information. The idea is that if a litigant directly discloses information to an opponent, the opponent can form better trial strategies; but an opponent cannot exploit disclosure to a mediator.

The article's outline is as follows. Section 2 sets out the model. Section 3 characterizes equilibrium outcomes for a given level of discovery. Section 4 analyzes how the extent of discovery affects total litigation costs. Section 5 shows how the optimal extent of discovery varies with the parameters describing the legal environment. Section 6 studies impact from discovery's cost structures, which can be determined by the court's decision on what information sources to target in discovery. Section 7 contrasts outcomes in our strategic game with a single decision-maker's problem of dividing information acquisition into two stages. Section 8 considers extentions such as more rounds of discovery, fixed procedural trial costs, trial cost savings from successful discovery and costs of trial incurred by the court. Section 9 explores discovery in a screening setting. Section 10 concludes. Most proofs are in an Appendix.

2 The model

We model pre-trial discovery, and its impact on the likelihood of ex ante and interim settlement in negotiations between a defendant and plaintiff, where the defendant has private information about the probability it will be found liable by the court in a trial for damages it may have imposed on the plaintiff. The level of potential damage liability is public information and normalized to 1. The defendant has private information about the probability that it would be found *not* liable at a trial. Conditional on this private information, the probability the defendant is found *not* liable at a trial is $\theta \in [\underline{\theta}, \bar{\theta}] \subset (0, 1)$ where $\theta \sim F(\cdot)$. That is, the defendant privately observes a signal of the strength of the evidence against itself and the higher is θ, the stronger is the defendant.

All parties are risk neutral. Thus, a defendant seeks to minimize the sum of its expected payment to the plaintiff plus its own legal costs, and a plaintiff seeks to maximize the expected payment from the defendant less its own legal costs.

Our base model features one round of discovery. A defendant has two opportunities, pre- and post-discovery, to make settlement offers to the plaintiff. At $t = 1$, prior to discovery, a defendant can make a take-it-or-leave-it settlement offer. If the plaintiff accepts, then they settle and the suit is withdrawn. If the offer is rejected, then the legal process moves on to discovery, where a defendant's private information is publicly revealed with probability $\pi \in [0, 1]$. We say that discovery succeeds if a defendant's private information is revealed; otherwise discovery is said to fail. At $t = 2$, post-discovery, a defendant can again make a settlement offer. If the plaintiff accepts, then they settle and the suit is withdrawn. If the offer is rejected, then the case proceeds to trial, where a court determines whether the defendant is liable and must pay damages to the plaintiff.

The defendant and plaintiff incur investigation costs as they proceed through the judicial process. We initially assume that all legal costs are informational in nature, dealing solely with the gathering and assessing of information about whether a defendant is liable. If a case proceeds all the way to trial, the plaintiff would incur a total investigation cost of $c_p > 0$, and the defendant would incur $c_d > 0$. We assume that $1 - \bar{\theta} > c_p$ so that the plaintiff always has a case with a positive expected value from trial.[8] Investigation costs are incurred both in discovery and at trial. The

[8]Effectively, these are cases that have survived motions to dismiss.

extent of discovery is characterized by the probability π that discovery uncovers the defendant's private information, θ. More extensive discovery costs more: the proportion of the maximum investigation costs incurred in discovery, $\rho(\pi) \in [0,1]$, is a strictly increasing, twice differentiable function of π, with $\rho(0) = 0$ and $\rho(1) = 1$. Thus, after discovery, if the two parties do not settle, the plaintiff would incur additional costs of $(1 - \rho(\pi))c_p$ at trial, and the defendant would incur $(1-\rho(\pi))c_d$.[9] We first analyze a linear cost structure where the proportion of investigation costs incurred in discovery rises one-for-one with the extent of discovery, i.e., $\rho(\pi) = \pi$. A strictly convex discovery cost structure captures the possibility that discovery focuses on less costly (per unit of information) sources of information, leaving more costly sources to a trial; a strictly concave cost structure captures the possibility that discovery focuses on more costly sources of information. We first take the level of discovery—the probability π that discovery succeeds—and the cost structure of discovery as given, and solve for how they affect the timing of when different defendant types make their first offers, and the sizes of these offers. We then characterize the discovery level that minimizes the expected total litigation costs of both litigants. It helps to define $c \equiv c_p + c_d$.

Let d denote the discovery outcome: $d = \theta$ if discovery uncovers θ, and $d = \emptyset$ otherwise. The defendant's strategy consists of a pair of settlement offer strategies x_1 and x_2, where, $x_1 : \theta \mapsto R_+ \bigcup \{N\}$ gives the pre-discovery settlement offer (with N denoting no offer) and $x_2 : (\theta, d, x_1) \mapsto R_+ \bigcup \{N\}$ gives the post-discovery settlement offer when there is no pre-discovery settlement.

The plaintiff's strategy consists of two rejection probability functions, p_1 and p_2. Here $p_1 : x_1 \mapsto [0,1]$ (with $p_1(N) = 1$) gives the probability that the plaintiff rejects a pre-discovery settlement offer x_1, and $p_2 : (x_1, d, x_2) \mapsto [0,1]$ (with $p_2(\cdot,\cdot,N) = 1$) gives the probability that the plaintiff rejects a post-discovery settlement offer x_2.

[9] In our base model, discovery outcomes do not affect the level of investigation costs that would be incurred at trial. This approximates a setting where the total amount of information (private or non-private) to be uncovered does not depend on whether discovery succeeds. We later study the case when successful discovery reduces trial costs.

We let $b_1(x_1)$ represent a plaintiff's beliefs about a defendant's private information after seeing x_1; and $b_2(x_1, d, x_2)$ represents his beliefs after seeing x_1, x_2 and the discovery outcome d. Abusing notation slightly, we let $b_1(x_1)$ and $b_2(x_1, d, x_2)$ denote point beliefs when beliefs are degenerate.

Definition: A profile $(x_1^*, x_2^*, p_1^*, p_2^*, b_1^*, b_2^*)$ forms an equilibrium if and only if (1) given beliefs (b_1^*, b_2^*), the strategy (x_1^*, x_2^*) maximizes the defendant's expected future payoff at any point in time, (2) given $(x_1^*, x_2^*, b_1^*, b_2^*)$, the strategy p_1^*, p_2^* maximizes the plaintiff's expected future payoff at any point in time, and (3) the plaintiff's beliefs (b_1^*, b_2^*) obey Bayes' Rule whenever possible.

In our dynamic model, "full separation" does not mean that all types separate at the very beginning of the game; but rather that separation occurs before the trial. Thus, the equilibrium is fully separating if the plaintiff's beliefs about the defendant's type become degenerate for each θ prior to the trial. We next characterize a fully-separating equilibrium, establishing the existence and essential uniqueness of a "universally divine equilibrium" using the refinements in Banks and Sobel (1987). We say "essentially" unique, because there is latitude in specifying off-equilibrium beliefs, as well as latitude in specifying offers that are always rejected along the equilibrium path.

Preliminaries: In any equilibrium, with take-it-or-leave-it offers, a defendant's settlement offer at $t = 2$ extracts all surplus when discovery uncovers θ, leaving the plaintiff indifferent between accepting the offer and going to trial, i.e., $x_2(\theta, \theta, x_1) = 1 - \theta - (1 - \rho(\pi))c_p$, and the plaintiff always accepts this offer. Thus, $p_2(x_1, \theta, x_2) = 1$ for $x_2 \geq 1 - \theta - (1 - \rho(\pi))c_p$ and $p_2(x_1, \theta, x_2) = 0$ otherwise. To reduce notation, we abuse notation slightly, and let $p_2(x_2) = p_2(N, \emptyset, x_2)$ denote the rejection probability when discovery fails and no offer was made prior to discovery. We first derive a basic property of the cost function $\rho(\pi)$:

Lemma 1. *If $\rho(\pi)$ is strictly convex, then (a) $\pi > \rho(\pi)$, and (b) $\frac{\rho(\pi)}{\pi}$ is strictly increasing in π. If $\rho(\pi)$ is strictly concave, then (a) $\pi < \rho(\pi)$, and (b) $\frac{\rho(\pi)}{\pi}$ is strictly decreasing in π.*

Proof: See the Appendix. □

3 Separation via offer amount and timing

A separating equilibrium is described by a cutoff type $\hat{\theta} \in [\underline{\theta}, \bar{\theta}]$ such that prior to discovery, any weaker defendant $\theta < \hat{\theta}$ makes an offer and settles with positive probability at $t = 1$, whereas any stronger defendant $\theta > \hat{\theta}$ waits until after discovery to make its first (acceptable) offer. Thus, defendant types partition themselves into groups of weak and strong defendants, $[\underline{\theta}, \hat{\theta}]$ and $(\hat{\theta}, \bar{\theta}]$. If discovery uncovers θ, then a strong defendant makes an (accepted) offer that extracts all surplus. Otherwise, within each group, types separate further via their proposed settlement amounts, with weaker types proposing more generous settlements. These offers leave a plaintiff indifferent between accepting and rejecting each offer (given separating beliefs), and the plaintiff's probability of rejection declines with the size of the settlement offer in a way that makes it incentive compatible for defendants to reveal their types via their settlement offers. This further separation within the populations of weak and strong defendants is in the spirit of Reinganum and Wilde (1986), who analyze settlement offers by a privately-informed plaintiff when there is no discovery. We prove the existence and uniqueness of a universally divine equilibrium (Banks and Sobel 1987) in the Appendix; the proof builds on that in Reinganum and Wilde. We recursively solve the game, first analyzing the post-discovery stage and then the pre-discovery stage.

Post-discovery settlement offers. First consider a type that did not make a pre-discovery offer ($x_1 = N$). Absent a post-discovery settlement at $t = 2$, the parties will go to trial, incurring additional trial costs of $(1-\rho)c_p$ and $(1-\rho)c_d$, where we omit the dependence of ρ on π where it does not cause confusion. If discovery uncovers θ, then the defendant makes the (accepted) offer $x_2(\theta, \theta, x_1) = (1-\theta) - (1-\rho)c_p$, extracting all surplus given that information from the plaintiff.

Now suppose discovery does not reveal θ. If the defendant's offer of x_2 is accepted, its payoff is $-x_2$. If its offer is rejected, the two parties go to trial. At trial the defendant expects to pay $1-\theta$ to the plaintiff and incur trial costs $(1-\rho)c_d$. Thus, a type θ defendant's expected payoff when the plaintiff rejects its post-discovery settlement offer x_2 with probability $p_2(x_2) = p_2(N, \emptyset, x_2)$ is:

$$(1 - p_2(x_2))[-x_2] + p_2(x_2)[-(1 - \theta) - (1 - \rho)c_d].$$

A type θ defendant's settlement offer x_2 maximizes this payoff. The associated first-order condition is:

$$p_2'(x_2)[-(1 - \theta) - (1 - \rho)c_d + x_2] - 1 + p_2(x_2) = 0.$$

The defendant's equilibrium settlement offer leaves the plaintiff indifferent between accepting and rejecting given *separating* beliefs. Therefore, the defendant's payoff must be maximized by $x_2(\theta, \emptyset, N) = (1 - \theta) - (1 - \rho)c_p$. Substituting this offer into the first-order condition yields

$$-p_2'(x_2)(1 - \rho)c - 1 + p_2(x_2) = 0, \tag{1}$$

where we recall that $c \equiv c_d + c_p$. The weakest type that did not make an offer prior to discovery is $\hat{\theta}$. The boundary condition reflects that the separating equilibrium is efficient, and hence $\hat{\theta}$'s offer must be rejected with probability 0, i.e., $p_2((1 - \hat{\theta}) - (1 - \rho)c_p) = 0$. Solving the differential equation (1) for the probability with which the plaintiff rejects the defendant's offer yields

$$p_2(x_2) = 1 - \exp\Big(\frac{x_2 - ((1 - \hat{\theta}) - (1 - \rho)c_p)}{(1 - \rho)c}\Big).$$

Substituting for $x_2(\theta, \emptyset, N) = (1 - \theta) - (1 - \rho)c_p$ yields the equilibrium probability that a type $\theta \in [\hat{\theta}, \bar{\theta}]$ has its offer rejected when discovery fails to reveal its private information. Using $r_2(\theta)$ to denote this equilibrium probability of rejection for $\theta \in [\hat{\theta}, \bar{\theta}]$, we have

$$r_2(\theta) = 1 - \exp\Big(-\frac{\theta - \hat{\theta}}{(1 - \rho)c}\Big). \tag{2}$$

This probability of rejection rises with θ, reflecting that stronger defendant types make less generous offers, and to make it unattractive for weaker defendant types to mimic those less generous offers, they must face a higher probability of rejection.

Strong defendants are always better off when discovery reveals their private information because settlement offers are unaffected by the discovery outcome; but when discovery fails, a plaintiff must sometimes reject offers, in order to make it incentive compatible for a defendant's offer to reveal the true strength of its case. As we discussed earlier, if a defendant's private information concerns the nonexistence of evidence against itself, then even though a stronger defendant wants to reveal this lack of evidence in order to avoid having its settlement offers rejected, it has no incentive compatible way to do so directly. This is because a weaker defendant always has an incentive to under-report/conceal unfavorable evidence, mimicking a stronger defendant. As a result, a defendant can only credibly convey its private information indirectly via the timing and size of its settlement offer; and to provide the correct truth-telling incentives, a plaintiff must be more likely to reject less generous offers, causing the parties to inefficiently incur trial costs.

Now consider a weak defendant type, $\theta < \hat{\theta}$, whose pre-discovery offer was rejected, when discovery failed to uncover its private information. On the equilibrium path, its pre-discovery offer reveals its type. As a result, on the equilibrium path, its post-discovery offer is $x_2(\theta, \emptyset, x_1^*) = 1 - \theta - (1-\rho)c_p$. Off-equilibrium path beliefs are not uniquely pinned down. For convenience, we assume that when discovery fails, the plaintiff's belief after a $t = 2$ offer is unchanged from that after the $t = 1$ offer. Then at $t = 2$ the defendant will make a take-it-or-leave-it offer that leaves the plaintiff indifferent between accepting and rejecting given the belief based on the first offer alone; or go to trial if it was a strong-enough type that, off-the-equilibrium path, mistakenly made an excessively generous pre-discovery offer x_1, associated with a weak type $\check{\theta}$, where

$$x_1 = (1 - \check{\theta}) - (1-\rho)c_p < (1-\theta) - (1-\rho)c_d \Leftrightarrow \theta - \check{\theta} > (1-\rho)(c_d + c_p).$$

Pre-discovery settlement offers. On the equilibrium path, strong defendant types $\theta > \hat{\theta}$ do not make offers prior to discovery, so consider the settlement offers of weaker types $\theta < \hat{\theta}$. If a pre-discovery offer x_1 is rejected and discovery reveals θ, then the defendant's post-discovery offer will be $x_2(\theta, \theta, x_1) = (1-\theta) - (1-\rho)c_p$. If discovery fails to uncover θ, the plaintiff's belief (on the equilibrium path) is the same as that based only on the first offer. Because pre-discovery offers fully separate those types that make offers, x_1 leads to degenerate beliefs. Thus, on the equilibrium path, $b_1^*(x_1(\theta)) = \theta$, so the defendant's post-discovery offer of $1-b_1^*(x_1)-(1-\rho)c_p$ when discovery fails is the same as that when discovery uncovers θ. It follows that the separating pre-discovery offer that leaves the plaintiff indifferent between accepting and rejecting it is $x_1(\theta) = (1-\theta)-c_p$. Therefore, $b_1^*(x_1) = 1 - x_1 - c_p$. Then, given a pre-discovery offer x_1 (potentially off-the-equilibrium path), the post-discovery offer would be $x_2(\theta, \emptyset, x_1) = (1 - b_1^*(x_1)) - (1-\rho)c_p = x_1 + \rho c_p$ (as long as x_1 was not mistakenly so generous that the defendant prefers to go to trial).

Thus, along the equilibrium path, a type θ defendant's expected payoff when the plaintiff rejects its pre-discovery settlement offer x_1 with probability $p_1(x_1)$ is:

$$(1 - p_1(x_1))[-x_1] + p_1(x_1)[\pi(-(1-\theta) + (1-\rho)c_p) + (1-\pi)(-x_1 - \rho c_p) - \rho c_d].$$

That is, the plaintiff accepts x_1 with probability $1-p_1(x_1)$, and with probability $p_1(x_1)$ the plaintiff rejects it and the case proceeds to discovery. After discovery, if θ is not revealed, the defendant offers $x_1 + \rho c_p$, which is accepted. A weak defendant's optimal pre-discovery settlement offer x_1 maximizes its expected payoff, solving the first-order condition:

$$p_1'(x_1)[\pi(-(1-\theta) + (1-\rho)c_p) + (1-\pi)(-x_1 - \rho c_p) - \rho c_d + x_1] - 1 + p_1(x_1)\pi = 0.$$

Substituting in the equilibrium pre-discovery offer $x_1(\theta) = (1-\theta) - c_p$, yields

$$-p_1'(x_1)\rho(c_p + c_d) - 1 + p_1(x_1)\pi = 0.$$

The boundary condition reflects that the pre-discovery settlement offer of the weakest type $\underline{\theta}$ is always accepted. Solving this differential equation yields the probability that the plaintiff rejects pre-discovery settlement offer x_1:

$$p_1(x_1) = \frac{1}{\pi}[1 - \exp\Big(\frac{x_1 - (1 - \underline{\theta} - c_p)}{\frac{\rho}{\pi}c}\Big)].$$

Substituting for $x_1(\theta) = (1-\theta) - c_p$ yields the equilibrium probability $r_1(\theta)$ that a weak defendant type $\theta < \hat{\theta}$'s pre-discovery offer is rejected. For $\rho > 0$, we have

$$r_1(\theta) \;=\; \frac{1}{\pi}[1 - \exp\Big(-\frac{\theta - \underline{\theta}}{\frac{\rho}{\pi}c}\Big)]. \tag{3}$$

The pre-discovery probability of rejection $r_1(\theta)$ differs from its post-discovery counterpart $r_2(\theta)$ in two ways. First, the pre-discovery probability of rejection $r_1(\theta)$ is scaled up by $\frac{1}{\pi}$. This is because the discovery signal is weaker than the trial "signal": it reveals θ with probability π, rather than with probability 1. A stronger future signal makes separation easier by reducing a weaker defendant types gains from mimicking a stronger type, thereby reducing the required probability of rejection. Second, the denominator in the exp term for pre-discovery is $\frac{\rho}{\pi}c$, whereas that for post-discovery is $(1-\rho)c$. These terms represent the cost per unit of revelation probability of the next signal. The "discovery signal" costs ρc and reveals θ with probability π, so the per unit cost of information revelation is $\rho c/\pi$. Following failed discovery, the cost of discovery is sunk, so the "trial signal" costs $(1-\rho)c$ and it reveals the private information with probability one, so the per unit cost of the trial signal is $(1-\rho)c$. The higher is the cost per unit of information of the next "signal", the greater is the cost of rejection to a defendant, making it easier to separate types, and thus lowering the probability of rejection required to induce incentive compatible signaling. The sum of the investigation costs to both the defendant *and* the plaintiff enter because the defendant indirectly bears the plaintiff's discovery costs whenever a settlement fails: the defendant can no longer use these sunk investigation costs to threaten a plaintiff when making a post-discovery offer.

Cutoff type, $\hat{\theta}(\pi)$. Defendant type $\hat{\theta}$ is indifferent between making a pre-discovery offer that reveals its type and waiting until after discovery to make its first offer, where, as it is the weakest type that waits, its offer is always accepted. $\hat{\theta}$'s payoff from its equilibrium pre-discovery offer is:

$$r_1(\hat{\theta})[-(1-\hat{\theta})+(1-\rho)c_p-\rho c_d]+(1-r_1(\hat{\theta}))[-(1-\hat{\theta})+c_p] = -(1-\hat{\theta})+c_p-\rho(c_p+c_d)r_1(\hat{\theta}).$$

That is, if $\hat{\theta}$ makes a pre-discovery settlement offer then it incurs discovery costs ρc_d and it fails to extract the plaintiff's discovery costs ρc_p only when its offer is rejected. If, instead, $\hat{\theta}$ waits until after discovery to make its first offer, then $\hat{\theta}$ receives

$$-\rho c_d-(1-\hat{\theta})+(1-\rho)c_p = -(1-\hat{\theta})+c_p-\rho(c_p+c_d).$$

That is, if $\hat{\theta}$ delays making an offer until after discovery, it always incurs discovery costs ρc_d and it always fails to extract ρc_p. For $\hat{\theta}$ to be indifferent between making a pre-discovery offer and waiting until after discovery to make a first offer, its pre-discovery offer must always be rejected: $r_1(\hat{\theta})=1$. If the extent of discovery π satisfies $\pi < 1-\exp\Big(-\frac{\bar{\theta}-\underline{\theta}}{\frac{\rho(\pi)}{\pi}c}\Big)$, then $\hat{\theta}$ solves $r_1(\hat{\theta})=1$:

$$\frac{1}{\pi}[1-\exp\Big(-\frac{\hat{\theta}-\underline{\theta}}{\frac{\rho(\pi)}{\pi}c}\Big)] = 1.$$

Writing the rejection probability and cutoff type as functions of π to emphasize their dependence on π we now show that more discovery reduces the probability that pre-discovery offers are rejected:

Lemma 2. *The rejection probability $r_1(\theta,\pi)$ is strictly decreasing in π for any $\theta > \underline{\theta}$.*

Proof: See the Appendix. □

Inspection of $r_1(\theta,\pi) = \frac{1}{\pi}[1-\exp\Big(-\frac{\theta-\underline{\theta}}{\frac{\rho(\pi)}{\pi}c}\Big)]$ reveals that greater discovery affects the pre-discovery probability of rejection $r_1(\theta,\pi)$ in two ways. First, a higher probability of uncovering θ in discovery directly reduces a weaker defendant's gain from mimicking a stronger defendant, which reduces the rejection probability needed to induce separation, reflected in the term $\frac{1}{\pi}$. Second, if the fraction of total investigation costs incurred in discovery is strictly convex in the extent of discovery, increases in π raise the per-unit information cost of discovery, $\frac{\rho(\pi)}{\pi}$ (Lemma 1), harshening the consequences of having offers rejected. In turn, this further reduces the pre-discovery rejection probabilities needed to induce separation. If, instead, the fraction of total investigation costs incurred in discovery is strictly concave in the extent of discovery, the effect through $\frac{\rho}{\pi}$ raises the required rejection probability. However, this effect is dominated by the direct effect through $\frac{1}{\pi}$.

Let $\bar{\pi}$ be the solution to $r_1(\overline{\theta}, \pi) = 1$. Lemma 2 implies that for any $\pi \geq \bar{\pi}$, $r_1(\overline{\theta}, \pi) \leq 1$. Notice that $r_1(\theta, \pi)$ strictly increases in θ. Thus, for any $\pi \geq \bar{\pi}$ and any $\theta \in (\underline{\theta}, \overline{\theta})$, $r_1(\theta, \pi) < 1$. This implies all defendant types prefer to make the first acceptable offer prior to discovery, i.e. $\hat{\theta} = \bar{\theta}$ for any $\pi \geq \overline{\pi}$. Also, because $r_1(\overline{\theta}, 1) < 1$, we have $\bar{\pi} < 1$. Thus, the cutoff type is given by,

$$\hat{\theta}(\pi) \equiv \begin{cases} \underline{\theta} - \ln(1-\pi)\frac{\rho(\pi)}{\pi}c, & \text{if } \pi \in [0, \bar{\pi}), \\ \bar{\theta}, & \text{if } \pi \geq \bar{\pi}, \end{cases}$$

Proposition 1. *For a fixed π, define $\bar{x}_1 \equiv (1-\underline{\theta}) - c_p$, $\underline{x}_1 \equiv (1-\hat{\theta}(\pi)) - c_p$, $\bar{x}_2 \equiv (1-\hat{\theta}(\pi)) - (1-\rho)c_p$, $\underline{x}_2 \equiv (1-\bar{\theta}) - (1-\rho)c_p$.*

The equilibrium values of $(x_1^*, x_2^*, p_1^*, p_2^*)$ are uniquely determined. Equilibrium settlement offers are given by

$$\begin{aligned} x_1^*(\theta) &= (1-\theta) - c_p, \ \textit{if} \ \theta \in [\underline{\theta}, \hat{\theta}(\pi)]; \quad x_1^*(\theta) = N, \ \textit{if} \ \theta > \hat{\theta}(\pi) \\ x_2^*(\theta, \emptyset, x_1) &= (1-\theta) - (1-\rho)c_p, \ \textit{if} \ \theta \in [\underline{\theta}, \bar{\theta}]. \end{aligned}$$

The equilibrium rejection probabilities with which the plaintiff rejects a defendant's offers are:

$$p_1^*(x_1) = \frac{1}{\pi}[1 - \exp\Big(\frac{x_1 - (1 - \underline{\theta} - c_p)}{\frac{\rho}{\pi}c}\Big)],\ x_1 \in [\underline{x}_1, \bar{x}_1]$$

$$p_2^*(x_2) = p_2^*(N, \emptyset, x_2) = 1 - \exp\Big(\frac{x_2 - (1 - \hat{\theta}(\pi) - (1-\rho)c_p)}{(1-\rho)c}\Big),\ x_2 \in [\underline{x}_2, \bar{x}_2]$$

$$p_i^*(x_i) = 1,\ x_i < \underline{x}_i; \quad p_i^*(x_i) = 0,\ x_i > \bar{x}_i, \ \ i = 1, 2.$$

Proof: See the Appendix. □

The separating equilibrium uniquely pins down the sizes of accepted offers and their probabilities of acceptance, as well as the beliefs following those offers:

$$b_1^*(x_1) = 1 - x_1 - c_p, \text{ if } x_1 \in [\underline{x}_1, \bar{x}_1]; \qquad b_1^*(N) = F(\theta|\theta > \hat{\theta}(\pi)),$$

$$b_2^*(N, \emptyset, x_2) = 1 - x_2 - (1-\rho)c_p, \text{ if } x_2 \in [\underline{x}_2, \bar{x}_2]; \qquad b_2^*(x_1, \emptyset, x_2) = b_1^*(x_1), \text{ if } x_1 \neq N.$$

There is, however, some freedom in specifying the sizes of offers that are always rejected, and in specifying a plaintiff's beliefs following offers that are not made on the equilibrium path ($x_i > \bar{x}_i$ or $x_i < \underline{x}_i$). In the Appendix, we show that any "universally divine" equilibrium must be fully separating. Figure 1 illustrates the equilibrium when $\pi \in (0, \bar{\pi})$.

When discovery is more informative (π is higher), more types make pre-discovery offers—the equilibrium cutoff $\hat{\theta}(\pi)$ is higher. With enough discovery, i.e., if $\pi \geq \bar{\pi}$, then $\hat{\theta} = \bar{\theta}$: all types make pre-discovery offers, immediately revealing their types. When $\pi = 0$, then $\hat{\theta} = \underline{\theta}$: all types wait until after discovery to make offers—this case essentially reduces to the equilibrium in Reinganum and Wilde (1986), save that there are two pre-trial dates and no defendant makes a pre-discovery offer, supported by the belief that any defendant that does so is the weakest type.

Recall that cutoff type $\hat{\theta}(\pi)$'s pre-discovery offer is always rejected: $r_1(\hat{\theta}(\pi), \pi) = 1$. The rejection probability rises with θ and falls with the extent of discovery, π. Therefore, $\hat{\theta}(\pi)$ must rise with the extent of discovery in order to keep the rejection probability equal to one:

Corollary 1. *The cutoff type $\hat{\theta}(\pi)$ is strictly increasing in π for $\pi < \bar{\pi}$.*

Collectively, Lemma 2, Proposition 1 and Corollary 1 have significant empirical content. They imply that: (1) cases settle early with positive probability, with weaker cases being more likely to settle (with more generous settlement terms for the plaintiff); (2) cases that are rejected pre-discovery are weak and will settle before trial; (3) defendants who hold off making a first offer until after discovery, make less generous offers, and may end up being rejected, in which case there is a trial; and (4) greater discovery induces more defendants to make pre-discovery settlement offers, raises the probability that any given pre-settlement offer is accepted, and reduces the number of cases that go to trial. Our analysis also highlights selection issues that affect inference. Because stronger defendants are more likely to go to trial, one must be careful not to infer from a high defendant success rate that most plaintiff cases were weak.

4 The impact of discovery on expected litigation costs

In this section we derive how the level of discovery affects the total investigation costs that the two parties expect to incur in equilibrium given an arbitrary defendant type θ. We also determine how discovery affects the likelihood that a case proceeds all the way to trial.

We begin with the observation that no discovery and full discovery lead to the same effective outcome: all information is revealed at once, either at the trial (with no discovery), or at discovery (with full discovery). That is, fully exhaustive discovery that uncovers all information is akin to shifting the timing of a trial forward, and hence delivers the same expected litigation costs as a legal system with no discovery, where litigants go directly to trial if they fail to settle. Consequently,

Lemma 3. *Total litigation costs are the same with no discovery as with full discovery.*

Proof: See the Appendix. $\square$

For any θ, define $\hat{\pi}(\theta)$ as the solution of $\hat{\theta}(\pi) = \theta$. Thus, given discovery π, a defendant type θ is weak, i.e., $\theta \leq \hat{\theta}(\pi)$, if and only if the extent of discovery $\pi \geq \hat{\pi}(\theta)$. In other words, when discovery is extensive relative to a type, it induces the defendant to make an acceptable pre-discovery settlement offer.

From Lemma 1, we have,

Corollary 2. *The cutoff $\hat{\pi}(\theta)$ strictly increases in θ and $\hat{\pi}(\bar{\theta}) = \bar{\pi}$.*

In the remainder of this section, we suppose that $\rho(\pi) = \pi$, i.e., discovery costs rise one-for-one with the extent of discovery. We defer analysis with more general (i.e., non-linear) discovery cost functions to Section 6.

We begin by decomposing the impact of discovery according to whether the defendant type θ makes offer pre- or post-discovery.

Weak defendants. When discovery is extensive relative to θ, i.e., when $\pi \geq \hat{\pi}(\theta)$, type θ is a weak defendant type that always settles prior to a trial. Litigants incur costs of $\rho(\pi)(c_p + c_d) = \rho(\pi)c$ only if a pre-discovery offer is rejected, and the probability of rejection is $r_1(\theta, \pi)$. Thus, expected total litigation costs incurred with a type θ weak defendant are:

$$C^w(\theta, \pi) \equiv r_1(\theta, \pi)\rho(\pi)c = \frac{\rho(\pi)}{\pi}(1 - \exp\Big(-\frac{\theta - \underline{\theta}}{\frac{\rho(\pi)}{\pi}c}\Big))c.$$

Substituting $\rho(\pi) = \pi$ into $C^w(\theta, \pi)$ reveals that $C^w(\theta, \pi)$ depends only on the type, and *not* on the particular extent of discovery:

Proposition 2. *(Weak defendants with linear discovery) When $\rho(\pi) = \pi$, the extent of discovery does not affect expected total litigation costs for a given type θ if $\pi \geq \hat{\pi}(\theta)$.*

A weak defendant's pre-discovery settlement offer reveals its type. As a result, post-discovery, its settlement offer is always accepted, so the case never goes to trial and at most only discovery costs are incurred. However, because discovery with $\pi < 1$ is less costly and less revealing than a trial in the absence of discovery, the plaintiff must also be more likely to reject a pre-discovery settlement offer than a pre-trial offer with no discovery. These two opposing effects cancel out for weak types when $\rho(\pi) = \pi$.[10]

Strong defendants. If discovery is limited relative to a defendant's type θ, i.e., if $\pi < \hat{\pi}(\theta)$, then θ is a strong defendant type, who does not make acceptable settlement offers until after discovery. With strong defendants, litigants always incur discovery costs $\rho(\pi)c$. If discovery uncovers θ, no more legal costs are incurred as they settle after discovery. So, too, if discovery fails but a defendant's settlement offer is accepted, no additional legal costs are incurred. However, if discovery fails and a defendant's post-discovery offer is rejected, they go to trial, incurring additional costs of $(1 - \rho(\pi))c$. Thus, expected total litigation costs with discovery for a strong defendant θ are:

$$C^s(\theta, \pi) = [\rho(\pi) + (1 - \pi)r_2(\theta, \pi)(1 - \rho(\pi))]c.$$

Unlike with weak defendants, limited discovery reduces expected litigation costs with strong defendants even when $\rho(\pi) = \pi$:

Proposition 3. *(Strong defendants with linear discovery) When $\rho(\pi) = \pi$, expected total litigation costs with a type θ defendant are strictly lower with limited discovery, i.e., $\pi \in (0, \hat{\pi}(\theta))$, than with no discovery, $\pi = 0$, or extensive discovery, $\pi \geq \hat{\pi}(\theta)$.*

To see why, we rewrite expected litigation costs with and without discovery in a way that is

[10] Our analysis in Section 6 reveals that these two effects do not typically cancel: convex discovery costs raise the cost savings of limited discovery relative to full or no discovery, whereas concave costs reduce the cost savings.

more comparable to the expected litigation costs with discovery:

$$
\begin{aligned}
\text{With discovery:}\frac{1}{c}C^s(\theta,\pi) \quad &=\rho(\pi)+(1-\rho(\pi))\Big(1-\exp\Big(-\frac{\theta-\hat{\theta}(\pi)}{(1-\pi)c}\Big)\Big)(1-\pi). && (4)\\
\text{Without discovery:}\frac{1}{c}C^s(\theta,0) \quad &=1-\exp\Big(-\frac{\theta-\underline{\theta}}{c}\Big)\\
&=1-\exp\Big(-\frac{\theta-\hat{\theta}(\pi)}{c}\Big)\exp\Big(-\frac{\hat{\theta}(\pi)-\underline{\theta}}{c}\Big) \qquad \text{for any } \pi\\
&=\rho(\pi)+(1-\rho(\pi))\Big(1-\exp\Big(-\frac{\theta-\hat{\theta}(\pi)}{c}\Big)\Big). \qquad \text{for any } \pi && (5)
\end{aligned}
$$

Equality (5) follows from substituting for $\exp\big(-\frac{\hat{\theta}(\pi)-\underline{\theta}}{c}\big)$ using $r_1(\hat{\theta},\pi)=1$, adding and subtracting $\rho(\pi)$ and re-arranging. We can isolate two common components in litigation costs with or without discovery as above. As a result, comparisons of expected litigation costs with and without discovery ((4) and (5)) revolve around comparisons of $1-\exp\big(-\frac{\theta-\hat{\theta}(\pi)}{c}\big)$ with $\big(1-\exp\big(-\frac{\theta-\hat{\theta}(\pi)}{(1-\pi)c}\big)\big)(1-\pi)$.

To understand these terms, define

$$R(s)\equiv 1-\exp\left(-\frac{s}{c}\right)$$

as the probability a settlement offer must be rejected to obtain incentive compatible revelation of θ when the 'distance' between θ and the lowest type from which it must separate is s and the trial cost is c. The solid curve in Figure 2 shows that this rejection probability rises *concavely* with the separation distance s.

The construction of $\big(1-\exp\big(-\frac{\theta-\hat{\theta}(\pi)}{(1-\pi)c}\big)\big)(1-\pi)$ in (4) reflects that when discovery succeeds, there is no longer a need to signal information, eliminating all distortion in rejection probabilities. But discovery fails with probability $1-\pi$, in which case the subsequent settlement rejection probability, $1-\exp\big(-\frac{\theta-\hat{\theta}}{(1-\pi)c}\big)$, must be higher than if there were no discovery because the remaining trial cost is lower, so having an offer rejected is not as costly—it is as if the distance to separate is inflated by a factor of $\frac{1}{1-\pi} > 1$. To see how these two effects play out observe that, given θ,

$$1 - \exp\big(-\frac{\theta - \hat{\theta}}{c}\big) \quad = \quad R(\theta - \hat{\theta})$$

and

$$\begin{aligned}\big(1 - \exp\big(-\frac{\theta - \hat{\theta}}{(1-\pi)c}\big)\big)(1-\pi) \quad &= \quad (1-\pi)R\Big(\frac{\theta - \hat{\theta}}{1-\pi}\Big) \\ &= \quad (1-\pi)R\Big(\frac{\theta - \hat{\theta}}{1-\pi}\Big) + \pi R(0),\end{aligned}$$

because $R(0) = 0$. A strong type's separation distance with no discovery, $\theta - \hat{\theta}$, is a linear combination of those with (linear) discovery, 0 and $\frac{\theta-\hat{\theta}}{1-\pi}$. That is,

$$\theta - \hat{\theta} = \pi \times 0 + (1-\pi) \times \frac{\theta - \hat{\theta}}{1-\pi}.$$

Concavity of $R(\cdot)$ then means that expected distortion costs with no discovery exceed those with discovery, i.e.,

$$R(\theta - \hat{\theta}) > \pi R(0) + (1-\pi)R\Big(\frac{\theta - \hat{\theta}}{1-\pi}\Big).$$

This is reflected in Figure 2. The vertical difference between the concave function and the linear function over $\theta - \hat{\theta}$ shows the benefit of discovery to a strong type. The intuition is that separation via discovery is linear, affecting all types in exactly the same way, but endogenous separation via the size of a settlement offer is only concave, embodied in the concavity of $1 - \exp(\cdot)$, which is everywhere pointwise above the linear separation via discovery.

The value of discovery. With linear discovery, litigants do not benefit when discovery is extensive relative to θ (Proposition 2), but they do benefit when discovery is limited (Proposition 3). Thus,

Corollary 3. *(Value of discovery) Suppose $\rho(\pi) = \pi$. Then with a type θ defendant, discovery* *benefits litigants (relative to no discovery) if and only if it is limited relative to* θ*, i.e.,* $\pi \in (0, \hat{\pi}(\theta))$.

We now establish that not only are strong defendants the only ones to gain from discovery, but that among strong defendants, it is the stronger ones who benefit more from discovery. Denoting the benefit of discovery for strong defendant types by $\Delta^s(\theta,\pi) \equiv C^s(\theta,0) - C^s(\theta,\pi)$, we have:

Proposition 4. *(Stronger types gain more from linear discovery.) When* $\rho(\pi) = \pi$*, the reduction in expected litigation costs,* $\Delta^s(\theta,\pi)$*, rises with* $\theta \geq \hat{\theta}(\pi)$.

Proof:

$$\frac{1}{c}\Delta^s(\theta,\pi) = (1-\pi)[R(\theta-\hat{\theta}) - (1-\pi)R\Big(\frac{\theta-\hat{\theta}}{1-\pi}\Big)].$$

Differentiating with respect to θ yields:

$$\frac{1}{c}\Delta^s_\theta(\theta,\pi) = (1-\pi)[R'(\theta-\hat{\theta}) - R'\Big(\frac{\theta-\hat{\theta}}{1-\pi}\Big)] > 0,$$

where the inequality follows from the concavity of $R(\cdot)$. $\square$

The concavity of the rejection probability $R(s)$ means that when the separation distance s is greater, a marginal increase in s leads to a smaller marginal increase in rejection rates. In other words, a stronger type is less hurt by an inflation of the separation distance. As a result, a stronger type (among the strong ones) gains more from linear discovery. We show in Section 6, that this comparative static result carries over to general cost structures.

Discovery outcomes do not affect a *plaintiff's* payoffs—a plaintiff is always indifferent between settling early and having the litigation go to trial. The plaintiff's indifference reflects a defendant's ability to make take-it-or-leave-it offers. Thus, all gains from reductions in litigation costs due to discovery accrue to the defendant. There is an important caveat to this result: Section 8 shows that if successful discovery reduces trial costs, then a plaintiff gains from discovery because it strengthens her bargaining position even when the defendant makes offers. In fact, a plaintiff can gain so much that discovery can harm a defendant even with linear discovery costs, $\rho(\pi) = \pi$.

Discovery reduces the likelihood that the litigants go to trial. With discovery, weak defendants never go to trial—even if their pre-discovery offers are rejected, their post-discovery offers are always accepted. In contrast, absent discovery, any $\theta > \underline{\theta}$ faces a strictly positive probability that its offer is rejected, in which case it goes to trial. If discovery succeeds, strong defendant types do not go to trial either. What is more interesting is that, even conditional on discovery *failing* to reveal θ, discovery reduces the probability of trial for at least some strong types. In particular, when discovery fails, the rejection probability needed for a strong type $\theta = \hat{\theta} + \epsilon$, $\epsilon > 0$, but small, to separate from $\hat{\theta}$ is close to zero; but, absent discovery, separating from worse types (e.g., $\theta = \underline{\theta}$) demands a higher probability of rejection, and hence trial. Summarizing, we have:

Corollary 4. *Even conditional on discovery failing to reveal θ, positive discovery reduces the probability of a trial for all weak defendants and at least some (weaker) strong defendants.*

5 The optimal extent of discovery

To measure social welfare, we use the total expected payoffs of the defendant and plaintiff.[11] The optimal extent of discovery maximizes these expected total payoffs. Note that, due to the fully separating nature of the equilibrium given any level of discovery, the expected damage paid by a defendant, whether awarded in a trial or settlement, perfectly reflects his private information. Thus, maximizing the litigants' joint payoff does not conflict with a broader social purpose of serving justice.

Because damages paid by a defendant to the plaintiff represent a transfer, maximizing total payoffs is equivalent to minimizing the ex ante expected total litigation costs of the two litigants. Because a social planner/judge does not know the defendant's private type, she must integrate over

[11] The court (and society at large) may also incur discovery and trial costs in which case the social welfare should consider these costs as well. See Section 8.

the defendant's possible types. Before we investigate this ex ante optimal level of discovery, we first identify the optimal extent of discovery given an arbitrary defendant type θ.

Optimal π given θ. Denote the discovery level that minimizes expected total litigation costs given a type θ defendant by $\pi^*(\theta)$. Corollary 3 shows that when $\rho(\pi) = \pi$, $\pi^*(\theta)$ is strictly positive, but does not exceed $\hat{\pi}(\theta)$. But how does $\pi^*(\theta)$ vary with θ?

We now establish that in the neighborhood of the optimal level of discovery, stronger (strong) types benefit more from greater discovery. Using subscripts to denote partial derivatives, we have:

Lemma 4. *(Stronger types gain more from more discovery) When $\rho(\pi) = \pi$, $\Delta^s_{\pi,\theta}(\theta, \pi^*(\theta)) > 0$.*

Proof: See the Appendix. □

Proposition 5 then follows directly:

Proposition 5. *When $\rho(\pi) = \pi$, the optimal extent of discovery given a type $\pi^*(\theta)$ increases in θ.*

Proof: Because $\pi^*(\theta)$ is interior, we have $C^s_{\pi,\pi}(\theta, \pi^*(\theta)) > 0$. From Lemma 4, $C^s_{\pi,\theta}(\theta, \pi^*(\theta)) < 0$. From the implicit function theorem, $\pi^*_\theta(\theta) > 0$. □

Were a strong type θ's post-discovery settlement offer always accepted, then it would be best to have no discovery. Were its post-discovery settlement offer always rejected, so that trial costs $(1-\pi)c$ are incurred whenever discovery fails, then it would be best to break information acquisition into two equal steps, because $argmax\{\pi+(1-\pi)(1-\pi)\} = \frac{1}{2}$. In equilibrium, post-discovery settlement offers are sometimes, but not always, rejected. Therefore, the optimal extent of discovery is between 0 and $\frac{1}{2}$. A stronger type's settlement offer gets rejected more often. This calls for greater discovery, i.e., discovery that is closer to $\frac{1}{2}$. This effect on $\pi^*(\theta)$ dominates the countervailing effect that, with stronger types, greater discovery reduces by more the probability that its settlement offer is rejected, which calls for less discovery.

Optimal π ex ante. The optimal extent of discovery π^* minimizes ex ante expected total investigation costs (integrating over the distribution F of defendant types θ). Thus, π^* solves

$$\min_{\pi\in[0,1]} C(\pi) \equiv \int_{\underline{\theta}}^{\hat{\theta}(\pi)} C^w(\theta,\pi)dF(\theta) + \int_{\hat{\theta}(\pi)}^{\bar{\theta}} C^s(\theta,\pi)dF(\theta).$$

The following properties of the optimal extent of discovery, π^*, follow from the properties of $\pi^*(\theta)$:

Corollary 5. *When $\rho(\pi) = \pi$, the optimal extent of discovery satisfies $0 < \pi^* < \bar{\pi}$, and $\pi^* < \frac{1}{2}$.*

The policy implication is that a judge should have positive, but sharply limited, discovery. That is, absent other considerations, with no settlement, most information acquisition should occur at trial. We next derive how the distribution of defendant types affects the optimal extent of discovery. We address: when a defendant is expected to be stronger ex ante, is it better to have more discovery or less? To do this, we compare distributions of defendant types that are ordered according to conditional first-order stochastic dominance, so there is a well-defined notion of a better distribution of defendant types. Distribution $F_2 \succeq_{CFOSD} F_1$ if $F_1(\theta|\theta > \mu) \geq F_2(\theta|\theta > \mu)$ for all μ in the support of F_1, strict for all $\mu \in (\underline{\theta}, \bar{\theta})$.[12] Let x_j^* be the optimal extent of discovery given distribution F_j for $j = 1, 2$. Lemma 4 shows that increased discovery benefits stronger types by more. Therefore,

Proposition 6. *(Better defendants make more discovery optimal) When $\rho(\pi) = \pi$, and distributions F_1 and F_2 of defendant types are ordered by $F_2 \succeq_{CFOSD} F_1$, then the optimal extent of discovery is higher when defendants are more likely to be stronger, i.e., $\pi_2^* > \pi_1^*$.*

Proof: See the Appendix. □

Our characterization of the optimal extent of discovery presumes that the social planner/judge cares equally about the litigation costs incurred by each type of defendant. If, instead, the social planner weighs costs incurred by "good" (i.e., higher θ) defendants by more, then our analysis

[12] $F_2 \succeq_{CFOSD} F_1$ if and only if the hazards are ordered, $\frac{f_2(\mu)}{1-F_2(\mu)} \leq \frac{f_1(\mu)}{1-F_1(\mu)}$.

indicates that greater discovery is optimal. So, too, if the social planner is concerned about the care taken by defendants to avoid imposing harm on a party (and greater care is associated stochastically with stronger defendant types), then our analysis reveals that even greater discovery is optimal. [13]

Extending this insight that limited discovery is optimal to a broader setting, it suggests that for any disputes that involve investigations, total costs would be reduced if the parties were given a chance to settle prior to the investigation being concluded.

6 General discovery cost structures

We now consider more general cost structures, exploring how the curvature of discovery (convex or concave) affects expected litigation costs and the optimal design of discovery. The discovery cost structure may be nonlinear for many reasons. For example, for a plaintiff who does not have strong evidence that the defendant is guilty, any discovery would be a "fishing expedition". The return to searching may rise over time as earlier discovery then helps direct attention to more fruitful areas. This would deliver a concave discovery cost structure: the marginal cost of uncovering information falls as more discovery takes place. The law on the matter, starting from the 1978 decision of the Court of Appeal in Dufault v. Stevens, 6 B.C.L.R. 199 (C.A.) is that a party applying for an order for the production of documents must satisfy the court that the application is not a "fishing expedition". In the words of Madam Justice Southin: "Perhaps it is not too fanciful to say that a litigant cannot have a licence to fish in his opponent's private swimming pool unless he can provide some evidence from which it can be inferred that there may be fish in that pool. If there is no such evidence, the defendant need not let him in to see if there is a fish." This section will address whether the court should allow fishing expeditions.

[13] This analysis presumes that the distribution of defendant types does not affect the discovery cost structure. Cases with extreme distributions may lead to non-linear discovery cost structures. We show in the next section that this can create a countervailing force for the optimal level of discovery.

In practice, judges have extensive discretion over which sources of information to target in discovery. On a case by case basis, a judge usually has significant qualitative signals at his disposal about which directions for discovery are more or less promising to uncover information on a per unit cost basis, and he can make decisions on that basis. Some discovery requests—like document review—tend to be expensive and often fail to generate relevant information; but other discovery requests—like taking deposition of key personnel— can be much more informative. Concretely, one can imagine two classes of documents, Class A and Class B, where Class A is more relevant than Class B. Ordering discovery to start with Class A and leaving Class B for a potential trial would imply that $\rho(\pi)$ is convex: the marginal cost of obtaining more information rises with the amount of information. Ordering discovery in the opposite way would imply that $\rho(\pi)$ is concave. The question becomes—when taking into account the strategic behavior of the plaintiff and defendant, should the court first direct discovery toward lower or higher cost sources of information?

We first consider a convex cost structures:

Proposition 7. *(Strictly convex $\rho(\pi)$) Expected total litigation costs are strictly lower with any limited discovery π than with no discovery. For any given θ, the optimal extent of discovery, $\pi^*(\theta)$, is less than $\hat{\pi}(\theta)$. That is, for any θ, its expected litigation costs are minimized when the extent of discovery makes it want to wait until after discovery to make its first acceptable settlement offer. Stronger types benefit more from any given amount of discovery.*

Proof: See the Appendix. □

Our qualitative findings that limited discovery will benefit with linear discovery are reinforced if discovery first targets less costly sources of information. Convex costs bring two benefits to strong defendants who make their first offers post-discovery: (a) their costs incurred in discovery are lower, and (b) trial costs are raised, increasing the incentive compatible probabilities with which their post-discovery offers are accepted. Convex costs also benefit weak defendants who make their first offers pre-discovery because the efficiency of early discovery allows them to avoid trial costs without sacrificing as much of the signaling power of a pre-discovery settlement offer. Thus, relative to linear costs, convex costs reduce litigation costs of both weak and strong defendants.

If, instead, discovery first focuses on higher marginal cost sources of information, we have:

Proposition 8. *(Strictly concave $\rho(\pi)$) There exists a $\tilde{\theta}(\pi) > \hat{\theta}(\pi)$ such that all types $\theta < \tilde{\theta}(\pi)$ are worse off with discovery π than with no discovery. If any defendant types benefit from concave discovery, then it is the stronger types that benefit by more.*

No discovery may be optimal: given any strictly concave discovery cost function $\rho(\cdot)$, for any limited level of discovery $\pi \in (0,1)$, there exists an absolutely continuous density $g(\theta)$ over defendant types such that expected total litigation costs are higher with discovery than without.

Proof: See the Appendix. $\square$

Concave discovery costs have the opposite impact of convex costs, hurting all weak defendants, and at least some strong defendants—expected litigation costs are higher with discovery than without for these types. Indeed, with concave discovery costs, all defendant types may be harmed by discovery. Thus, if discovery targets high marginal cost sources of information, it may be optimal to forgo limited discovery—expected litigation costs may be lower with no discovery or full discovery.

To gain further insights, we parameterize the degree of concavity/convexity:

Proposition 9. *(Convexity of costs) If $\rho(\pi) = \pi^z$ with $z > 0$, then increases in the convexity parameter, z, reduce expected total litigation costs.*

Proof: See the Appendix. $\square$

Figure 3 illustrates how expected litigation costs vary with the extent of discovery π and the convexity z of the discovery cost function. It reveals that the more concave (less convex) are discovery costs, the more likely defendants are to settle: shifting information acquisition costs forward by focusing on expensive sources of information that are unlikely to bear fruit induces *more* defendant types to make pre-discovery settlement offers and *raises* the probability that those offers are accepted. One might therefore think that a judge should grant such discovery requests.

In fact, our analysis indicates that the *opposite* is optimal. The results reveal that a judge should direct discovery toward cost-effective sources and discourage "fishing expeditions", i.e., discovery sought on suspicion, surmise or vague guesses, even if such discovery may create leads for more effective sources of information.[14] When the cost structure seems concave, for example, with a security litigation case alleging a company concealed information without any basis, a judge should simply deny discovery. This is exactly what the Private Securities Litigation Reform Act (PSLRA), passed in 1995, tries to achieve. Prior to the PSLRA, a plaintiff could start a case with minimal evidence and use pre-trial discovery to search for more. After PSLRA, plaintiffs must present evidence of fraud before any pretrial discovery takes place.[15]

We pose our model in the context of discovery. However, as our introduction suggests, one can more generally interpret discovery as an activity that breaks information gathering into stages, where the relevant private information can be uncovered at any stage. Our analysis then suggests that earlier stages should have a selective focus on the evidence considered; and that the proper design can reduce information acquisiton expenses. Many of our insights extend to characterize the evolution of trials, where information gradually becomes public as a trial takes its course, and there are multiple opportunities to settle. However, trials give judges less ability to control which sources of information get revealed first than does discovery. Importantly, our analysis indicates that the ability of a judge to determine the sequencing of which sources of information are considered first is a crucial determinent of its welfare properties.

[14] The definition of "fishing expedition" is taken from West's Encyclopedia of American Law, edition 2.

[15] See "Madoff Victims Face Grim Prospects in Court", Jane Bryant Quinn, BLOOMBERG.COM.

7 A related problem of acquiring information in steps

A judge seeking to minimize litigation costs must account for the fact that litigants are *strategic* when setting discovery: defendants internalize the level of discovery when deciding whether to make settlement offers, and the level of discovery affects the probability a plaintiff accepts an offer. We just saw that whether *any* discovery was optimal depended on the curvature of discovery costs.

One can glean insights from contrasting this problem with that of a single *non-strategic* agent who can divide information acquisition into two stages, stopping after the first stage if he succeeds, and only continuing if he fails. The agent chooses a probability y of learning the state at stage 1 at an information cost of $\rho(y)$, where $\rho(\cdot)$ has the same structure as the discovery cost function. If he fails, which happens with probability $1-y$, he proceeds to stage 2, where he incurs additional information costs of $1-\rho(y)$ to learn the state. Thus, stage 1 search corresponds to discovery, and stage 2 corresponds to a trial. The agent chooses y to minimize expected search costs $K(y) \equiv \rho(y)+(1-y)(1-\rho(y)) = 1-y+y\rho(y)$. Let $\Delta(y) = K(0)-K(y)$ denote the benefit of breaking up the information acquisition, and let y^* denote the optimal level of y.

Lemma 5. *(Nonstrategic search benchmark). A single searcher always benefits from breaking up information acquisition into steps:* $\Delta(y) > 0$ *for any* $y \in (0,1)$ *and any* $\rho(\cdot)$. *If* $\rho(y) = y$, $y^* = \frac{1}{2}$.

Proof: See the Appendix. $\square$.

Regardless of the curvature of search costs, a decision maker *always* benefits from breaking up information acquisition into steps: if the first step succeeds, it saves future effort. Thus, there is a stark difference between this outcome and that with concave discovery costs, where positive discovery can raise litigation expenses for *every* defendant type. The difference reflects that given discovery π, the probability that a plaintiff accepts a defendant's offer must make it incentive compatible for the defendant to make an offer that reveals his type. Relative to linear or convex costs, concave costs raise the cost of a given level of discovery and reduce the cost of trial. Thus, weak defendants pay more discovery costs when their pre-discovery offers are rejected and save less on trial costs. Even though pre-discovery offers are less likely to be rejected than with linear or convex costs, the net effect of discovery with concave costs is to hurt weak defendants. Concave costs hurt strong defendants who make their first offers post-discovery in two ways: (a) costs incurred in discovery are higher and (b) lower trial costs raise the likelihoods with which their post-discovery offers must be rejected to preserve incentive compatibility. Thus, concave costs can increase the litigation costs of both weak and strong defendants to the point that no discovery can become optimal.

Our model differs from this non-strategic benchmark in other important ways. First, the remaining trial costs are unaffected by whether discovery succeeds or fails (an assumption that we revisit). Instead, successful discovery provides incentives to settle immediately, in which case the litigants avoid a trial and its costs. Second, litigants may *not* incur discovery costs; but a searcher necessarily incurs some search costs. This difference reflects that the litigants may settle pre-discovery. Third, even after a failed discovery, litigants may avoid trial costs by settling. For these reasons, even though in the benchmark, interior levels of search $y \in (0,1)$ always yield positive benefits $\Delta(y) > 0$, extensive discovery $\pi > \hat{\pi}(\theta)$ provides no benefits when $\rho(\pi) = \pi$.

A non-strategic searcher minimizes $\rho(y)+(1-y)(1-\rho(y))$. A social planner/judge minimizes $\int_{\theta<\hat{\theta}} \rho(\pi) r_1(\theta,\pi) F(d\theta) + \int_{\theta>\hat{\theta}} \rho(\pi)+(1-\pi) r_2(\theta,\pi)(1-\rho(\pi)) F(d\theta)$. Even in a hypothetical setting where the judge "knows" θ, but must uncover it, he will choose the limited discovery ($\pi < \hat{\pi}(\theta)$), that minimizes $\rho(\pi)+(1-\pi)(1-\rho(\pi)) r_2(\theta,\pi)$. There is some chance that remaining trial costs can be avoided, which provides an incentive to reduce discovery relative to the search benchmark. Reduced discovery makes a trial more costly, which raises the probability a post-discovery settlement offer is accepted, thereby avoiding the trial costs. However, greater discovery raises the critical cutoff type, reducing the separation distance in the signaling game, which increases the probability that a pre-trial settlement offer is accepted. Even in this hypothetical situation with fixed θ, the net effect is that it is optimal to have less discovery than in the two-step search benchmark.

Proposition 10. *(Optimal discovery vs. optimal search) With linear information costs $\rho(\pi) = \pi$, for any distribution over θ, it is always optimal to have less discovery than search: $\pi^* < y^*$.*

Proof: See the Appendix. $\square$.

More generally, regardless of the curvature of information acquisition costs, a single searcher always wants to break information acquisition into steps (Lemma 5). In contrast, as Section 6 shows, with concave discovery costs, the optimal level of discovery can be 0 or 1. Thus,

Corollary 6. *(Benefits of discovery vs. search) Expected information costs of a single agent searcher are always lower for $y \in (0, 1)$ than for $y = 0$. In contrast, if discovery costs $\rho(\pi)$ are strictly concave then for any level of positive discovery $\pi \in (0, \bar{\pi})$, there exist uniformly continuous distributions of defendant types for which expected litigation costs are higher with discovery than without.*

Even though discovery provides benefits vis à vis the no discovery benchmark, Proposition 10 and Corollary 6 can be interpreted as being anti-discovery vis à vis the benchmark of a single agent's non-strategic search. The strategic and signaling considerations imply that earlier information revelation should be suppressed relative to what is optimal for a single information acquirer.

8 Additional real world features

In this section, we describe how integrating additional real world features qualitatively affects results. The detailed formal analysis is available online.[16]

Reduced trial costs due to successful discovery. We have supposed that the success or failure of discovery does not affect the level of costs incurred at trial. This is consistent with a scenario

[16]http://www2.warwick.ac.uk/fac/soc/economics/staff/mdbernhardt/online_appendix.pdf

in which the total information to be uncovered in the legal process—both the defendant's private information and the information that neither party knows—does not depend on the timing of when each type of information is uncovered. However, it may be that successful discovery that uncovers a defendant's private information *reduces* the information costs that must be incurred in a trial.

To highlight the qualitative consequences, suppose that successful discovery eliminates *all* trial costs. In our base model, there is never a trial following successful discovery, so one might conjecture that if successful discovery eliminates trial costs, this just results in a transfer from the defendant to a plaintiff via a higher settlement, and does not affect litigation costs. This reasoning about the transfer is correct, but incomplete. The reduction in trial costs due to successful discovery reduces the surplus a defendant can extract. This raises the cost of mimicking a better type's pre-discovery offer because such mimicking raises the chance that the offer is rejected, raising the chance of successful discovery. This makes pre-discovery separation of types easier—the rejection probability required to induce incentive compatible pre-discovery separation is reduced. Moreover, because more defendant types make pre-discovery offers, it eases the separation of stronger types. As a result, positive, but limited, discovery reduces expected total legal costs incurred with any defendant type θ relative to when successful discovery does not affect trial costs.

This does not imply that defendants benefit when successful discovery reduces trial costs. When successful discovery reduces trial costs, plaintiffs gain from a defendant's reduced ability to extract lower settlements via a threat to go to trial. Even though total surplus is raised when successful discovery eliminates trial costs, defendants may be hurt. In particular, with linear discovery costs, only the stronger of strong defendant types ever benefit from the improved separation that discovery facilitates, and it can be that all defendant types would prefer no discovery to a discovery that removes the threat of trial costs used to extract surplus from plaintiffs.

Indeed, our online appendix shows that because plaintiffs now gain from successful discovery, a plaintiff prefers to have more discovery than is socially optimal—if a plaintiff could, she would request socially excessive levels of discovery. This means that a judge may serve an important role in filtering a plaintiff's discovery requests, and possibly denying some of them.

Fixed procedural trial costs. In our base model, trial costs only reflect information discovery. In practice, trials have fixed procedural costs that are unrelated to the presentation of evidence. Our online appendix shows that such costs serve only to raise the optimal extent of discovery: positive fixed trial costs favor greater discovery because it encourages defendants to settle prior to trial.

Court incurs costs at trial. The court (and society at large) also incur costs with a trial that are not internalized by the litigants. These include the opportunity costs of a jury's time, and of the judge's and clerks' time that can be allocated to other cases. The socially optimal level of discovery should reflect these costs. Our online appendix shows that the optimal level of discovery is increased as a result. The intuition is simple: discovery reduces the probability of a trial and the likelihood that these court costs are incurred. Still, we caution that sometimes discovery itself can be costly for the court, for example in the form of a Summary Jury Trial or Court-Annexed Arbitration that we discuss in the introduction. Such forms of "discovery" are essentially abbreviated trials. If the share of costs incurred by the court and the litigants is the same for a summary jury trial and a full trial, then the optimal extent of discovery is the same as when the court incurs no costs.

Multiple rounds of discovery. Our base model features one round of discovery. In practice, discovery itself is a process, and litigants have multiple chances within discovery to settle. Our online appendix shows that the cost savings due to discovery are enhanced when the discovery process is further divided into more rounds. Such division better facilitates separation via the timing of settlement offers, reducing the inefficiencies associated with the higher rates with which settlement offers must be rejected in order to induce incentive compatible revelation of a defendant's type.

How dividing discovery into more rounds affects equilibrium outcomes is somewhat subtle: with more rounds of discovery, *more* defendant types are prepared to wait until after the end of *all* discovery to make their first offers (unless discovery succeeds). Intuitively, *conditional* on one round of discovery failing, the next round becomes more cost effective, revealing θ with a higher *conditional* probability per unit cost of discovery, and this less costly signal makes separation harder, reducing the number of defendant types that make pre-trial settlement offers (when discovery fails). Nonetheless, expected total litigation costs are reduced when discovery is divided into more rounds for all strong defendant types, whereas those of weak types who make offers immediately are unchanged, as their incentives are unaffected by what stronger types later do.

The intuition for why multiple rounds of discovery reduce expected investigation costs with stronger defendants, is similar to that for why one round of partial discovery is better than none. Discovery costs are linear in π, whereas endogenous separation via settlement offers requires rejection probabilities to rise concavely with θ, i.e., faster than linearly in order to induce weaker types not to mimic stronger types. The inefficiency in endogenous separation via higher rejection probabilities shrinks in the "distance" from the weakest type separating in a given round of settlement offering, making division into more rounds of discovery optimal.

Allowing for pre-discovery settlement. One can interpret discovery more generally as a stage of an investigation of a dispute, followed by an opportunity to settle, with the investigation continuing absent a settlement, where there is uncertainty about who is at fault, and one side has private information that the investigation may uncover. Our analysis indicates that the parties should have a chance to settle both prior to the investigation, as well as at stages during the investigation.

For example, a tax audit is an investigation into a dispute between a tax payer and the IRS. Unlike in our discovery setting, however, a "defendant" (the tax payer) cannot "settle" prior to an audit. The moment that a tax payer learns of a dispute is when he or she receives notice of an audit. The tax payer can try to settle after the audit, but absent a settlement, he or she goes to the tax court with the IRS. One can place a tax audit in our setting, by assuming that a defendant cannot make offers prior to discovery, but can make offers after some discovery occurs. Total litigation costs in this setting exceed those in our benchmark model for any positive level of discovery for all defendant types. To see this, observe that types that would have settled post-discovery must now separate from worse types than before because no type can signal via a pre-discovery offer. This raises the probability of rejection for post-discovery settlement offers needed to induce truthful revelation, raising litigation costs for these types. In addition, weaker types that would have settled pre-discovery, if given the chance, now incur higher costs to settle post-discovery.

This logic suggests that the IRS should give a tax payer a chance to settle before starting an audit. This insight applies to other disputes outside the court, such as those within an organization.

9 Plaintiff makes settlement offers

Our analysis has focused on the impact of discovery in settings where an informed defendant signals his type by choosing the timing and size of settlement offers to make to an uninformed plaintiff. The design of discovery also matters in screening settings, where an uninformed plaintiff tries to separate defendant types by making take-it-or-leave-it settlement offers pre- and post-discovery. We now characterize the optimal design of discovery in such screening settings.

To ease presentation, we assume that (a) discovery costs are linear, $\rho(\pi) = \pi$; and (b) a defendant's type is drawn from a uniform distribution on $[\underline{\theta}, \bar{\theta}]$ with $c < \overline{\theta} - \underline{\theta}$. Our model builds on Bebchuk (1984). For $\pi \in (0, 1)$, a plaintiff makes screening offers both pre- and post-discovery:

Proposition 11. *(Screening with discovery) Suppose $\rho(\pi) = \pi$ and $\theta \sim U[\underline{\theta}, \bar{\theta}]$. A plaintiff's pre-discovery offer is $x_1^* = 1 - \underline{\theta} - \pi c + c_d - (1-\pi)^2 c$. Defendant types $\theta \in [\underline{\theta}, \underline{\theta} + \pi c]$ accept the offer. Stronger types reject it and go on to discovery. If discovery succeeds, the parties settle post-discovery at the full information settlement offer. If discovery fails, the plaintiff's post-discovery offer is $x_2^* = 1 - \underline{\theta} - \pi c - (1-\pi)c_p$. Defendant types $\theta \in [\underline{\theta} + \pi c, \underline{\theta} + c]$ accept the offer, whereas stronger defendant types reject it and go to trial.*

Proof: See the Appendix. $\square$

The solution only relies on the uniform distribution of weaker types, $\theta \in [\underline{\theta}, \underline{\theta} + c]$ (assuming that the first-order conditions continue to describe the solution, for which log-concavity of F suffices). In the special case of no discovery, i.e., $\pi = 0$, the plaintiff's settlement offer reduces to $x^* = 1 - \underline{\theta} - c_p$. With no discovery, all defendant types $\theta < \theta^* \equiv \underline{\theta} + c$ accept the offer and settle; and all $\theta > \theta^*$ reject the offer and go to trial. Total expected litigation costs in this case are $\frac{\bar{\theta}-\underline{\theta}-c}{\bar{\theta}-\underline{\theta}}c$. Discovery has two effects on expected litigation costs relative to no discovery:

1. Discovery sometimes succeeds for strong types $\theta \in [\underline{\theta} + c, \bar{\theta}]$ that would go to trial absent discovery, who comprise fraction $\frac{\bar{\theta}-(\underline{\theta}+c)}{\bar{\theta}-\underline{\theta}}$ of defendants. Discovery succeeds with probability π, and when it does, a plaintiff's post-discovery settlement offer is accepted, saving trial costs of $(1 - \pi)c$. Hence, the expected savings are

$$\frac{\bar{\theta} - (\underline{\theta} + c)}{\bar{\theta} - \underline{\theta}}\pi(1 - \pi)c = \frac{\bar{\theta} - \underline{\theta} - c}{\bar{\theta} - \underline{\theta}}\pi(1 - \pi)c.$$

2. Discovery causes some defendant types to strategically delay settlement until after discovery. Intermediate types $\theta \in [\underline{\theta} + \pi c, \underline{\theta} + c]$ would settle immediately absent discovery, but with discovery they reject the initial offer, incur discovery costs πc and then settle. The probability of these types is $\frac{(\underline{\theta}+c)-(\underline{\theta}+\pi c)}{\bar{\theta}-\underline{\theta}} = \frac{(1-\pi)c}{\bar{\theta}-\underline{\theta}}$, so discovery raises costs for these types by

$$\frac{(1 - \pi)c}{\bar{\theta} - \underline{\theta}}\pi c = \frac{c}{\bar{\theta} - \underline{\theta}}\pi(1 - \pi)c.$$

Subtracting yields the expected net gain or loss from discovery π vis à vis no discovery:

$$\frac{\bar{\theta} - \underline{\theta} - c}{\bar{\theta} - \underline{\theta}}\pi(1 - \pi)c - \frac{c}{\bar{\theta} - \underline{\theta}}\pi(1 - \pi)c = \frac{\bar{\theta} - \underline{\theta} - 2c}{\bar{\theta} - \underline{\theta}}\pi(1 - \pi)c.$$

Thus, the net gain or cost of discovery vis à vis no (or full) discovery is proportional to $\pi(1 - \pi)c$. When $\bar{\theta}-\underline{\theta}-2c > 0$, expected litigation costs are minimized by $\pi = \frac{1}{2}$, which maximizes $\pi(1-\pi)$; and when $\bar{\theta} - \underline{\theta} - 2c < 0$, no discovery is optimal, as it minimizes $\pi(1 - \pi)$. To summarize:

Proposition 12. *(Value of discovery with screening) When $\rho(\pi) = \pi$ and $\theta \sim U[\underline{\theta}, \bar{\theta}]$, discovery reduces total expected litigation costs if and only if $\bar{\theta} - \underline{\theta} - 2c > 0$. In that case, the optimal extent of discovery is $\pi^* = \frac{1}{2}$.*

The condition $\bar{\theta} - \underline{\theta} - 2c > 0$ is equivalent to $(1 - \underline{\theta}) - \frac{\bar{\theta}+\underline{\theta}}{2} > c$. This says that the difference between the liability of the worst type and the average type exceeds the cost of information acquisition about types. The happens when the cost is low enough relative to the uncertainty about types, which is natural for cases where the defendant has significant private information. This condition may fail for cases where most private information was resolved before the onset of the settlement negotiation, for example in previous cases involving the same defendant.

Corollary 7. *When $\rho(\pi) = \pi$ and $\theta \sim U[\underline{\theta}, \bar{\theta}]$, stronger defendant types are less harmed by discovery: strong defendants $\theta > \underline{\theta} + c$ are unaffected, types $\theta \in [\underline{\theta} + \pi c, \underline{\theta} + c]$ expect to pay an additional $\pi(\underline{\theta} + c - \theta)$ relative to no discovery and weak defendant types $\theta \in [\underline{\theta}, \underline{\theta} + \pi c]$ expect to pay an additional $(1 - \pi)\pi c$. Plaintiffs prefer greater discovery than is socially efficient.*

Proof: See the Appendix. $\square$

Very strong defendant types are indifferent to the extent of discovery: they will go to trial without discovery and will either go to trial or settle under full information with discovery with the plaintiff extracting all surplus. Again, if successful discovery reduces trial costs, they would gain at the plaintiff's expense.

For intermediate defendant types, discovery means that they face less harsh settlement demands when discovery fails because the remaining trial costs are lower; but they also face harsher settlement demands when discovery succeeds because their private information is eliminated. Taking into account the extra cost of discovery, in expectation, their payoffs fall. Weaker defendants are also hurt by discovery: the prospect of losing their private information when discovery succeeds makes them willing to accept a lower settlement offer. Thus, a defendant always loses from discovery. It follows that when discovery reduces total litigation costs, the plaintiff gains by more than the reduction in litigation costs. In fact, the plaintiff always strictly prefers more discovery than is socially optimal. This implies that the court should grant less discovery than a plaintiff requests.

To show how changes in the prior distribution of defendant types affect the merits of discovery, we consider a simple class of distributions: with small probability q, the defendant is of type $\overline{\theta}$ and with probability $1-q$, the type $\sim U[\underline{\theta}, \bar{\theta}]$. Proposition 11 extends due to the uniform bottom tail of the distribution. The total benefit of discovery for strong defendants $\theta \in [\underline{\theta}+c, \overline{\theta}]$ becomes

$$\left[(1-q)\frac{\bar{\theta}-\underline{\theta}-c}{\bar{\theta}-\underline{\theta}}+q\right]\pi(1-\pi)c.$$

The total loss from discovery for intermediate defendants $\theta \in [\underline{\theta}+\pi c, \underline{\theta}+c]$ is

$$(1-q)\frac{c}{\bar{\theta}-\underline{\theta}}\pi(1-\pi)c.$$

Therefore, a stronger pool of defendants, i.e., a higher q, raises the social value of discovery.

If the plaintiff is not allowed to make a settlement offer prior to the discovery at level $\pi > 0$, then a segment of types, $[\underline{\theta}, \underline{\theta}+\pi c]$, would have to delay their settlement after the discovery, causing additional total litigation costs of πc with probability $\frac{\pi c}{\bar{\theta}-\underline{\theta}}$. This shows that the parties should be allowed to settlement prior to the discovery.

We have assumed that the sole criterion for optimal discovery is to minimize expected litigation costs. In fact, accuracy also matters—the court wants an award to reflect a defendant's true liability. With signaling, a defendant's settlement offer always reflects its true liability damage. However, with screening, offers only perfectly reflect a defendant's private information if discovery succeeds; and settlements before the discovery or after a failed discovery lump defendant types together.

Comparing screening and signaling. The settings share key features. In both settings, limited discovery is always optimal, and when realistic features are introduced, the uninformed plaintiff wants more discovery than is socially optimal. The consequences for settlement are also similar: weak defendants settle prior to discovery, stronger types settle after discovery, and the strongest ones go to trial, unless discovery succeeds, in which case they settle. In both settings, discovery reduces the probabilty of trial and raises settlement rates. Finally, the distributional effects on defendant payoffs are similar: stronger defendants gain more from discovery, or are hurt less.

There are also important differences. First, in screening settings, with limited uncertainty about defendant types relative to the potential costs of litigation, discovery is not socially optimal even when discovery costs rise one-for-one with the probability of uncovering private information. This reflects that in screening settings, defendant types that settle after incurring discovery costs are not more likely to settle relative to no discovery, so the social benefits of discovery come solely from strong defendant types that would go to trial absent discovery, but settle when discovery succeeds. In contrast, in signaling settings, defendant types that settle after discovery do not need to separate away from weaker types that settled pre-discovery, so the social benefits of discovery come from more than just breaking the information acquisition into steps. Second, in both settings, properly-designed discovery helps the party that makes offers when successful discovery does not reduce trial costs; and it can help the party receiving offers when successful discovery reduces trial costs. However, who makes offers differs.[17] Third, in a signaling setting, the settlement offers of all types reflect their true private information either via an offer that reveals the private information or through discovery or trial that directly reveals the information; but in a screening setting, there is a "loss of accuracy" because the uninformed's offers lump (similar) types together. However, even then, there is improved accuracy when discovery succeeds; and when it fails, on average, offers are better targetted at a defendant's true type because of the increased separation.

Collective policy implications include that discovery should be sharply limited, a judge should limit requests by uninformed parties for discovery (because they may have incentives to seek ex-

[17] That discovery can hurt a defendant in screening settings is also seen in Sobel (1989) and Schwartz and Wickelgren (2009).

cessive discovery), discovery should be directed toward low marginal cost sources of information and parties should be allowed to settle prior to discovery.

10 Conclusion

In this article, we characterize how the process of publicly-gathering information via discovery that may reveal a defendant's private information affects strategic interactions in litigation between a plaintiff and a defendant. We endogenize the timing and size of settlement offers made by a privately-informed defendant throughout the litigation process, and the equilibrium probabilities with which a plaintiff accepts these offers.

We show how the discovery process provides defendants an additional channel with which to signal—the timing of their initial settlement offers. With limited discovery, weaker defendants make offers prior to discovery, but stronger defendants wait to make offers until after discovery. Limited discovery facilitates separation of defendant types, reducing the inefficiently high rates with which plaintiffs must reject settlement offers in order to induce truthful revelation of a defendant's private information. As a result, the privately-informed defendants gain from properly-designed *limited* discovery, with stronger defendants gaining more.

We derive how the optimal extent of discovery hinges on the distribution of defendant types—more discovery is optimal when the privately-informed defendants tend to have stronger cases *and* when discovery can be directed to specific fruitful areas of search.

In practice, considerations other than the ones we analyze may also affect the optimal extent of discovery. For example, discovery may create new costs, rather than simply moving information acquisition costs earlier from the trial. For example, shareholder litigation against a merger or an acquisition may delay the merger or acquisition. More discovery will increase these delay costs for the defendant. This would call for less discovery.

Our analysis starts after a defendant has disclosed all verifiable good news. A defendant wants to reveal any verifiable good news at the outset. Thus, one can view the distribution over the defendant's type in our model as being given by the posterior distribution after the defendant has chosen the verifiable information to disclose at the outset. However, some private information remains because some information (e.g., the absence of bad news) cannot be credibly revealed absent a thorough discovery.

Our analysis implicitly assumes that the court commits to the legal process and its design of discovery. In particular, the court does not impose settlements based on any information revealed by the settlement offers themselves: settlement offers must sometimes be rejected and the threats of discovery and trial must be real to preserve a defendant's incentive to reveal information in settlement offers. In practice, the court often does not observe settlement offers. More philosophically, the court has dual roles of (1) structuring incentives for litigants to settle; and, (2) should the parties fail to settle, to deliver "justice" by uncovering the relevant information, both private information, and information that neither party initially has. This latter consideration precludes using the information in settlement terms to shortcut the legal process.

Our model of discovery can be interpreted more broadly. We decompose the uncovering of a defendant's information into stages of costly information acquisition. The stages are distinguished by the abilities of litigants to make and accept settlement offers prior to each stage, where litigants incur costs if they fail to settle prior to a stage. Our analysis describes other settings with this structure, including summary jury trials and court-annexed arbitration. One can even view a trial itself as a process similar to the discovery-and-then-trial process that we model in that private information becomes public as a trial takes its course, and there are multiple opportunities to settle during a trial. However, there is a key distinction: in a trial, a judge has less ability to control which sources of information get revealed first. Importantly, our analysis shows that a judge's ability to determine the sequencing of different information sources is crucial to its welfare properties.

References

Banks, J. and J. Sobel. “Equilibrium Selection in Signaling Games.” Econometrica, Vol. 55(1987), pp. 647-661.

Bebchuk, L. A. “Litigation and Settlement under Imperfect Information.” Rand Journal of Economics, Vol. 15(1984), pp. 404-415.

Border, K. C., and Sobel, J. “Samurai Accountant: A Theory of Auditing and Plunder.” Review of Economic Studies, Vol. 54(1987), pp. 525-540.

Cooter, R. and D. Rubinfeld. “An Economic Model of Legal Discovery.” Journal of Legal Studies, Vol. 23(1994), pp. 435-463.

Doornik, K. “A Rationale for Mediation and its Optimal Use.” International Review of Law and Economics, Vol.38(2014) pp. 1-10.

Farmer, A. and P. Pecorino. “Civil Litigation with Mandatory Discovery and Voluntary Transmission of Private Information.” Journal of Legal Studies, Vol. 35(2005) pp. 137-159.

Glaser, W. A. “Pretrial Discovery and the Adversary System.” Russell Sage Foundation, 1968.

Hay, B. “Civil Discovery: Its Effects and Optimal Scope.” Journal of Legal Studies, Vol. 23(1994) pp. 481-415.

Hannaford-Agor, P. “Short, Summary & Expedited: The Evolution of Civil Jury Trials.” National Center for State Courts. 2012.

Lambros, T. D. “Summary Jury Trial - An Alternative Method of Resolving Disputes.” Judicature, Vol. 69(1985), pp. 286-290.

Lambros, T. D. “Summary Jury Trial: An Effective Aid to Settlement.” Judicature, Vol. 77(1993), pp. 6-8.

Landes, W. M. “Sequential versus Unitary Trials: An Economic Analysis.” The Journal of Legal Studies, Vol. 22(1993), pp. 99-134.

Lind, E. A., and J. Shapard. “Evaluation of Court-Annexed Arbitration in Three Federal District Courts.” Federal Judicial Center. 1983.

Mookherjee, D., and I. Png. "Optimal Auditing, Insurance, and Redistribution." Quarterly Journal of Economics, Vol. 104 (1989), pp. 399-415.

Posner, R. A. "The Summary Jury Trial and other Methods of Alternative Dispute Resolution: Some Cautionary Observations." The University of Chicago Law Review, Vol. 53(1986), pp. 366-393.

Reinganum, J. and L. Wilde. "Income Tax Compliance in a Principal-Agent Framework." Journal of Public Economics, Vol. 26(1985), pp. 1-18.

Reinganum, J. and L. Wilde. "Settlement, Litigation, and the Allocation of Litigation Costs." RAND Journal of Economics, Vol. 17(1986), pp. 557-566.

Rosenberg, M. "Federal Rules of Civil Procedure in Action: Assessing Their Impact." University of Pennsylvania Law Review, Vol. 137(1989) pp. 2197-2213.

Schrag, J. "Managerial Judges: An Economic Analysis of the Judicial Management of Legal Discovery." RAND Journal of Economics, Vol. 30(1999) pp. 305-323.

Schwartz, W. and A. Wickelgren. "Credible Discovery, Settlement, and Negative Expected Value Suits." RAND Journal of Economics, Vol. 40(2009) pp. 636-657.

Shavell, Steven. "Sharing of Information Prior to Settlement or Litigation." RAND Journal of Economics, Vol. 20(1989), pp. 183-195.

Sobel, J. "An Analysis of Discovery Rules." Law and Contemporary Problems, Vol. 52(1989) pp. 133-159.

Webber, C. F. "Mandatory Summary Jury Trial: Playing by the Rules?" The University of Chicago Law Review, Vol. 56(1989), pp. 1495-1522.

White, M. J. "Asbestos Litigation: Procedural Innovations and Forum Shopping." Journal of Legal Studies, Vol. 35(2006), pp. 365-398.

11 Appendix

Proof of Lemma 1: Strict convexity implies $(1-\pi)\rho(0)+\pi\rho(1)>\rho(\pi)$ for any $\pi\in(0,1)$. That is, $\pi>\rho(\pi)$. Strict convexity and differentiability of ρ imply that $\rho''(\pi)>0$ for $\pi\in(0,1)$, so, $\pi\rho''(\pi)+\rho'(\pi)>\rho'(\pi)$, i.e., the derivative of $\rho'(\pi)\pi$ exceeds that of $\rho(\pi)$. Further, because $\rho'(0)0=\rho(0)$, we have $\rho'(\pi)\pi>\rho(\pi)$. Therefore, $\left(\frac{\rho(\pi)}{\pi}\right)'=\frac{\rho'(\pi)\pi-\rho(\pi)}{\pi^2}>0$ for $\pi\in(0,1)$. The case for strictly concave $\rho(\pi)$ follows with the inequalities reversed. $\square$

Proof of Lemma 2: Let $(\exp)$ denote $\exp\left(-\frac{\theta-\underline{\theta}}{\frac{\rho(\pi)}{\pi}c}\right)$. The derivative of r_1 with respect to π is:

$$\frac{\partial}{\partial\pi}r_1(\theta,\pi)=-\frac{1}{\pi^2}(1-(\exp))-\frac{1}{\pi}(\exp)\frac{\theta-\underline{\theta}}{\frac{\rho}{\pi}c}\frac{(\frac{\rho}{\pi})'}{\frac{\rho}{\pi}}.$$

Case 1. strictly convex or linear costs. Then $(\frac{\rho}{\pi})'\geq 0$ directly implies $\frac{\partial}{\partial\pi}r_1(\theta,\pi)<0$. We are done.

Case 2. strictly concave costs. Then $(\frac{\rho}{\pi})'<0$. First, we claim $-(1-(\exp))-(\exp)\left(-\frac{\theta-\underline{\theta}}{\frac{\rho(\pi)}{\pi}c}\right)>0$ for any $\theta>\underline{\theta}$. To see that, define within this proof $f(x)\equiv-(1-\exp(x))-\exp(x)x$. Then $f'(x)=-\exp(x)x<0$, which means $f\left(-\frac{\theta-\underline{\theta}}{\frac{\rho(\pi)}{\pi}c}\right)>f(0)=0$. Second, we have,

$$-\frac{1}{\pi^2}-\frac{1}{\pi}\frac{(\frac{\rho}{\pi})'}{\frac{\rho}{\pi}}=-\frac{1}{\pi^2}\frac{\rho'}{\rho}\pi<0.$$

That is, $\frac{1}{\pi^2}>-\frac{1}{\pi}\frac{(\frac{\rho}{\pi})'}{\frac{\rho}{\pi}}$. Then, for any $\theta>\underline{\theta}$, because $1-(\exp)>0$, we have,

$$\begin{aligned}\frac{\partial}{\partial\pi}r_1(\theta,\pi)&=\frac{1}{\pi^2}(-(1-(\exp)))-\left(-\frac{1}{\pi}\frac{(\frac{\rho}{\pi})'}{\frac{\rho}{\pi}}\right)(\exp)\left(-\frac{\theta-\underline{\theta}}{\frac{\rho}{\pi}c}\right)\\&<-\frac{1}{\pi}\frac{(\frac{\rho}{\pi})'}{\frac{\rho}{\pi}}\left[-(1-(\exp))-(\exp)\left(-\frac{\theta-\underline{\theta}}{\frac{\rho}{\pi}c}\right)\right]<0.\quad\square\end{aligned}$$

Proof of Proposition 1. *Existence:* To complete the description of the equilibrium we specify off-equilibrium beliefs:

$$b_1^*(x_1) = 1 - x_1 - c_p, \text{ if } x_1 \in [\underline{x}_1, \bar{x}_1]; \qquad b_1^*(N) = F(\theta|\theta > \hat{\theta}),$$
$$b_2^*(N, \emptyset, x_2) = 1 - x_2 - (1-\rho)c_p, \text{ if } x_2 \in [\underline{x}_2, \bar{x}_2]; \qquad b_2^*(x_1, \emptyset, x_2) = b_1^*(x_1), \text{ if } x_1 \neq N.$$

(i) Plaintiff's perspective. For any $x_1 \in [\underline{x}_1, \bar{x}_1]$, given the belief, the plaintiff will be offered $x_1 - \rho(\pi)c_p$ at $t = 2$. The plaintiff is indifferent between accepting and rejecting the offer, so $p_1^*(x_1)$ is optimal for the plaintiff. For $x_1 > \bar{x}_1$, $b_1^*(x_1) = \underline{\theta}$. Because the plaintiff is indifferent between accepting and rejecting if the belief is $\underline{\theta}$ and the offer is $\bar{x}_1$, she must strictly prefer accepting an offer $x_1 > \bar{x}_1$, so $p_1^*(x_1) = 0$ is optimal. For $x_1 < \underline{x}_1$, $b_1^*(x_1) = \hat{\theta}$. Because the plaintiff is indifferent between accepting and rejecting if the belief is $\hat{\theta}$ and the offer is $\underline{x}_1$, she must strictly prefer rejecting offers $x_1 < \underline{x}_1$, so $p_1^*(x_1) = 1$ is optimal. This shows the optimality of p_1^*. Similarly, p_2^* is optimal.

(ii) Defendant's perspective. First consider $\theta \leq \hat{\theta}$. Offer $x_1 > \bar{x}_1$ is dominated by offer $\bar{x}_1$, because both are accepted with probability one. For $x_1 \in [\underline{x}_1, \bar{x}_1]$, $p_1^*(x_1)$ is twice differentiable. Its construction gives the first-order condition for maximizing type θ's payoff among offers $x_1 \in [\underline{x}_1, \bar{x}_1]$. The second-order condition is

$$p_1''(x_1)[\pi(-(1-\theta) + c_p) + \pi x_1 - \rho(c_p + c_d)] + p_1'(x_1)2\pi.$$

Because

$$p_1''(x_1) = -\frac{1}{\pi}\exp\Big(\frac{x_1 - (1 - \underline{\theta} - c_p)}{\frac{\rho}{\pi}c}\Big)\frac{1}{c^2} = \frac{p_1'(x_1)}{c},$$

the second-order derivative evaluated at $x_1^*(\theta) = (1-\theta) - c_p$ is $p_1'(x_1)[2\pi - \rho(\pi)] < 0$ as $\rho(\pi) \leq \pi$ and $p_1'(x_1) < 0$. That is, $x_1^*(\theta)$ is a local strict maximizer. Because it is the only one satisfying the first-order condition over $[\underline{x}_1, \bar{x}_1]$ for type θ, it must be the maximizer over $[\underline{x}_1, \bar{x}_1]$. Offers $x_1 < \underline{x}_1$ are always rejected and thus are dominated by offering $\underline{x}_1$. If type θ does not make an offer, then at $t = 2$, offering $x_2 > \bar{x}_2$ is dominated by offering $\bar{x}_2$. The probability p_2^* is constructed so that for type θ, the payoff rises in x_2. Thus, offering $\bar{x}_2$ is optimal should $\theta < \hat{\theta}$ not make an offer at $t = 1$. This gives a payoff that is the same as making offer $\underline{x}_1$ at $t = 1$. Thus, not making an offer at $t = 1$ is suboptimal for $\theta < \hat{\theta}$.

Now consider $\theta > \hat{\theta}$. Given that the defendant does not make an offer at $t = 1$, x_2^* is optimal for the same reason that x_1^* is optimal for $\theta \leq \hat{\theta}$. Now consider the incentives of $\theta > \hat{\theta}$ at $t = 1$. The construction of p_1^* implies that if $\theta > \hat{\theta}$ makes an offer at $t = 1$, then the best offer is $\underline{x}_1$, which is always rejected and if discovery fails, he will receive the second round equilibrium payoff of type $\hat{\theta}$. If he does not make an offer at $t = 1$, he always receives the equilibrium payoff of type $\theta > \hat{\theta}$. Thus, not making an offer at $t = 1$ is optimal for $\theta > \hat{\theta}$.

(iii) It is clear that the beliefs are consistent with the strategies.

Uniqueness: Suppose $\pi < 1$. First, whenever discovery reveals θ, all types remaining at $t = 2$ will make offers that leave the plaintiff indifferent and the offers will be accepted with probability one.

Second, using the "universally-divine" equilibrium refinement, it cannot be that in equilibrium a positive measure of types make the same offer at $t = 1$ or $t = 2$ and that offer is accepted with positive probability. This is because the highest θ among those types will deviate to a lower (more defendant-favorable) offer. Semi-pooling is similarly ruled out (see Reinganum and Wilde (1986) for details). Therefore, if a positive measure of types make offers at $t = 1$ and their offers are accepted with positive probability, then these offers must be fully separating among these types.

Case 1. A positive measure of types makes offers that are accepted with positive probability at $t = 1$, but not all types.

Consider the subgame following failed discovery for θs whose pre-discovery offers do not reveal their types. Then by Reingannum and Wilde (1986), these types must separate in the subgame.

Let X_t denote the set of offers made at $t = 1, 2$, let $x_t \in X_t$ be an offer, and let $p_t(x_t)$ be the associated rejection probability. Let Θ_{x_1} denote the set of types θ_{x_1} whose offers are accepted with positive probability at $t = 1$. Let Θ_{x_2} denote the set of types θ_{x_2} who wait until $t = 2$ to make offers that are accepted with positive probability.

Step 1. Let $\dot{x}_1 = \inf\{x_1 : x_1 \in X_1\}$, $\ddot{x}_1 = \sup\{x_1 : x_1 \in X_1\}$, and $\dot{x}_2 = \inf\{x_2 : x_2 \in X_2\}$, $\ddot{x}_2 = \sup\{x_2 : x_2 \in X_2\}$. Then $p_1(x_1) = 0$ for $x_1 < \dot{x}_1$ and $p_2(x_2) = 0$ for $x_2 < \dot{x}_2$. $p_t(\cdot)$ is an increasing function over X_t (see Reinganum and Wilde 1986 for details).

Step 2. There do not exist $\theta_{x_1}, \theta_{x_2}$ such that $\theta_{x_2} < \theta_{x_1}$. Suppose there is. Then type θ_{x_2} has a strict incentive to mimic type θ_{x_1} because that offer improves beliefs, which will last to $t = 2$ as well and the $t = 1$ offer is accepted with positive probability. This is a contradiction. Therefore, such $\theta_{x_1}, \theta_{x_2}$ do not exist: X_1, X_2 are connected sets and $\ddot{x}_1 \leq \dot{x}_2$.

Step 3. $p_t(x_t)$ is differentiable on the interior of X_t (Reinganum and Wilde 1986).

Then by the optimality of the offer for each type, $p_1(x_1)$ and $p_2(x_2)$ must satisfy the differential equations with the boundary conditions detailed in the main text.

Case 2. No positive measure of types make acceptable offers at $t = 1$. Then the $t = 2$ analysis mirrors that in Reinganum and Wilde (1986).

Case 3. All types make offers that are accepted with positive probability at $t = 1$. Then the $t = 1$ analysis mirrors that in Reinganum and Wilde (1986).

Proof of Lemma 3: With no discovery, all types are strong, so total litigation costs are $r_2(\theta, 0)c$. With full discovery, all types are weak, so total litigation costs are $r_1(\theta, 1)c$:

$$r_1(\theta, 1) = 1 - \exp\left(-\frac{\theta - \underline{\theta}}{c}\right) = 1 - \exp\left(-\frac{\theta - \hat{\theta}(0)}{c}\right) = r_2(\theta, 0). \quad \square$$

Proof of Lemma 4: Let $(\exp) \equiv \exp\left(-\frac{\theta - \hat{\theta}}{(1-\rho)c}\right)$. Let $\exp_\theta$ denote the derivative of $(\exp)$ with respect to θ. When $\rho(\pi) = \pi$, incorporating $\hat{\theta}'(\pi) = \frac{c}{1-\pi}$,

$$C^s_\pi = 1 - 2(1 - \pi)(1 - (\exp)) - (\exp)(-\frac{\theta - \hat{\theta}}{c} + 1).$$

$$C^s_{\pi,\theta} = 2(1 - \pi)\exp_\theta - \exp_\theta(-\frac{\theta - \hat{\theta}}{c} + 1) + \frac{(\exp)}{c}$$

We have $\exp_\theta = (\exp)(-\frac{1}{(1-\rho)c}) < 0$. That is, $(\exp) = -(1 - \pi)c(\exp)$. Therefore,

$$C^s_{\pi,\theta} = 2(1 - \pi)\exp_\theta - \exp_\theta(-\frac{\theta - \hat{\theta}}{c} + 1) - (1 - \pi)(\exp_\theta) = \exp_\theta[\frac{\theta - \hat{\theta}}{c} - \pi]$$

Because $\exp_\theta < 0$, showing $C^s_{\pi,\theta}(\theta, \pi^*(\theta)) < 0$ is equivalent to showing $\pi^*(\theta) < \frac{\theta - \hat{\theta}(\pi^*(\theta))}{c}$. Because $\pi^*(\theta)$ is interior and it minimizes C^s, by the first order condition we have

$$(2\pi^*(\theta) - 1)(1 - (\exp)) = -\frac{\theta - \hat{\theta}(\pi^*(\theta))}{c}(\exp) \Rightarrow 1 - 2\pi^*(\theta) = \frac{\theta - \hat{\theta}(\pi^*(\theta))}{c}\frac{(\exp)}{1 - (\exp)}.$$

We prove the result by contradiction. Suppose not. That is, suppose that $\pi^*(\theta) \geq \frac{\theta-\hat{\theta}(\pi^*(\theta))}{c}$. Because the right-hand-side is decreasing in $\frac{\theta-\hat{\theta}(\pi^*(\theta))}{c}$, we have

$$1 - 2\pi^*(\theta) \geq \pi^*(\theta)\frac{\exp(-\frac{\pi^*(\theta)}{1-\pi^*(\theta)})}{1-\exp(-\frac{\pi^*(\theta)}{1-\pi^*(\theta)})}$$

Within this proof, define $f(\pi) \equiv (1-2\pi) - \pi\frac{\exp(-\frac{\pi}{1-\pi})}{1-\exp(-\frac{\pi}{1-\pi})}$. Function $f(\pi)$ is strictly decreasing for $\pi \in (0,1)$. Therefore, $f(\pi) < \lim_{\pi\to 0} f(\pi) = 0$. This forms a contradiction. $\square$

Proof of Proposition 6: Suppose $\rho(\pi) = \pi$. Because minimizing litigation costs maximizes discovery benefits, Proposition 2 implies that π_j^* solves:

$$\frac{\partial}{\partial\pi}\int_{\hat{\theta}(\pi)}^{\bar{\theta}} \Delta^s(\theta,\pi)dF_j(\theta) = 0.$$

Because $\Delta^s(\hat{\theta}(\pi),\pi) = 0$, π_j^* solves:

$$\int_{\hat{\theta}(\pi)}^{\bar{\theta}} \Delta_\pi^s(\theta,\pi)dF_j(\theta) = 0 \Leftrightarrow \int_{\hat{\theta}(\pi)}^{\bar{\theta}} \frac{\Delta_\pi^s(\theta,\pi)}{1-F_j(\hat{\theta}(\pi))}dF_j(\theta) = 0.$$

From Lemma 4, $\Delta_\pi^s(\theta,\pi)$ is increasing in θ. Therefore, because $F_2 \succeq_{CFOSD} F_1$, for any given $\hat{\theta}$, F_2 places relatively more probability mass on higher values of θ. The result follows. $\square$

Proof of Proposition 7: (1) Claim: $C^w(\theta,\pi) < C^w(\theta,1)$. Let $\zeta \equiv \frac{\rho(\pi)}{\pi}$. Define $h(\zeta) \equiv \zeta\big(1 - \exp\big(-\frac{\theta-\underline{\theta}}{\zeta c}\big)\big)c = C^w(\theta,\pi)$ for $\zeta < 1$. Note that as $\rho(1) = 1$, $C^w(\theta,\pi) < C^w(\theta,1)$ is equivalent to $h(\zeta) < h(1)$. With convex ρ, because ζ is strictly increasing in π for strictly convex $\rho(\pi)$ by Lemma 1, to prove $C^w(\theta,\pi) < C^w(\theta,1)$ ($h(\zeta) < h(1)$) for convex $\rho(\pi)$, it suffices to show that $h(\zeta)$ is strictly increasing in ζ. We have, $h'(\zeta) = c(1-\exp(-\delta)(1+\delta)) \equiv l(\delta)$, where $\delta = \frac{\theta-\underline{\theta}}{\zeta c}$. Note that $l(0) = 0$, and $l(\delta)$ increases in δ. Because $\delta > 0$ for $\theta > \underline{\theta}$, $l(\delta) > 0$. That is, $h'(\zeta) > 0$.

(2) Claim: $C^s(\theta,\pi) < C^s(\theta,0)$ for $\pi \in (0,\hat{\pi}(\theta))$. If $\rho(\pi)$ is strictly convex in the extent of discovery, the reduction in investigation costs associated with discovery is reinforced. To see this,

observe that because $\frac{\rho(\pi)}{\pi} < 1$, the convexity of $\exp(\cdot)$ implies that for $\pi \in (0, \bar{\pi})$,

$$\frac{\rho(\pi)}{\pi} \exp\Big(-\frac{\hat{\theta}(\pi) - \underline{\theta}}{\frac{\rho(\pi)}{\pi} c}\Big) + (1 - \frac{\rho(\pi)}{\pi}) \exp(0) > \exp\Big(-\frac{\hat{\theta}(\pi) - \underline{\theta}}{c} + 0\Big) = \exp\Big(-\frac{\hat{\theta}(\pi) - \underline{\theta}}{c}\Big). \quad (6)$$

Substituting the implicit solution for $\hat{\theta}$,

$$1 - \pi = \exp\Big(-\frac{\hat{\theta}(\pi) - \underline{\theta}}{\frac{\rho(\pi)}{\pi} c}\Big),$$

into the left-hand side of inequality (6) reveals that it simplifies to $1 - \rho(\pi)$. Therefore,

$$1 - \rho(\pi) > \exp\Big(-\frac{\hat{\theta}(\pi) - \underline{\theta}}{c}\Big).$$

Therefore,

$$\frac{1}{c} C^s(\theta, 0) > \rho(\pi) + (1 - \rho(\pi)) R(\theta - \hat{\theta})$$
$$\frac{1}{c} C^s(\theta, \pi) = \rho(\pi) + (1 - \rho(\pi))[\pi R(0) + (1 - \pi) R\Big(\frac{\theta - \hat{\theta}(\pi)}{1 - \rho(\pi)}\Big)].$$

$$\begin{aligned}\pi R(0) + (1 - \pi) R\Big(\frac{\theta - \hat{\theta}(\pi)}{1 - \rho(\pi)}\Big) &< \rho(\pi) R(0) + (1 - \rho(\pi)) R\Big(\frac{\theta - \hat{\theta}(\pi)}{1 - \rho(\pi)}\Big) \\ &< R\Big(\rho(\pi) 0 + (1 - \rho(\pi)) \frac{\theta - \hat{\theta}(\pi)}{1 - \rho(\pi)}\Big) = R(\theta - \hat{\theta}).\end{aligned}$$

The first inequality follows from $1 - \rho(\pi) > 1 - \pi$, and the second follows from concavity of $R(\cdot)$.

Therefore, $C^s(\theta, \pi) < C^s(\theta, 0)$.

(3) $\pi^*(\theta) < \hat{\pi}(\theta)$. First, C^w strictly increases in π. Second, we show $C^s_\pi(\theta, \hat{\pi}(\theta)) > 0$:

$$C^s_\pi(\theta,\pi) = \frac{d}{d\pi}[\rho + (1-\pi)(1-\rho)] - \frac{d}{d\pi}[(1-\pi)(1-\rho)] \exp\left(-\frac{\theta-\hat{\theta}}{(1-\rho)c}\right) + (1-\pi)(1-\rho)\exp\left(-\frac{\theta-\hat{\theta}}{(1-\rho)c}\right)\left[\frac{\theta-\hat{\theta}}{(1-\rho)^2c}\rho' - \frac{\hat{\theta}'}{(1-\rho)c}\right].$$

When $\pi = \hat{\pi}(\theta)$, we have $\hat{\theta}(\hat{\pi}(\theta)) = \theta$. This implies that when $\pi = \hat{\pi}(\theta)$, we have $\exp\left(-\frac{\theta-\hat{\theta}}{(1-\rho)c}\right) = 1$. Therefore, the left derivative evaluated at $\hat{\pi}(\theta)$ is:

$$C^s_\pi(\theta, \hat{\pi}(\theta)) = \rho'(\hat{\pi}) - \frac{\rho(\hat{\pi})}{\hat{\pi}} > 0.$$

(4) Claim: the reduction in litigation costs rises in θ for $\theta < \hat{\theta}$. Strict convexity implies $\rho(\pi)/\pi < 1$.

$$\begin{aligned}\Delta^w(\theta,\pi) &= r_1(\theta,1)c - r_1(\theta,\pi)\rho(\pi)c\\ &= (1-\exp\Big(-\frac{\theta-\underline{\theta}}{c}\Big))c - \frac{\rho(\pi)}{\pi}(1-\exp\Big(-\frac{\theta-\underline{\theta}}{\frac{\rho(\pi)}{\pi}c}\Big))c. \Rightarrow\\ \Delta^w_\theta &= \exp\Big(-\frac{\theta-\underline{\theta}}{c}\Big) - \exp\Big(-\frac{\theta-\underline{\theta}}{\frac{\rho(\pi)}{\pi}c}\Big) > 0. \quad \square\end{aligned}$$

The following lemma is used in the proof of Proposition 8.

Lemma 6. $\Delta^s_\theta > 0$ *if costs are convex. If costs are concave and* $\Delta^s(\hat{\hat{\theta}}, \pi) > 0$ *for some* $\hat{\hat{\theta}}$*, then* $\Delta^s(\theta,\pi) > \Delta^s(\hat{\hat{\theta}}, \pi)\ \forall \theta > \hat{\hat{\theta}}$*: if any type gains from discovery, then stronger types gain even more.*

Proof:

$$\frac{1}{c}\Delta^s = \Big(1-\exp\Big(-\frac{\theta-\underline{\theta}}{c}\Big)\Big) - \rho(\pi) - (1-\rho(\pi))(1-\pi)\Big(1-\exp\Big(-\frac{\theta-\hat{\theta}}{(1-\rho(\pi))c}\Big)\Big)$$

$$\Delta^s_\theta = \exp\Big(-\frac{\theta-\underline{\theta}}{c}\Big) - (1-\pi)\exp\Big(-\frac{\theta-\hat{\theta}}{(1-\rho(\pi))c}\Big)$$

$$c\Delta^s_{\theta,\theta} = \frac{1-\pi}{1-\rho(\pi)}\exp\Big(-\frac{\theta-\hat{\theta}}{(1-\rho(\pi))c}\Big) - \exp\Big(-\frac{\theta-\underline{\theta}}{c}\Big).$$

Let $\tilde{\tilde{\theta}}$ solve $\Delta_\theta^s(\tilde{\tilde{\theta}}, \pi) = 0$. Then,

$$c\Delta_{\theta,\theta}(\tilde{\tilde{\theta}}, \pi) = (1-\pi)\left(\frac{1}{1-\rho(\pi)} - 1\right)\exp\Big(-\frac{\theta - \hat{\theta}}{(1-\rho(\pi))c}\Big) > 0.$$

Then $\tilde{\tilde{\theta}}$ is at most unique among $\theta > \hat{\theta}$. If $\tilde{\tilde{\theta}}$ exists, then let $\hat{\hat{\theta}} \equiv \max\{\tilde{\tilde{\theta}}, \hat{\theta}\}$. Then for all $\theta > \hat{\hat{\theta}}$, $\Delta_\theta(\theta, \pi) > 0$.

Note that for strictly convex costs of discovery, for all $\theta > \hat{\theta}$,

$$\Delta_\theta^s(\hat{\theta}, \pi) = \exp\Big(-\frac{\theta - \underline{\theta}}{c}\Big) - (1-\pi)\exp\Big(-\frac{\theta - \hat{\theta}}{(1-\rho(\pi))c}\Big) = \exp\Big(-\frac{\theta - \underline{\theta}}{c}\Big) - \exp\Big(-\frac{\theta - \underline{\theta}}{\frac{\rho(\pi)}{\pi}c}\Big) > 0.$$

Proof of Proposition 8: (1) Claim: $C^w(\theta, \pi) > C^w(\theta, 1)$. Let $\zeta \equiv \frac{\rho(\pi)}{\pi}$. Define $h(\zeta) \equiv \zeta\big(1 - \exp\big(-\frac{\theta - \underline{\theta}}{\zeta c}\big)\big)c = C^w(\theta, \pi)$ for $\zeta < 1$. Note that because $\rho(1) = 1$, $C^w(\theta, \pi) < C^w(\theta, 1)$ is equivalent to $h(\zeta) < h(1)$. Because ζ is strictly decreasing in π for strictly concave $\rho(\pi)$ by Lemma 1, to prove $h(\zeta) > h(1)$, it suffices to show that $h'(\zeta) > 0$. This was done in step 1 of the proof of Proposition 7.

(2) C^w is strictly decreasing over $\pi > \hat{\pi}(\theta)$. Also,

$$C_\pi^s(\theta, \hat{\pi}(\theta)) = \rho'(\hat{\pi}) - \frac{\rho(\hat{\pi})}{\hat{\pi}} < 0$$

Therefore, there are two cases regarding the minimizer $\pi^*(\theta)$: either $\pi^*(\theta)$ is either 0 or 1 (because $C^s(\theta, 0) = C^w(\theta, 1)$), or $\pi^*(\theta) \in (0, \hat{\pi}(\theta))$. In the former case, all types are hurt by discovery. That is $\Delta(\theta, \pi) < 0$ for any $\pi \in (0, 1)$. Set $\tilde{\theta}(\pi) = \overline{\theta}$ in this case. In the latter case, $\Delta(\theta, \pi) > 0$ for some $\theta > \hat{\theta}(\pi)$ and $\Delta(\theta, \pi) < 0$ for $\theta < \hat{\theta}(\pi)$. By Lemma 6, there exists a cutoff $\tilde{\theta}(\pi) > \hat{\theta}(\pi)$ such that $\Delta(\tilde{\theta}(\pi), \pi) = 0$. The rest follows from Lemma 6.

(3) If $\pi \in (0, \bar{\pi})$, then $\hat{\theta}(\pi) \in (\underline{\theta}, \bar{\theta})$. With concave costs, weak types have expected litigation costs that exceed those when $\pi \in \{0, 1\}$ by an amount that we can bound from below by some $\gamma > 0$, and the strongest type $\theta = \bar{\theta}$ benefits (if at all) by less than 1. Thus, any density $g(\theta)$ that places cumulative probability of at least $\frac{1}{1+\gamma}$ on $\theta \in [\underline{\theta}, \hat{\theta}(\pi)]$ has higher costs with discovery π than without. $\square$

Proof of Proposition 9: For $\pi \in (0,1)$ and $\theta > \underline{\theta}$,

$$\frac{\partial}{\partial z} C^w(\theta, \pi)/c = \pi^{z-1}(-\ln \pi)[\exp\left(-\frac{\theta - \underline{\theta}}{\pi^{z-1}c}\right)(\frac{\theta - \underline{\theta}}{\pi^{z-1}c} + 1) - 1] < 0.$$

Let $\delta \equiv \frac{\theta - \underline{\theta}}{\pi^{z-1}c}$. The inequality is because $f(\delta) \equiv \exp(-\delta)(\delta + 1) - 1$ decreases in δ and $f(0) = 0$.

When $\rho(\pi) = \pi^z$, we have,

$$\frac{1}{c} C^s(\theta, \pi) = \pi^z + (1 - \pi)(1 - \pi^z) \exp\left(-\frac{\theta - \hat{\theta}}{(1 - \pi^z)c}\right)$$

Let $\lambda = \frac{\theta - \hat{\theta}}{(1-\pi^z)c}$, incorporating $\frac{d}{dz}\hat{\theta} = (-\ln \pi)\pi^z \frac{\ln(1-\pi)}{\pi}$,

$$\frac{1}{c} C^s_z(\theta, \pi) = (-\ln \pi)\pi^z[-1 + (1 - \pi) \exp(-\lambda)(1 + \lambda + \frac{\ln(1 - \pi)}{\pi})]$$

Note that $\exp(-\lambda)(1 + \lambda + a)$ is strictly decreasing in λ for any $a \geq 0$. Therefore,

$$-1 + (1 - \pi) \exp(-\lambda)(1 + \lambda + \frac{\ln(1 - \pi)}{\pi}) < -1 + (1 - \pi) \exp(-0)(1 + 0 + \frac{\ln(1 - \pi)}{\pi}) < 0.$$

This implies $C^s_z(\theta, \pi) < 0$. $\square$

Proof of Lemma 5: First note that because $\rho(y) \in [0,1]$ and $y \in [0,1]$, $\rho(y) + (1-y)(1-\rho(y)) \leq 1$ for any $y \in [0,1]$. There are only two solutions to $\rho(y) + (1 - y)(1 - \rho(y)) = 1$: $y = 0$ or $y = 1$. This implies that for all $y \in (0,1)$, $\rho(y) + (1 - y)(1 - \rho(y)) < 1$. That is, $\Delta(y) > 0$. When $\rho(y) = y$, $K'(y) = 2y - 1$ and $K''(y) = 2 > 0$, so the minimizer of $K(y)$ solves $K'(y) = 0$. $\square$

Proof of Proposition 10: Differentiating C^s with respect to π yields:

$$\begin{aligned} C^s_\pi(\theta, \pi) = &\frac{d}{d\pi}[\rho + (1 - \pi)(1 - \rho)] - \frac{d}{d\pi}[(1 - \pi)(1 - \rho)] \exp\left(-\frac{\theta - \hat{\theta}}{(1 - \rho)c}\right) \\ &+ (1 - \pi)(1 - \rho) \exp\left(-\frac{\theta - \hat{\theta}}{(1 - \rho)c}\right) \left[\frac{\theta - \hat{\theta}}{(1 - \rho)^2 c}\rho' - \frac{\hat{\theta}'}{(1 - \rho)c}\right]. \end{aligned}$$

Evaluate this derivative at the benchmark level $\pi = y^*$. Because $\frac{d}{d\pi}[\rho + (1 - \pi)(1 - \rho)]|_{\pi = y^*} = 0$, we have $\frac{d}{d\pi}[(1-\pi)(1-\rho)]|_{\pi=y^*} = -\rho'(y^*)$. Coupled with $\hat{\theta}' = \frac{\rho c}{\pi(1-\pi)}$, we have:

$$C^s_\pi(\theta, y^*) = [\rho'(y^*) + \frac{1-y^*}{1-\rho(y^*)}\frac{\theta-\hat{\theta}(y^*)}{c}\rho'(y^*) - \frac{\rho(y^*)}{y^*}] \exp\left(-\frac{\theta-\hat{\theta}(y^*)}{(1-\rho(y^*))c}\right).$$

With linear discovery, $\rho(\pi) = \pi$, we have

$$C^s_\pi(\theta, y^*) = \frac{\theta-\hat{\theta}(y^*)}{c}\exp\left(-\frac{\theta-\hat{\theta}(y^*)}{(1-y^*))c}\right) > 0.$$

By definition of y^*, if $\pi > y^*$, then $\rho' + \frac{d}{d\pi}[(1-\pi)(1-\rho)] > 0$. Let $(\exp) \equiv \exp\left(-\frac{\theta-\hat{\theta}}{(1-\rho)c}\right) \leq 1$.

$$C^s_\pi(\theta, \pi) = [\rho' - \frac{\rho}{\pi}](\exp) + [\rho' + \frac{d}{d\pi}[(1-\pi)(1-\rho)]](1-(\exp)) + \frac{1-\pi}{1-\rho}\frac{\theta-\hat{\theta}}{c}\rho'(\exp).$$

With linear discovery, $\rho' - \frac{\rho}{\pi} = 0$, so $C^s_\pi(\theta, \pi) > 0$. This implies that for any $\pi > y^*$, the litigation cost is increasing. Therefore, the litigation cost must be minimized at $\pi^*(\theta) < y^*$. Because it holds for any θ, the optimal given any distribution of θ must have the same property. $\square$

Proof of Proposition 11: Given discovery π, denote the pre-discovery offer by x_1 and, when discovery fails, denote the post-discovery offer by x_2. Let θ_1 denote the cutoff type below which the defendant accepts x_1, and θ_2 denote the cutoff above which the defandant rejects x_2. It is equivalent to think of the plaintiff choosing cutoffs θ_1 and θ_2, with x_1 and x_2 uniquely determined by θ_1 and θ_2.

Post-discovery asymmetric information. Because θ_2 is indifferent between accepting the offer x_2 and going to trial with the additional cost of $(1-\pi)c_d$, $x_2 = 1-\theta_2+(1-\pi)c_d$. Let F_{θ_1} denote the truncated distribution over $[\theta_1, \overline{\theta}]$. Given θ_1, the plaintiff chooses θ_2 to maximize:

$$x_2 F_{\theta_1}(\theta_2) - (1-F_{\theta_1}(\theta_2))(1-\pi)c_p + \int_{\theta_2}^{\overline{\theta}}(1-\theta)dF_{\theta_1}(\theta).$$

Solving the associated first-order condition,

$$F_{\theta_1}(\theta_2) = f_{\theta_1}(\theta_2)(1-\pi)c \Rightarrow \frac{F(\theta_2)-F(\theta_1)}{f(\theta_2)} = (1-\pi)c.$$

When the bottom tail of f is uniform on $[\underline{\theta}, \underline{\theta}+c]$, this simplifies to $\theta_2 = \theta_1 + (1-\pi)c$.

Post-discovery full information. Suppose that discovery succeeds in revealing the defendant's type. Then, the plaintiff's offer of $1-\theta+(1-\pi)c_d$ extracts all surplus from the defendant.

Pre-discovery. If cutoff type θ_1 accepts the offer, her payoff is $-x_1$. If θ_1 rejects the offer, her payoff is $-\pi c_d - \pi(1-\theta+(1-\pi)c_d) - (1-\pi)x_2$. This expression reflects that if x_1 is rejected, the defendant must pay the discovery cost; if discovery then succeeds the defendant pays the full information offer; and if discovery fails the defendant pays the post-discovery offer under asymmetric information. This implies $x_1 = \pi(1-\theta_1) + (1-\pi)(1-\theta_2) + c_d$.[18]

The plaintiff's optimal pre-discovery cutoff, θ_1, maximizes:

$$\begin{aligned} &- c_p + \pi[F(\theta_1)(1-\theta_1) + F(\theta_1)c + \int_{\theta_1}^{\bar{\theta}} (1-\theta)dF(\theta) + (1-F(\theta_1))(1-\pi)c] + \\ &\qquad (1-\pi)[F(\theta_2)(1-\theta_2) + F(\theta_1)c + \int_{\theta_2}^{\bar{\theta}} (1-\theta)dF(\theta) + (F(\theta_2)-F(\theta_1))(1-\pi)c] \end{aligned}$$

The term $(1-F(\theta_1))(1-\pi)c$ at the end of the first line reflects that when discovery succeeds, the plaintiff extracts the remaining trial costs from defendant types $\theta \in [\theta_1, \bar{\theta}]$ and saves on his own remaining trial costs, gaining a total of $(1-\pi)c$. The final term $(F(\theta_2)-F(\theta_1))(1-\pi)c$ reflects that even when discovery fails, intermediate types $\theta \in [\theta_1, \theta_2]$ still settle allowing the plaintiff to again extract the remaining trial costs of the defendant and save his own remaining trial costs, gaining a total of $(1-\pi)c$. The objective can be rewritten as:

$$\begin{aligned} &- c_p + \pi[F(\theta_1)(1-\theta_1+c) + \int_{\theta_1}^{\bar{\theta}} (1-\theta)dF(\theta)] + \\ &\qquad (1-\pi)[F(\theta_2)(1-\theta_2+c) + \int_{\theta_2}^{\bar{\theta}} (1-\theta)dF(\theta)] + \pi(1-\pi)(1-F(\theta_2))c. \end{aligned}$$

Note that θ_2 is implicitly a function of θ_1. Plugging in $F(\theta_2) = F(\theta_1) + (1-\pi)f(\theta_2)c$, the

[18] If successful discovery eliminates the remaining trial cost, then x_1 would be lower by $\pi(1-\pi)c_d$, reflecting that there is less trial costs to be extracted from the defendant. As a result, the plaintiff's payoff will be reduced by $\pi(1-\pi)(1-F(\theta_1))c_d$. This would give the plaintiff an incentive to increase the cutoff θ_1 to reduce the chance that it cannot extract the remaining trial costs. When c_p is relatively small, the defendant may gain from discovery.

derivative with respect to θ_1 is:

$$\pi(-F(\theta_1) + f(\theta_1)c) - (1-\pi)F(\theta_1)\theta_2'(\theta_1) = 0$$

When the bottom tail $[\underline{\theta}, \underline{\theta} + c]$ is uniformly distributed, $\theta_2'(\theta_1) = 1$, so the first order condition becomes $F(\theta_1) = f(\theta_1)\pi c$. Therefore, $\theta_1^* = \underline{\theta} + \pi c$ and $\theta_2^* = \underline{\theta} + c$. This implies $x_1^* = 1 - \underline{\theta} - \pi c + c_d - (1-\pi)^2 c$ and $x_2^* = 1 - \underline{\theta} - \pi c - (1-\pi)c_p$.

Proof of Corollary 7: First, we consider the defendant's payoff.

Strong defendant types $\theta > \underline{\theta} + c$ reject pre-discovery offers, incurring discovery cost πc_d. If discovery succeeds they pay the full information settlement offer $1 - \theta + (1-\pi)c_d$ and if it fails, they incur trial cost $(1-\pi)c_d$ and pay out $1-\theta$ at trial. In both cases, the total payment is $c_d + 1 - \theta$, regardless of π. Therefore, discovery does not affect their payoffs.

Defendants types $\theta \in [\underline{\theta} + \pi c, \underline{\theta} + c]$ also incur discovery cost πc_d. If discovery succeeds, they settle at $1-\theta+(1-\pi)c_d$. If discovery fails, they pay the post-discovery offer of $1-(\underline{\theta}+c)+(1-\pi)c_d$. Thus, they expect to pay $\pi(1-\theta+c_d) + (1-\pi)(1-(\underline{\theta}+c)+c_d) = 1 - \pi\theta - (1-\pi)(\underline{\theta}+c) + c_d$. Without discovery, they settle, paying $(1 - (\underline{\theta} + c) + c_d)$. Thus, a type θ defendant expects to pay an additional $\pi(\underline{\theta} + c - \theta)$ relative to when there is no discovery.

Lastly, the pre-discovery offer to defendants $\theta < \underline{\theta} + \pi c$ is $1 - \underline{\theta} - \pi c + c_d - (1-\pi)^2 c$. Without discovery, the settlement offer will be $1 - \underline{\theta} - c_p$. Therefore, each type pays $\pi(1-\pi)c$ more with discovery.

Now consider the payoff for the plaintiff. We have established that the expected reduction in total litigation costs associated with any strong type $\theta > \underline{\theta} + c$ is $\pi(1-\pi)c$, and the benefits of this reduction all accrue to the plaintiff who extracts them in his settlement offer; and the increased settlement by any type $\theta < \underline{\theta} + \pi c$ is also $\pi(1-\pi)c$. For the remaining types $\theta \in [\underline{\theta} + \pi c, \underline{\theta} + c]$, relative to no discovery, the defendant pays an additional $\pi(\underline{\theta} + c - \theta) - \pi c_d$ (which may not be positive) and the plaintiff incurs a cost of πc_p. Therefore, for this intermiediate type, the plaintiff incurs a cost of $\pi(\theta - \underline{\theta})$. The benefits for the plaintiff from the weak and the strong segments of types intergrate to $\pi(1-\pi)c\frac{\overline{\theta}-\underline{\theta}-(1-\pi)c}{\overline{\theta}-\underline{\theta}}$. The costs from the intermediate segment intergrate to $\pi(1-\pi)c\frac{(1+\pi)c}{2(\overline{\theta}-\underline{\theta})}$. Therefore, the net gain for the plaintiff is $\pi(1-\pi)c\frac{\overline{\theta}-\underline{\theta}-2c+\frac{3}{2}(1+\pi)c}{\overline{\theta}-\underline{\theta}}$. When $c \in (\frac{\overline{\theta}-\underline{\theta}}{2}, \frac{\overline{\theta}-\underline{\theta}}{2-\frac{3}{2}(1+\pi)})$, the total litigation costs increase but the plaintiff's payoff increases under discovery. The payoff of the plaintiff benefits an additional $\pi(1-\pi)c\frac{\frac{3}{2}(1+\pi)c}{\overline{\theta}-\underline{\theta}}$ from discovery than the sum of the payoffs of the plaintiff and the defendant. The first order derivative of this term has the same sign as $(1-2\pi)(1+\pi)+\pi(1-\pi)$. (1) When $\overline{\theta}-\underline{\theta} > 2c$, the socially optimal level of discovery is $\frac{1}{2}$. At $\pi = \frac{1}{2}$, the first order derivative of $\pi(1-\pi)(1+\pi)$ is $\frac{1}{4}$, i.e. positive, so the payoff of the plaintiff is maximized at $\pi > \frac{1}{2}$. (2) When $\overline{\theta}-\underline{\theta} \leq 2c$, the first order derivative of $\pi(1-\pi)(1+\pi)$ is 1, i.e. positive, so the payoff of the plaintiff is maximized at $\pi > 0$.

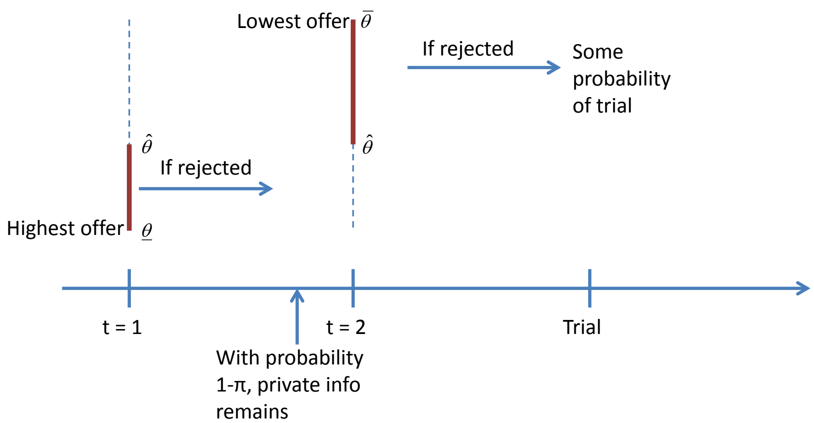

Figure 1: Separating by offer amount and timing under limited discovery $\pi \in (0, \bar{\pi})$.

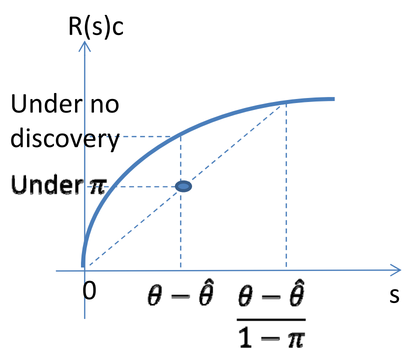

Figure 2: Discovery reduces post-discovery signaling distortions for strong defendants if $\rho(\pi) = \pi$.

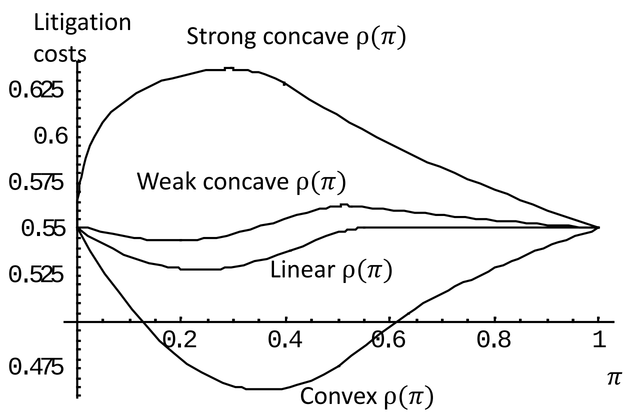

Figure 3: Total expected litigation costs with different discovery cost functions. The convex cost function is $\rho(\pi) = \pi^{\frac{3}{2}}$, the "strong concave" cost function is $\rho(\pi) = \pi^{\frac{1}{2}}$ and the "weak concave" cost function is $\rho(\pi) = \pi^{0.9}$. Parameters: $\underline{\theta} = 0$, $\theta = 0.8$, $c = 1$.