The Optimal Extent of Discovery *

Dan Bernhardt and Frances Zhiyun Xu

March 26, 2014

Abstract

We characterize how the process of publicly-gathering information via discovery affects strategic interactions between plaintiffs and privately-informed defendants. We endogenize the timing and size of settlement offers, and their equilibrium probabilities of acceptance. Discovery allows defendants to signal the strengths of their cases via the offer timing. Weaker defendants attempt to settle pre-discovery, while stronger defendants wait until later. Discovery facilitates separation of defendant types, reducing the likelihoods that plaintiffs reject settlements. Conventional wisdom about welfare impacts is overturned: privately-informed defendants gain from limited discovery, with stronger defendants gaining more. Greater discovery is optimal when more defendants have stronger cases.

*We thank Odilon Camara, Wing Suen, Satoru Takahashi and the audience at National University of Singapore.

1 Introduction

The law and economics literature typically compresses the litigation process in a civil law suit—the information gathering, litigation costs and legal outcome—into a single trial stage, possibly combined with a pre-trial settlement offer stage. In practice, the litigation process takes place over time, and litigants can attempt settlements prior to discovery, during discovery or post-discovery, prior to a trial that will come to a final legal determination of a defendant's liability. During discovery, defendants and plaintiffs uncover information about whether a defendant is liable, and the extent to which the defendant is liable. The extent of discovery is such that most litigation expenses—gathering and reviewing case materials, preparing reports, hiring experts, pre-trial depositions, etc.—are incurred in the process of discovery, rather than in the trial itself.[1]

Despite discovery's prominent real world role, there has been little analysis of how the design of discovery affects decision-making by litigants. The analysis that exists suggests that the possibility of discovery harms the privately-informed party (see Hay 1994, Sobel 1989, Schrag 1999, or Shavell 1989). In this paper, we endogenize the timing of settlement offers throughout the process of discovery, and show that conventional wisdom about the effects of discovery is overturned. In our analysis, we answer the questions—how does the discovery process affect settlement offers and the timing of settlement? how does the division of information gathering between discovery and trial affect the litigation costs incurred in equilibrium? and, who benefits when from discovery?

At the outset of many litigations, a plaintiff and defendant are often very asymmetrically informed about their prospects in a trial. We focus on a setting in which the defendant alone has private information. As the parties go through the litigation process, (a) these information asymmetries may be reduced via costly discovery that may uncover a defendant's private information; and, (b) a defendant will have multiple opportunities to make settlement offers that endogenously convey further information to a plaintiff about the defendant's information. A pre-discovery settlement offer that is rejected does not mean that the parties must go to trial, but only that the litigants will proceed to discovery. Exactly how much information is conveyed by an offer will depend not only on the offer itself, but also on its timing relative to discovery, and on the extent of discovery.

Our model of the process of discovery builds on Reinganum and Wilde (1986). Reinganum

[1] Final Colorado Civil Access Pilot Project Overview 8-31-11, Website of the Colorado Supreme Court.

and Wilde consider a plaintiff that is privately-informed about the expected damage award were the case to go to trial.[2] In their model, the informed party 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. Reinganum and Wilde derive how the size of a settlement offer is related in equilibrium to the private information about the plaintiff's chances at a trial. Equilibrium offers are separating, and, to induce the informed party to truthfully reveal its private information, the uninformed party must be more likely to accept more generous settlement offers, so that less generous offers are more likely to result in a costly trial.

With public discovery, both the extent of information revelation and its costs are spread out over time. In our model, each round of discovery is characterized by (a) the probability that a defendant's private information will be uncovered publicly, and (b) the costs that the defendant and plaintiff incur in that round of discovery. More extensive discovery that is more likely to uncover a defendant's private information incurs a greater share of the total litigation costs. Before each round of discovery, a defendant can make a settlement offer; if the plaintiff rejects the offer, 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 all rounds of discovery, absent a settlement, a trial determines whether the 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. Discovery opens a new channel of signaling: the timing of a settlement offer. We find that 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 settlement offers of defendant types making offers at the same point in the litigation process 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 may wait until after discovery to make their first offers; with even stronger types willing to wait for more rounds of discovery, unless earlier discovery uncovered their private information, in which case they settle immediately.

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

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 succeeds in uncovering 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, and hence delivers 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. Our base model has one round of discovery, and should the litigants go to trial, the trial costs only depend on the extent of discovery, and not on whether discovery succeeds or fails (of course, after successful discovery, cases are always settled prior to trial). More extensive discovery reduces the likelihood that a pre-discovery settlement offer is rejected, but raises the discovery costs incurred when those offers are rejected. When the discovery technology is such that the share of costs incurred in discovery rise one-for-one with the extent of discovery, these two effects exactly cancel out for weak defendants that make their first offers prior to discovery—so expected total litigation costs incurred with weak defendants do not vary with the extent of discovery. If, instead, discovery selectively targets lower marginal cost sources of information, then limited discovery reduces expected litigation costs for weak defendants who make their first offers prior to the discovery.

In contrast, regardless of its cost structure, discovery always facilitates signaling by those stronger defendant types that wait until after discovery to make their first offers, reducing expected litigation costs for all strong defendants. Separation of strong defendants via discovery is linear, affecting all types in exactly the same way: after successful discovery, each defendant makes the settlement associated with its true type, and it is *accepted*. However, when discovery fails, a strong type must signal via its settlement offer. Unlike with successful discovery, with endogenous separation via settlement offers, a defendant that mimics a better type's offer is perceived as a better type even when its offer is rejected, so it must face a higher probability of rejection—its expected litigation expenditures must be higher—in order to preserve incentive compatible revelation of the strength of its case. In effect, discovery reduces the inefficiencies in the rates at which settlement offers by strong defendant types must be rejected by a plaintiff, reducing expected litigation costs.

Thus, the conventional wisdom that discovery that reveals one party's private information harms that party is overturned: in our base model, sufficiently limited discovery *always* reduces the expected total costs incurred in litigation, and the informed defendant extracts all of these gains. Indeed, this qualitative finding is reinforced by more rounds of discovery—dividing limited discovery into more rounds facilitates signaling, reducing the expected litigation costs incurred, benefiting the informed defendant. Intuitively, "shrinking the distance" from the worst type to be separated from in a given round of settlement offering reduces the distortion required. We also show that dividing discovery into more rounds induces more defendant types to "play chicken": with more rounds, more defendant types are willing to wait until after *all* discovery ends, making their first offer just before a trial whenever discovery fails.

We then characterize the extent of discovery that minimizes expected litigation costs, which, in turn, maximizes a defendant's ex-ante payoffs. With more extensive discovery, more defendant types choose to make offers pre-discovery to avoid incurring the discovery costs; and if discovery is so extensive that all defendants make pre-discovery offers, expected litigation outcomes are the same with discovery as without. Thus, limited discovery is always optimal. The tradeoffs are clearest when discovery costs rise one-for-one with the extent of discovery, so that only strong types that wait until after discovery to make their first offers benefit from discovery. We establish that among those defendants that wait until after discovery to make their first settlement offers, the stronger is the defendant, the more it benefits from any given level of discovery. This directly implies that greater discovery is optimal when the distribution of defendant types is better/stronger in the conditional first order stochastic dominance sense.

Our base model supposes that all litigation costs are purely informational in nature, concerning only the costs of uncovering information. That is, we initially abstract away from any fixed procedural costs associated with trials in order to show that they are not necessary to make positive, but limited, discovery optimal. Our qualitative findings extend when there are fixed trial costs—trial costs that are unrelated to the presentation of evidence. To the extent that trials feature such non-information based costs, more extensive discovery becomes optimal, as more extensive discovery always reduces the probability that a case ultimately goes to trial.

Our results also extend when successful discovery that uncovers the defendant's private information reduces the costs that litigants must incur *if* they go to trial. One might conjecture that expected litigation costs would be unaffected by this possibility because the litigants always settle pre-trial following successful discovery. That is, one might conjecture that such potential savings at trial from successful discovery *only* represent a transfer from the defendant to the plaintiff, as a defendant in its take-it-or-leave-it offers can no longer use the trial costs as a threat to support a lower settlement. In fact, we show that such trial savings *reduce* expected total litigation costs by facilitating separation of weak defendant types that make pre-discovery offers: the cost to a weaker defendant type of mimicking a stronger type's offer is raised because if its offer is rejected and discovery succeeds, then the defendant's post-discovery bargaining position is worsened. We establish that the plaintiff now benefits from discovery because a successful discovery reduces her trial cost, reducing what the defendant can threaten it with in the post-discovery settlement. Indeed, a defendant may be hurt by discovery that reduces trial costs, even though the two parties collectively benefit.

The Discovery Literature. Hay (1994) assumes that greater discovery raises a plaintiff's chance of winning against a defendant, so that more discovery always hurts the defendant. 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, while post-discovery, the plaintiff can make an offer. This exogenous structure precludes the possibility of using the timing of when an offer is made to signal, 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, which implies that mandatory discovery again hurts the informed plaintiff. Schrag (1999) endogenizes the extent of discovery chosen by a plaintiff whose expenditures on discovery will hinge on how likely it believes discovery to uncover dirt about a defendant. A semi-pooling equilibrium may exist in which both a strong defendant (who is sure to win at trial) and weak one (who may win at trial) refuse to settle before discovery. Schrag shows that if the court exogenously limits the extent of discovery, it can raise the chances of settlement before discovery. In sharp contrast to this entire discovery literature, in our model, rather than hurting the informed party whose private information discovery reveals, discovery *always helps* it.

We model discovery as an inevitable step in the litigation process unless the litigants settle. Schwartz and Wickelgren (2009) model discovery as a conscious choice by the uninformed party. In their screening model, an uninformed defendant makes settlement offers pre- and post-discovery. The defendant's low pre-discovery offer keeps the threat of discovery credible in case its offer is rejected. Farmer and Percorino (2005) incorporate mandatory disclosure as a conscious choice of the uninformed party in the Reinganum and Wilde 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, which would raise 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.[3] 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 love to convey that non-existence, but it has no way to directly do so: its own selective disclosure would be suspected since a defendant with damaging evidence would like to conceal it and mimic the harmless non-information disclosed by a defendant with little to hide. In this circumstance, discovery that lets an uninformed plaintiff examine a subset of the documents of its *own* choosing is one way to (stochastically) and credibly uncover the existence or non-existence of damaging evidence. Alternatively, a defendant can *indirectly* convey its private information by signaling via its settlement offer.

The paper is organized as follows. Section **??** sets out the model for a given level of discovery. Section **??** characterizes the equilibrium. Section **??** analyzes how discovery affects the payoffs of the two litigating parties and characterizes how the optimal extent of discovery depends on the parameters describing the legal environment. Section **??** considers variations of the basic model including an extension to multiple rounds of discovery. Section **??** concludes. Most proofs are in an Appendix.

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

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. Given defendant's private information, the expected probability of the defendant being not liable at the trial is $\theta \in [\underline{\theta}, \bar{\theta}] \subset (0,1)$ where $\theta \sim F(\cdot)$. That is, the defendant privately observes 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, while 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.[4] At $t = 2$, post-discovery, the 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 purely informational in nature, dealing solely with the gathering and assessing of information about whether the defendant is liable. If the case proceeds all the way to trial, the plaintiff would incur a total investigation cost of $c_p > 0$, while 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. These costs of investigation are incurred both in discovery and at trial. The extent of discovery is characterized by the probability π that discovery

[4]Even when discovery fails, it may reveal relevant information to the lawsuit. It is just that it does not reveal the private part of the information.

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 investigation costs of $(1 - \rho(\pi))c_p$ at trial, while the defendant would incur $(1 - \rho(\pi))c_d$.[5] We sometimes focus on the special case where the share of investigation costs incurred in discovery rises one-for-one with the extent of discovery, $\rho(\pi) = \pi$, which we refer to as a linear discovery cost function. A strictly convex discovery cost function captures the possibility that discovery focuses on less costly (per unit of information) sources of information first.We discuss the less plausible scenario in which $\rho(\pi)$ is strictly concave in Section **??**. We first take the design of discovery—the probability π that discovery succeeds, and the share $\rho(\pi)$ of investigation costs incurred in discovery—as given, and solve for how the design of discovery affects the timing of when different defendant types make their first offers, and the sizes of these offers. We then characterize the discovery design that minimizes expected total litigation costs of both litigants. It is useful to define $c \equiv c_p + c_d$.

Let d denote the discovery outcome: $d = \theta$ if discovery uncovers θ, and $d = \emptyset$ otherwise. A strategy for the defendant 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.

A strategy for the plaintiff 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.

We let $b_1(x_1)$ represent the plaintiff's beliefs about the defendant's private information after observing x_1; and $b_2(x_1, d, x_2)$ represents the plaintiff's beliefs after observing 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 the belief is 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

[5] In our base model, discovery outcomes do not affect the level of investigation costs that would be incurred were there a trial. This approximates a setting where (1) the total amount of information to be uncovered does not depend on whether discovery succeeds, and (2) whether the defendant's information is uncovered at discovery or at trial is randomly determined. We later show how outcomes are affected when successful discovery reduces trial costs.

(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; 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 possible 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.

3 Separation via offer amount and timing

The 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 $t = 1$, while 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 from the defendant. 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.

Post-discovery settlement offers. First consider a type that did not make a pre-discovery offer ($x_1 = N$). Absent a settlement at $t = 2$ post-discovery, 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 a defendant type θ 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 and 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 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 (**??**) 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

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

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, the plaintiff sometimes rejects 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 would like to reveal the minimal evidence against it in order to avoid having its settlement offers rejected, it has no incentive compatible way to directly do so. Unfortunately for a strong defendant, a weaker defendant always has an incentive to under-report/conceal evidence against it, mimicking a stronger defendant, so that a plaintiff cannot trust selected information provided by a 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 additional trial costs.

Now consider a weak defendant type, $\theta < \hat{\theta}$, whose $t = 1$ pre-discovery offer was rejected, when discovery failed to uncover its private information. On the equilibrium path, its pre-discovery settlement 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, and 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 the defendant will make a take-it-or-leave-it offer that leaves the plaintiff indifferent between accepting and rejecting holding 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 some weak type $\tilde{\theta}$, i.e., if

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

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 settlement 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. Since 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". A stronger future signal makes separation easier by reducing the gains to a weaker defendant type from mimicking a stronger type, thereby reducing the required probability of rejection. Second, the denominator in the exp term for pre-discovery is $\frac{\rho}{\pi}c$, while that for post-discovery is $(1-\rho)c$. These terms represent the cost per unit of revelation probability of the next signal. Pre-discovery, the next "discovery signal" costs ρc and reveals θ with probability π, so the per unit cost of information revelation is $\rho c/\pi$. At $t = 2$, following failed discovery, the next "trial signal" costs $(1-\rho)c$ and it accords in expectation with θ, "on average" revealing the private information, 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 total 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 if it rejects its post-discovery offer.

The cutoff type, $\hat{\theta}(\pi)$. The critical cutoff defendant type $\hat{\theta}$ is indifferent between making an equilibrium offer pre-discovery that reveals its type and waiting until after discovery to make its first offer, where, as the weakest type among those that wait, its offer is always accepted. If $\hat{\theta}$ offers its equilibrium pre-discovery offer, its payoff 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 is small enough that $\pi < 1 - \exp\Big(-\frac{\bar{\theta}-\underline{\theta}}{\frac{\rho(\pi)}{\pi}c}\Big)$, then $\hat{\theta}$ solves $r_1(\hat{\theta}) = 1$:

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

If the extent of discovery is higher, so that $\pi \geq 1 - \exp\Big(-\frac{\bar{\theta}-\underline{\theta}}{\frac{\rho(\pi)}{\pi}c}\Big)$, then no defendant type wants to wait until after discovery to make its first offer.

Define $\bar{\pi} \equiv 1 - \exp\Big(-\frac{\bar{\theta}-\underline{\theta}}{\frac{\rho(\bar{\pi})}{\bar{\pi}}c}\Big)$. Then,

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

Proposition 1. *Fixing 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. The equilibrium rejection probabilities with which the plaintiff rejects a defendant's offers are given by

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

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

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

and equilibrium settlement offers are given by

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

Proof: See the Appendix. □

The separating equilibrium uniquely pins down the size 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 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 **??** illustrates the equilibrium when $\pi \in (0, \bar{\pi})$.

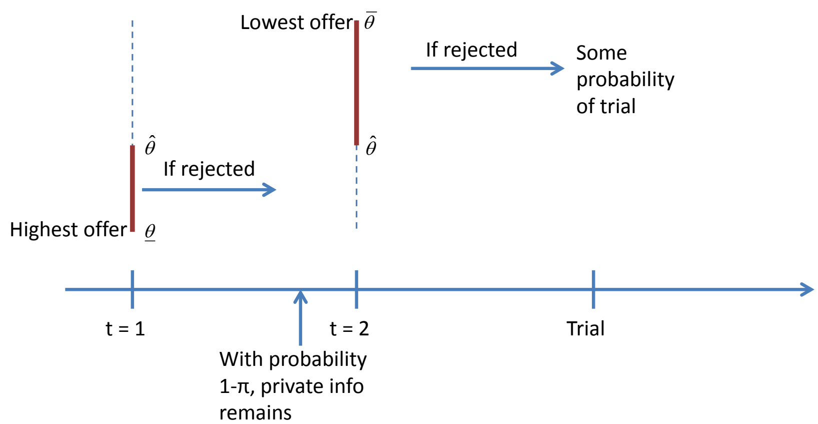

Figure 1: Separating by offer amount and timing under limited discovery $\pi \in (0, \bar{\pi})$.

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 $\pi > \bar{\pi}$, 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.

We write the rejection probabilities as $r_1(\theta, \pi)$ and $r_2(\theta, \pi)$ (see equations (**??**) and (**??**)), to emphasize their dependence on π.

Lemma 1. *The rejection probability $r_1(\hat{\theta}, \pi)$ is strictly decreasing in π.*

Proof: See the Appendix. $\square$

Inspection of equation (**??**) 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 type, which reduces the rejection probability required to induce separation. Second, if the share 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, $\frac{\rho(\pi)}{\pi}$ (by Lemma **??** in the Appendix), harshening the consequences of having offers rejected. In turn, this further reduces the pre-discovery rejection probabilities needed to induce separation. If, instead, the share of total investigation costs incurred in discovery are strictly convex 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 $\frac{1}{\pi}$ when θ is close to $\hat{\theta}$.

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, around $\hat{\theta}$. Therefore, $\hat{\theta}$ 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 $\pi < \bar{\pi}$.*

4 The impact of discovery on expected litigation costs

In this section we derive how the design of discovery affects the total investigation costs that the two parties expect to incur in equilibrium given any 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). Consequently, total litigation costs are the same:

Lemma 2. *Total litigation costs are the same with no discovery ($\pi = 0$) as full discovery ($\pi = 1$).*

Proof: See the Appendix. $\square$

In the remainder of this section, we focus attention on a linear discovery technology, where discovery costs rise one-for-one with the extent of discovery, i.e., $\rho(\pi) = \pi$. We defer analysis of more general discovery technologies to Section **??**.

For any θ, define $\hat{\pi}(\theta)$ to be the cutoff on the extent of discovery that determines whether type θ is a strong defendant who waits until after discovery to make a first settlement offer, or a weak defendant who makes an acceptable settlement offer prior to discovery. That is, $\hat{\pi}(\theta)$ solves $\hat{\theta}(\pi) = \theta$. 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 π is at least $\hat{\pi}(\theta)$. In other words, discovery is extensive relative to a type if and only if it induces the defendant to make an acceptable pre-discovery settlement offer. Otherwise, discovery is limited relative to the (strong) defendant's type.

From Lemma **??**, we have,

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

We break the analysis of the impact of discovery on expected litigation costs incurred according to whether whether the defendant type θ is strong or weak given the extent π of 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 provided discovery is extensive enough to make a type θ defendant weak, costs depend only on the type, and not on the particular extent of discovery:

Proposition 2. *(Weak defendants with linear discovery.) When $\rho(\pi) = \pi$ and the defendant type θ is weak, i.e., $\pi \geq \hat{\pi}(\theta)$, the extent of discovery does not affect expected total litigation costs. Thus, there is no difference between no/full discovery and sufficiently extensive discovery.*

A weak defendant's pre-discovery settlement offer fully reveals its type. As a result, its post-discovery settlement offer is always accepted, so that the case never goes to trial. Without discovery, there is always some chance that the pre-trial settlement offer is rejected and the full trial costs are incurred. Extensive discovery is less costly than the trial. However, since discovery is less costly and less revealing than the trial, the plaintiff must be more likely to reject a pre-discovery settlement offer than a pre-trial offer (when there is no discovery). These two opposing effects cancel out for weak types when $\rho(\pi) = \pi$, so there is no cost difference between no discovery and extensive discovery. We will show in Section **??** that these two effects do not typically cancel: A convex discovery technology favors searching for information in steps to reach the low-hanging fruit first, while a concave discovery technology favors searching for all information in one stage.

Strong defendants. When discovery is limited relative to a defendant's type θ, i.e., when $\pi < \hat{\pi}(\theta)$, 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 additional legal costs are incurred as a settlement is then always reached. 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 $(1-\rho(\pi))c$. Expected total investigation costs with discovery for a strong defendant type are:

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

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

Unlike with weak defendants, limited discovery reduces expected litigation costs with strong defendants even when $\rho(\pi)=\pi$. To see why, we rewrite expected litigation costs with no discovery in a way that is comparable to those with discovery:

$$\begin{aligned}
\frac{1}{c}C^s(\theta,0) &= 1-\exp\Big(-\frac{\theta-\underline{\theta}}{c}\Big)\\
&= 1-\exp\Big(-\frac{\theta-\hat{\theta}}{c}\Big)\exp\Big(-\frac{\hat{\theta}-\underline{\theta}}{c}\Big)\\
&= \rho(\pi)+(1-\rho(\pi))\Big(1-\exp\Big(-\frac{\theta-\hat{\theta}}{c}\Big)\Big).\\
\frac{1}{c}C^s(\theta,\pi) &= \rho(\pi)+(1-\rho(\pi))\Big(1-\exp\Big(-\frac{\theta-\hat{\theta}}{(1-\pi)c}\Big)\Big)(1-\pi).
\end{aligned}$$

The third line follows from substituting for $\exp\big(-\frac{\hat{\theta}-\underline{\theta}}{c}\big)$ using $r_1(\hat{\theta}) = 1$, adding and subtracting $\rho(\pi)$ and re-arranging. The common term $\rho(\pi)$ reflects that (a) strong types wait until after discovery to make settlement offers, so that both litigants have to pay discovery costs, but, (b) post discovery, a settlement offer only has to separate θ from the cutoff type $\hat{\theta}(\pi)$ instead of from the worst type $\underline{\theta}$, which reduces litigation costs. With linear discovery, these two effects offset. As a result, comparisons of expected litigation costs with and without discovery revolve around comparisons of $1-\exp\big(-\frac{\theta-\hat{\theta}}{c}\big)$ with $\big(1-\exp\big(-\frac{\theta-\hat{\theta}}{(1-\pi)c}\big)\big)(1-\pi)$.

To understand these terms, recognize that

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

is 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. Figure **??** shows that this rejection probability rises concavely with the separation distance s.

Figure 2: Discovery reduces post-discovery signaling distortions for strong defendants with a linear discovery technology.

The construction of $\big(1-\exp\big(-\frac{\theta-\hat{\theta}}{(1-\pi)c}\big)\big)(1-\pi)$ 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 that having an offer rejected is less costly—it is as if the distance to separate is inflated by factor $\frac{1}{1-\pi} > 1$. To see how these two effects play out observe that

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

and

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

since $R(0) = 0$. With linear discovery, the strong type's distance of separation with no discovery, $\theta - \hat{\theta}$, is a linear combination of those with 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 implies that expected distortion costs with no discovery exceed those with discovery, i.e.,

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

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. That is, rejection probabilities (proportional to $1 - \exp(\cdot)$) must rise concavely, above linearly, in order to discourage weaker types from mimicking a stronger type's settlement offer and thereby receiving its better payoff. Phrased differently, with separation via public discovery, each defendant makes the settlement associated with its true type. In contrast, with endogenous separation via settlement offers, a defendant that mimics a better type's offer is perceived as a better type even when its offer is rejected, so it must face a higher probability of rejection—its expected litigation expenditures must be higher—to preserve incentive compatible revelation of type.

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

Corollary 3. *When $\rho(\pi) = \pi$, and there is a type θ defendant, discovery benefits litigants (relative to no discovery) if and only if $\pi \in (0, \hat{\pi}(\theta))$.*

We next establish that not only are strong defendants the only ones to gain from discovery, but that among strong defendants, it is the stronger defendants who benefit more from discovery. Let $\Delta^s(\theta,\pi) \equiv C^s(\theta,0) - C^s(\theta,\pi)$ denote the benefit of discovery for strong defendant types. 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 distance to separate s is greater, a marginal increase in s leads to a smaller marginal increase in rejection rates. As a result, a stronger (strong) type gains more from linear discovery. We show in Section **??**, that this effect is amplified by a convex discovery technology, $\rho(\pi)$, but is at least partially offset if $\rho(\pi)$ is concave.

Discovery outcomes do not affect a plaintiff's payoffs—the plaintiff is always indifferent between settling early and having the litigation go all the way to trial. The plaintiff's indifference reflects a defendant's ability to make take-it-or-leave-it offers. Therefore, all gains from reductions in investigation costs due to discovery accrue to the defendant. There is an important caveat to this result: Section **??** shows that if a successful discovery reduces trial costs relative to an unsuccessful one, then a plaintiff gains from discovery because it strengthens her bargaining position. In fact, the plaintiff can gain so much that discovery can actually harm the defendant.

Discovery reduces the likelihood that litigation goes all the way to trial. With discovery, weak defendants never go to trial—even if their pre-discovery offers are rejected, their subsequent 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. When the discovery cost is linear or convex, limited discovery reduces expected total investigation costs relative to no discovery. Since litigants always incur positive discovery costs with such discovery, it follows directly that they must be less likely to incur trial costs. What is more interesting is that, even conditional on discovery *failing* to reveal θ, discovery must reduce the probability of trial for at least some strong types. In particular, for a strong type $\theta \sim \hat{\theta}$, even when discovery fails, the rejection probability needed to separate from $\hat{\theta}$ is close to zero. However, absent discovery, separation 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

Because damages paid by the defendant to the plaintiff represent a transfer, a utilitarian social planner wants to design discovery to minimize the ex ante total expected litigation costs of the two litigants. Since the social planner does not know the defendant's private type, she must integrate over the defendant's possible types. Before we investigate this ex ante optimal level of discovery, we first identify the optimal extent of discovery given some 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 **??** shows that when $\rho(\pi) = \pi$, then $\pi^*(\theta)$ is strictly positive, but does not exceed $\hat{\pi}(\theta)$. But how precisely does $\pi^*(\theta)$ vary with θ?

We next 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 3. *When $\rho(\pi) = \pi$, $\Delta^s_{\pi,\theta}(\theta, \pi^*(\theta)) > 0$.*

Proof: See the Appendix. $\square$

Proposition **??** then follows directly:

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

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

Given a particular type θ, if its pre-trial settlement offer is always accepted, then it is best to have no discovery and let the settlement offer bear all information costs. If its pre-trial settlement offer is always rejected, then it is best to set an intermediate level of discovery of $argmax\{\pi + (1-\pi)(1-\pi)\} = \frac{1}{2}$, to break information acquisition into two equal steps. Since, in equilibrium, pre-trial settlement offer are sometimes, but not always, rejected, the optimal extent of discovery is somewhere between 0 and $\frac{1}{2}$. A stronger type's settlement offer gets rejected more often. This calls for greater discovery, discovery that is closer to $\frac{1}{2}$. This effect on $\pi^*(\theta)$ dominates the countervailing effect that, with stronger types, marginal increases in discovery reduce by more the probability that the settlement offer is rejected, which calls for less discovery.

Optimal π ex ante. The socially 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 π^* follow from those of $\pi^*(\theta)$:

Corollary 5. *When $\rho(\pi) = \pi$,*

$$0 < \pi^* < \bar{\pi}, \textit{ and } \quad 0 < \pi^* < \frac{1}{2}.$$

We next investigate how the distribution of defendant types affects the optimal extent of discovery. To do this, we compare distributions of defendant types that are ordered according to conditional first-order stochastic dominance. A 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 $\mu \in (\underline{\theta}, \bar{\theta})$.[6] We ask: when defendants are more likely to be strong, is it better to have more discovery or less? Lemma **??** shows that increased discovery benefits stronger types by more.

Proposition 6. *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., $\pi_2^* > \pi_1^*$.*

6 General discovery technologies

We next consider more general discovery structures, exploring how the curvature of discovery (convex or concave) affects expected litigation costs and the optimal design of discovery. In practice, a judge has substantial discretion over the directions that he allows discovery to go. The question becomes—when taking into account the strategic incentives of the plaintiff and defendant, should the court direct discovery toward lower or higher cost sources of information?

If the court dictates that discovery starts with more promising sources of information (per dollar of discovery costs), then the marginal return to discovery declines with the extent of discovery. That is, the cost of discovery $\rho(\pi)$ is convex in π. In this case, we have

Proposition 7. *(Strictly convex $\rho(\pi)$.) Expected total litigation costs are strictly lower with any limited discovery π than with no discovery. For any 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 a strong defendant who waits until after discovery to make a first acceptable settlement offer. Stronger types benefit more from any given amount of discovery.*

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

Proof: See the Appendix. □

Our qualitative findings with linear discovery, where discovery costs rise one-for-one with the extent of discovery $\rho(\pi)$, are reinforced when discovery first targets less costly sources of information. When $\rho(\pi)$ is strictly convex, even weak defendants benefit from discovery. This is because the two forces in Proposition **??** no longer offset: discovery now brings a benefit because the efficiency of early discovery allows weak defendants to avoid trial costs without sacrificing as much of the signaling power of the pre-discovery settlement offer. For strong defendant types, convex discovery reduces the distortion in the signaling following the discovery by even more than with linear discovery: $\pi > \rho(\pi)$ implies that relatively more weight is put on the event that discovery succeeds, obviating the need to separate post-discovery via a settlement offer.

If, instead, discovery focuses on inefficient sources of information, then the marginal return of discovery *rises* with the level of discovery. That is, the cost of discovery $\rho(\pi)$ is a convex function of π. Then we have:

Proposition 8. *(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 technology $\rho(\cdot)$, for any positive 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. □

With concave discovery costs, the two forces in Proposition **??** do not offset: discovery now increases litigation costs because the *inefficiency* of early discovery means that the discovery reduces the signaling power of the pre-discovery settlement offer without providing litigants enough savings on trials costs. Even strong defendant types may no longer benefit from discovery: concave discovery may not reduce the distortion in signaling post discovery because $\pi < \rho(\pi)$ implies that relatively more weight is put on the event that discovery fails.

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

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

Proof: See the Appendix. $\square$

Figure **??** illustrates how expected litigation costs vary with the extent of discovery π and the convexity z of the discovery technology.

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

The results in this section reveal that a judge should actively direct discoveryfirst toward more cost-effective sources. Further, the judge should 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.[7]

7 The model

Reduction in trial costs due to successful discovery. We have supposed that the success or failure of discovery to uncover a defendant's private information does not affect the level of investigation costs incurred at trial. This is consistent with a scenario in which the total amount of information to be uncovered in the legal process—both the private information that the defendant knows and the information that neither party knows—is unaffected by the timing of when each type of information is actually uncovered. Still, it is important to understand how outcomes are affected by an informational cost structure in which successful discovery that uncovers a defendant's private information *reduces* the information costs that must be incurred at trial.

To highlight the qualitative consequences, consider the extreme scenario in which successful discovery eliminates *all* trial costs. One might conjecture that such cost savings would not affect expected litigation costs relative to the base model precisely because successful discovery *always* induces the parties to settle prior to trial. That is, in the equilibrium of our base model, there is

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

never a trial following successful discovery, so one might conjecture that when successful discovery eliminates trial costs, this just results in a transfer from the defendant to the plaintiff via a higher settlement, and does not affect total investigation costs. While this reasoning about the transfer is correct, it turns out that the costs avoided by successful discovery affect not only division of the cost savings of discovery between the defendant and plaintiff, but also the expected cost savings. As we now show, the impact is a subtle one via the pre-discovery separation of different defendant types.

If a plaintiff accepts a pre-discovery offer x_1, she gets x_1. If she rejects it, they go to discovery. If discovery succeeds, they settle with an offer of $1-\theta$, since successful discovery reduces trial costs to zero (in this extreme scenario). If discovery fails, they settle at $(1-\theta)-(1-\rho)c_p$. Therefore, the pre-discovery offer that leaves the plaintiff indifferent between accepting and rejecting it under separating beliefs is

$$x_1(\theta) = \pi(1-\theta) + (1-\pi)((1-\theta)-(1-\rho)c_p) - \rho c_p = (1-\theta) - c_p + \pi(1-\rho)c_p.$$

Thus, when successful discovery reduces trial costs, it causes the defendant to lose $\pi(1-\rho)c_p$, i.e., the probability of successful discovery times the lost trial costs that the defendant can no longer threaten a plaintiff with to extract a reduced settlement in its post-discovery offer.

Proposition 10. *Suppose that successful discovery obviates the need to incur investigation costs at trial. Then, for any level of discovery $\pi \in (0,1)$, expected total investigation costs incurred with each defendant type θ are reduced.*

Proof: See the Appendix. □

The intuition is that the reduction in trial costs reduces the surplus a defendant can extract following successful discovery. This raises the cost of mimicking a better type's pre-discovery offer because such mimicking increases 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. In turn, more defendant types make pre-discovery offers, making separation of stronger types easier.

When successful discovery did not lower trial costs, plaintiffs were indifferent to the extent of discovery—the defendant extracted all of the cost savings generated by discovery. However, when successful discovery reduces investigation costs at trial, plaintiffs benefit from a defendant's reduced ability to extract a lower settlement via a threat to go to trial. Indeed, plaintiffs may gain at the defendant's expense:

Proposition 11. *Suppose that successful discovery eliminates trial investigation costs and that $\rho(\pi) = \pi$. Then, weak defendant types that make pre-discovery offers are hurt by positive discovery, as are some strong defendant types that wait until after discovery to make their first offers: either all strong types are hurt or there exists a $\tilde{\theta} > \hat{\theta}$ such that only sufficiently strong defendant types $\theta > \tilde{\theta}$ benefit from discovery for $\pi \in (0, 1)$.*

Proof: See the Appendix. $\square$

If, instead, the design of discovery targets lower marginal cost sources of information, then the elimination of trial costs by successful discovery hurts fewer defendant types—they gain from the eased separation from weak types facilitated by the convex discovery technology. Indeed, when $\rho(\pi)$ is sufficiently convex *and* the plaintiff's share of investigation costs $\frac{c_p}{c_p+c_d}$ is small enough, *all* defendant types benefit from discovery even if successful discovery eliminates all trial costs. This reflects that when $\frac{c_p}{c_p+c_d}$ is small, a defendant cares more about the reduction in the probability of incurring costs at trial, and less about the (modest) reduction in its threat following successful discovery.

Fixed trial costs. We have assumed that trial costs solely reflect information discovery. In practice, trials may have a fixed procedural cost component, unrelated to the presentation of evidence. We now show that positive fixed procedural trial costs do not alter our qualitative findings, serving only to raise the attraction of more extensive discovery, which reduces the likelihood that the parties go to trial. Denote these fixed procedural costs by k_p and k_d for the plaintiff and the defendant respectively, and let $k = k_p + k_d$ and $c = c_p + c_d$.

Proposition 12. *Suppose that $k > 0$ and $\rho(\pi) = \pi$. Then expected total litigation costs are lower with discovery than without ($\pi = 0$). They are lower with extensive discovery ($\pi \geq \bar{\pi}$) than limited discovery ($\pi < \bar{\pi}$) if and only if $k > \bar{\pi}c$. Moreover, the optimal extent of discovery is increased by positive fixed procedural trial costs.*

Proof: See the Appendix. $\square$

When $\rho(\pi) = \pi$, all levels of extensive discovery ($\pi > \bar{\pi}$) result in the same expected litigation costs. Positive fixed trial costs $k > 0$ favor extensive discovery that encourages all defendant types to settle before a trial, and with large enough fixed trial costs, extensive discovery becomes optimal.

Multiple rounds of discovery. Our base model feature a single round of discovery. We now show 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.

To illustrate the qualitative effects, we consider two rounds of discovery, where the extent of discovery in round $i = 1, 2$ is $\pi_i^2 > 0$, and the associated investigation cost share technology is linear, i.e., $\rho(\pi_i^2) = \pi_i^2$. We focus on the interesting case where the total amount of discovery, $\pi_1^2 + \pi_2^2 \equiv \pi$ is less than $\bar{\pi}$, so that not all types settle prior to the second round of discovery.

To highlight how the results extend to more rounds of (properly designed) discovery, we use notation that accommodates N rounds of discovery: let $\hat{\theta}_i^N$ denote the defendant type that is indifferent between making an offer just before round-i discovery, and waiting until after round-i discovery to make a first offer whenever prior discoveries fail, where we adopt the convention that $\hat{\theta}_0^N = \underline{\theta}$ and $\hat{\theta}_{N+1}^N = \bar{\theta}$; and let $r_i^N(\theta)$ be the equilibrium probability of rejection when prior discoveries fail, for types prepared to make offers just before round-i discovery, i.e., $\theta \in [\hat{\theta}_{i-1}^N, \hat{\theta}_i^N)$ for $i = 1, \ldots, N+1$.

Proposition 13. *Expected total investigation costs incurred under two rounds of discovery with $\pi_1^2 + \pi_2^2 = \pi$ are less than those incurred under one round of discovery with π.*

Proof: See the Appendix. $\square$.

This result is subtle. We first prove that $\hat{\theta}_1^1 > \hat{\theta}_2^2 > \hat{\theta}_1^2$: with two rounds of discovery, *more* defendant types are prepared to wait until after the end of *all* discovery to make their first offers. Intuitively, when the first round of discovery fails, the second round becomes more cost effective, revealing θ with probability $\frac{\pi_2^2}{1-\pi_1^2}$ at a cost of π_2^2. This less costly signal makes separation harder, reducing θ_2^2. That is, dividing discovery into more rounds induces *more* defendant types to wait until just before the trial to make their first offer (unless discovery succeeds). Nonetheless, expected total investigation costs are reduced when discovery is divided into two rounds for all strong defendant types $\theta > \hat{\theta}_1^2$, while expected investigation costs are the same for weak defendant types

$\theta < \hat{\theta}_1^2$ (since $\theta < \hat{\theta}_1^2$ must separate away from weaker types that are, in both instances, making offers immediately—their incentives are unaffected by stronger types $\theta > \hat{\theta}_1^2$).

The intuition for why, with stronger defendants, two rounds of discovery reduce expected investigation costs 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 ($\exp(\cdot)$ is convex) 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 further division of discovery rounds optimal.

This intuition extends: one can always reduce expected investigation costs by further dividing discovery into more rounds, i.e., so that there are $N + 1$ rounds of discovery rather than N rounds. Most obviously, the above analysis implies that dividing the "last" round of discovery into two rounds reduces expected total investigation costs.

8 Conclusion

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

We show that the process of discovery provides defendants an additional channel with which to signal—the timing of its initial settlement offer. With one round of discovery, weaker defendants make their initial offers prior to discovery, while stronger defendants wait until after limited discovery to do so. 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 defendant types. As a result, conventional wisdom about the impact of discovery on a privately-informed party's welfare is overturned: privately-informed defendants gain from *limited* discovery, with stronger defendants gaining more.

We continue to 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. We further show that when discovery targets low marginal cost sources of information first, leaving less cost-effective sources for trial, expected investigation costs are reduced by more, especially for stronger defendants. We conclude by establishing the ways in which our qualitative findings extend in settings with multiple rounds of discovery, fixed procedural trial costs, or when successful discovery reduces the investigation expenses that must be incurred at trial.

References

[1] Cooter, R. and D. Rubinfeld. 1994. An Economic Model of Legal Discovery. Journal of Legal Studies, 23:435-463.

[2] Farmer, A. and P. Pecorino. 2005. Civil Litigation with Mandatory Discovery and Voluntary Transmission of Private Information. Journal of Legal Studies, 35:137-159.

[3] Hay, B. 1994. Civil Discovery: Its Effects and Optimal Scope. Journal of Legal Studies, 23:481-415.

[4] Reinganum, J. and L. Wilde. 1986. Settlement, Litigation, and the Allocation of Litigation Costs. RAND Journal of Economics, 17:557-566.

[5] Schrag, J. 1999. Managerial Judges: An Economic Analysis of the Judicial Management of Legal Discovery. RAND Journal of Economics, 30:305-323.

[6] Schwartz, W. and A. Wickelgren. 2009. Credible Discovery, Settlement, and Negative Expected Value Suits. RAND Journal of Economics, 40: 636-657.

[7] Shavell, Steven. 1989. Sharing of Information Prior to Settlement or Litigation. RAND Journal of Economics, 20:183-195.

[8] Sobel, J. 1989. An Analysis of Discovery Rules. Law and Contemporary Problems, 52:133-159.

9 Appendix

Proof of Proposition ??. First, we prove that if $\pi < \bar{\pi}$, then $r_1(\bar{\theta}) > 1$; and if $\pi > \bar{\pi}$, then $r_1(\bar{\theta}) < 1$. Since $\bar{\pi}$ is defined from $r_1(\bar{\theta}) = 1$, it is sufficient to show that $r_1(\bar{\theta})$ is strictly decreasing in π. Clearly, $\frac{1}{\pi}$ strictly decreases in π. Further, $1 - \exp\Big(-\frac{\bar{\theta}-\underline{\theta}}{\frac{\rho(\pi)}{\pi}(c_p+c_d)}\Big)$ decreases in $\frac{\rho(\pi)}{\pi}$, which, in turn, increases (at least weakly) in π. Therefore, $r_1(\bar{\theta})$ is strictly decreasing in π. Since when $\pi = 1$, $r_1(\bar{\theta}) < 1$, it must be that $\bar{\pi} < 1$.

Existence: We complete the description of the equilibrium by specifying 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 $x_1 > \bar{x}_1$, $b_1^*(x_1) = \underline{\theta}$. Since 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 $x_1 < \underline{x}_1$, $b_1^*(x_1) = \hat{\theta}$. Since 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

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

Since

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

the second-order derivative evaluated at $x_1^*(\theta) = (1-\theta) - c_p$ is $p_1'(x_1)[2\pi - \rho(\pi)] < 0$ as $\rho(\pi) \leq \pi$. That is, $x_1^*(\theta)$ is a local strict maximizer. Since 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 in the event of discovery signaling being uninformative, 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 starting at $t = 2$ given that discovery has failed for those θ that did not make pre-discovery offers that revealed their types. Then by Reingannum and Wilde (1986), these types must separate in this subgame.

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

$t = 2$, and let θ_{x_2} be an associated element.

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 and 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 (see Reinganum and Wilde (1986) for details).

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 ??. Since the defendant type is strong, $\frac{\Delta^s(\theta,\pi)}{c_p+c_d} = (\pi + (1-\pi)\exp\Big(-\frac{\theta-\hat{\theta}}{(1-\pi)(c_p+c_d)}\Big) - \exp\Big(-\frac{\theta-\hat{\theta}}{c_p+c_d}\Big))(1-\pi)$. The first-order derivative Δ^s_π has the same sign as:

$$\begin{aligned}
&1 - 2\pi - 2(1-\pi)\exp\Big(-\frac{\theta-\hat{\theta}}{(1-\pi)(c_p+c_d)}\Big) \\
&\quad + (1-\pi)\exp\Big(-\frac{\theta-\hat{\theta}}{(1-\pi)(c_p+c_d)}\Big)\Big(\frac{1}{\exp\Big(-\frac{\theta-\hat{\theta}}{(1-\pi)(c_p+c_d)}\Big)(1-\pi)} - \frac{\theta-\hat{\theta}}{(1-\pi)^2(c_p+c_d)}\Big).
\end{aligned}$$

Define $\delta \equiv \frac{\theta-\hat{\theta}}{(1-\pi)(c_p+c_d)}$ within this proof. Then, differentiating the expression above with respective to θ, $\Delta^s_{\theta,\pi}$ must have the same sign as:

$$[(1-\pi)\exp(-\delta)(-\frac{1}{(1-\pi)(c_p+c_d)})](-2+\frac{1}{\exp(-\delta)(1-\pi)} - \frac{\theta-\hat{\theta}}{(1-\pi)^2(c_p+c_d)} - \frac{1}{(\exp(-\delta))^2(1-\pi)}) > 0. \quad \square$$

Proof of Proposition ??: Recall that $\hat{\theta}(\pi)$ is implicitly defined by $\frac{1}{\pi}[1 - \exp\Big(-\frac{\hat{\theta}(\pi)-\underline{\theta}}{\frac{\rho(\pi)}{\pi}(c_p+c_d)}\Big)] = 1$, so that

$$\exp\Big(-\frac{\hat{\theta}(\pi) - \underline{\theta}}{\frac{\rho(\pi)}{\pi}(c_p + c_d)}\Big) = 1 - \pi. \tag{4}$$

Next, decompose the plaintiff's probability of acceptance under no discovery as

$$\exp\Big(-\frac{\theta - \underline{\theta}}{c_p + c_d}\Big) = \exp\Big(-\frac{\theta - \hat{\theta}(\pi)}{c_p + c_d}\Big)\exp\Big(-\frac{\hat{\theta}(\pi) - \underline{\theta}}{c_p + c_d}\Big). \tag{5}$$

When discovery costs $\rho(\pi)$ are linear (so $\frac{\rho(\pi)}{\pi} = 1$), substitute using (**??**) to simplify equation (**??**) to

$$\exp\Big(-\frac{\theta - \underline{\theta}}{c_p + c_d}\Big) = \exp\Big(-\frac{\theta - \hat{\theta}(\pi)}{c_p + c_d}\Big)(1 - \pi).$$

Hence, the cost advantage of discovery is:

$$\frac{\Delta^s(\theta, \pi)}{c_p + c_d} = (1 - \pi)[\pi + (1 - \pi)\exp\Big(-\frac{\theta - \hat{\theta}(\pi)}{(1 - \pi)(c_p + c_d)}\Big) - \exp\Big(-\frac{\theta - \hat{\theta}(\pi)}{c_p + c_d}\Big)].$$

Since $\exp(\cdot)$ is convex, Jensen's Inequality reveals that the cost savings associated with discovery are strictly positive for all levels of discovery $\pi \in (0, \bar{\pi})$:

$$\pi + (1 - \pi)\exp\Big(-\frac{\theta - \hat{\theta}(\pi)}{(1 - \pi)(c_p + c_d)}\Big) - \exp\Big(-\frac{\theta - \hat{\theta}(\pi)}{c_p + c_d}\Big) > 0 \;\Rightarrow\; \frac{\Delta^s(\theta, \pi)}{c_p + c_d} > 0.$$

If the information costs of discovery $\rho(\pi)$ are strictly convex in the extent of discovery, the reduction in investigation costs associated with discovery is reinforced. To see this, observe that since $\frac{\rho(\pi)}{\pi} < 1$, the convexity of $\exp(\cdot)$ implies that for $\pi \in (0, \bar{\pi})$,

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

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

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

the left-hand side of inequality (**??**) simplifies to $1 - \rho(\pi)$. Therefore,

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

Multiplying both sides by $\exp\Big(-\frac{\theta-\hat{\theta}(\pi)}{c_p+c_d}\Big)$ yields

$$(1-\rho(\pi))\exp\Big(-\frac{\theta-\hat{\theta}(\pi)}{c_p+c_d}\Big) > \exp\Big(-\frac{\theta-\underline{\theta}}{c_p+c_d}\Big). \tag{7}$$

Comparing (**??**) with:

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

This reveals that

$$\frac{\Delta^s(\theta,\pi)}{c_p+c_d} > (1-\rho(\pi))[\pi+(1-\pi)\exp\Big(-\frac{\theta-\hat{\theta}(\pi)}{(1-\rho(\pi))(c_p+c_d)}\Big) - \exp\Big(-\frac{\theta-\hat{\theta}(\pi)}{c_p+c_d}\Big)].$$

Substituting $\rho(\pi)<\pi$ for π on the right-hand side yields

$$\frac{\Delta^s(\theta,\pi)}{c_p+c_d} > (1-\rho(\pi))[\rho(\pi)+(1-\rho(\pi))\exp\Big(-\frac{\theta-\hat{\theta}(\pi)}{(1-\rho(\pi))(c_p+c_d)}\Big) - \exp\Big(-\frac{\theta-\hat{\theta}(\pi)}{c_p+c_d}\Big)] > 0,$$

where the final inequality follows from the convexity of $\exp(\cdot)$ and Jensen's Inequality. □

Proof of Proposition ??: Let $p_1(x_1)$ be the probability that offer x_1 is rejected. The payoff of a type θ is:

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

Each θ chooses $x_1(\theta)$ to maximize this payoff. The associated first-order condition is:

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

Substituting $x_1(\theta)=(1-\theta)-c_p+\pi(1-\rho)c_p$ yields:

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

Coupled with the boundary condition for $\underline{\theta}$, this implies that:

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

Let $\dot{r_1}(\theta)$ denote the equilibrium probability that a pre-discovery offer by $\theta\in[\underline{\theta},\dot{\theta}(\pi)]$ is rejected when successful discovery eliminates trial costs:

$$\dot{r_1}(\theta)=\frac{1}{\pi}[1-\exp\Big(-\frac{\theta-\underline{\theta}}{\frac{\rho}{\pi}(c_p+c_d)+(1-\pi)(1-\rho)c_p}\Big)] < \frac{1}{\pi}[1-\exp\Big(-\frac{\theta-\underline{\theta}}{\frac{\rho}{\pi}(c_p+c_d)}\Big)] = r_1(\theta).$$

Comparing $\dot{r_1}$ with r_1 reveals that the probability of rejection is lower when successful discovery reduces trial costs. This lower probability of rejection implies that the cost advantage of discovery for a weak defendant type is raised: even with linear discovery costs, total investigation costs are reduced by discovery for weak defendant types. Moreover, the cutoff type rises: $\dot{\theta}(\pi) > \hat{\theta}(\pi)$. The trial costs avoided by successful discovery affect the cost advantage of strong types only via the cutoff type. The proof for Proposition **??** extends when we replace $\exp\left(-\frac{\theta-\underline{\theta}}{c_p+c_d}\right) = \exp\left(-\frac{\hat{\theta}-\underline{\theta}}{c_p+c_d}\right)\exp\left(-\frac{\theta-\hat{\theta}}{c_p+c_d}\right)$ with $\exp\left(-\frac{\theta-\underline{\theta}}{c_p+c_d}\right) < \exp\left(-\frac{\dot{\theta}-\underline{\theta}}{c_p+c_d+(1-\pi)^2c_p}\right)\exp\left(-\frac{\theta-\dot{\theta}}{c_p+c_d}\right)$. Thus, any reduction in trial costs due to successful discovery raises the cost advantage of discovery for all types (via the indifference of the cutoff type). $\square$

Proof of Proposition ??:

1. Weak Defendant. The combined total cost savings from discovery for both plaintiff and defendant with a weak defendant θ is:

$$\Delta^{wp} + \Delta^{wd} = [\exp\Big(-\frac{\theta-\underline{\theta}}{(c_p+c_d)+(1-\pi)^2c_p}\Big) - \exp\Big(-\frac{\theta-\underline{\theta}}{c_p+c_d}\Big)](c_p+c_d).$$

where Δ^{wp} denotes the cost savings for the plaintiff and Δ^{wd} denotes those for the defendant when the defendant is weak.

The gain from discovery for the plaintiff is:

$$\Delta^{wp} = \pi(1-\pi)c_p.$$

Note that $\Delta^{wp} + \Delta^{wd}$ increases in θ. That is, among weak defendants, the total gain is highest for the cutoff type, $\dot{\theta}$. Therefore, the total gain is bounded from above by:

$$[\exp\Big(-\frac{\dot{\theta}-\underline{\theta}}{(c_p+c_d)+(1-\pi)^2c_p}\Big) - \exp\Big(-\frac{\dot{\theta}-\underline{\theta}}{c_p+c_d}\Big)](c_p+c_d) = [(1-\pi) - \exp\Big(-\frac{\dot{\theta}-\underline{\theta}}{c_p+c_d}\Big)](c_p+c_d).$$

By Jensen's inequality:

$$\begin{aligned}\frac{c_p+c_d}{c_p+c_d+(1-\pi)^2c_p}\exp\Big(-\frac{\dot{\theta}-\underline{\theta}}{c_p+c_d}\Big) + (1 - \frac{c_p+c_d}{c_p+c_d+(1-\pi)^2c_p})\exp(0)& \\ > \exp\Big(-\frac{\dot{\theta}-\underline{\theta}}{(c_p+c_d)+(1-\pi)^2c_p}\Big) &= 1-\pi.\end{aligned}$$

Therefore,

$$\exp\Big(-\frac{\dot{\theta}-\underline{\theta}}{c_p+c_d}\Big) > \frac{c_p+c_d+(1-\pi)^2c_p}{c_p+c_d}(1-\pi) - \frac{(1-\pi)^2c_p}{c_p+c_d}.$$

This implies that $\Delta^{wp} + \Delta^{wd}$ is further bounded above by:

$$\begin{aligned}\Delta^{wp} + \Delta^{wd} &< [(1-\pi) - \frac{c_p + c_d + (1-\pi)^2 c_p}{c_p + c_d}(1-\pi) + \frac{(1-\pi)^2 c_p}{c_p + c_d}](c_p + c_d) \\ &= (1-\pi)^2 c_p - (1-\pi)^3 c_p = \pi(1-\pi)^2 c_p < \pi(1-\pi)c_p = \Delta^{wp}.\end{aligned}$$

Therefore, $\Delta^{wd} < 0$ for any weak $\theta \leq \dot{\theta}$.

2. Strong defendant. The combined total cost savings for both the plaintiff and the defendant with a strong defendant θ is:

$$\Delta^{sp} + \Delta^{sd} = [(\pi + (1-\pi)\exp\Big(-\frac{\theta - \dot{\theta}}{(1-\pi)(c_p + c_d)}\Big))(1-\pi) - \exp\Big(-\frac{\theta - \underline{\theta}}{c_p + c_d}\Big)](c_p + c_d). \quad (8)$$

The first-order derivative of these cost savings with respect to θ has the same sign as:

$$-(1-\pi)\exp\Big(-\frac{\theta - \dot{\theta}}{(1-\pi)(c_p + c_d)}\Big) + \exp\Big(-\frac{\theta - \underline{\theta}}{c_p + c_d}\Big).$$

The second-order derivative with respect to θ has the same sign as:

$$\exp\Big(-\frac{\theta - \dot{\theta}}{(1-\pi)(c_p + c_d)}\Big) + \exp\Big(-\frac{\theta - \underline{\theta}}{c_p + c_d}\Big) > 0.$$

Therefore, the cost savings are strictly convex in θ.

At $\theta = \dot{\theta}$, the first-order derivative with respect to θ is:

$$-(1-\pi) + \exp\Big(-\frac{\dot{\theta} - \underline{\theta}}{c_p + c_d}\Big) < 0,$$

since, by the definition of $\dot{\theta}$, $\exp\Big(-\frac{\dot{\theta} - \underline{\theta}}{(c_p + c_d) + (1-\pi)^2 c_p}\Big) = 1 - \pi$. That is, the total savings first decrease in θ. Since these savings are strictly convex, for large θ, the cost savings may eventually begin to increase in θ.

We have shown that for all $\theta < \dot{\theta}$ total savings are less than $\pi(1-\pi)c_p$, and for $\theta > \dot{\theta}$, total savings can exceed $\pi(1-\pi)c_p$ only for strongest θ for which total savings increase in θ.

Thus, if the total savings at $\theta = \bar{\theta}$ are less than $\pi(1-\pi)c_p$, then for all θ, total savings are less than $\pi(1-\pi)c_p$ (which implies $\Delta^{sd} < 0$); and if total savings at $\theta = \bar{\theta}$ exceed $\pi(1-\pi)c_p$ (which implies $\Delta^{sd} \geq 0$), then there exists $\tilde{\theta} \in (\dot{\theta}, \bar{\theta}]$ such that for all $\theta < \tilde{\theta}$, the total savings are less than $\pi(1-\pi)c_p$ and for all $\theta \geq \tilde{\theta}$, total savings exceed $\pi(1-\pi)c_p$. $\square$

Proof of Proposition ??. No discovery ($\pi = 0$) is now worse than extensive discovery ($\pi > \bar{\pi}$). With extensive discovery, all types are weak, while with no discovery, all types are strong. With extensive discovery, the cost saving from a settlement (relative to going to a trial) for a type θ is $(c + k) - (1 - \exp\Big(-\frac{\theta - \underline{\underline{\theta}}}{c}\Big))c = k + c\exp\Big(-\frac{\theta - \underline{\underline{\theta}}}{c}\Big)$. With no discovery, the cost saving is $(c + k)\exp\Big(-\frac{\theta - \underline{\underline{\theta}}}{c+k}\Big)$. Writing $k = k\ exp(0)$, and using the convexity of $\exp(\cdot)$, we have

$$k + c\exp\Big(-\frac{\theta - \underline{\underline{\theta}}}{c}\Big) = (c+k)[\frac{k}{c+k}\exp(0) + \frac{c}{c+k}\exp\Big(-\frac{\theta - \underline{\underline{\theta}}}{c}\Big)] > (c+k)\exp\Big(-\frac{\theta - \underline{\underline{\theta}}}{c+k}\Big).$$

Therefore, high discovery yields greater cost savings than no discovery.

Next compare limited positive discovery ($\pi < \bar{\pi}$) with no discovery. Because weak defendants never go to trial under discovery, fixed trial costs k reinforce the advantage of discovery over no discovery in terms of lower litigation costs. With strong defendants, the expected cost savings relative to going to trial is:

$$\begin{aligned}
&((1-\pi)c + k)[\pi\exp(0) + (1-\pi)\exp\Big(-\frac{\theta - \hat{\theta}}{(1-\pi)c + k}\Big)] \\
&> ((1-\pi)c + k)[\pi\exp(0) + \frac{(1-\pi)c + k}{c+k}\exp\Big(-\frac{\theta - \hat{\theta}}{(1-\pi)c + k}\Big)] \\
&> ((1-\pi)c + k)\exp\Big(-\frac{\theta - \hat{\theta}}{c+k}\Big).
\end{aligned}$$

Under no discovery, a settlement saves $c + k$. The expected saving in litigation costs is:

$$\begin{aligned}
(c+k)\exp\Big(-\frac{\theta - \underline{\underline{\theta}}}{c+k}\Big) &= (c+k)\exp\Big(-\frac{\hat{\theta} - \underline{\underline{\theta}}}{c+k}\Big)\exp\Big(-\frac{\theta - \hat{\theta}}{c+k}\Big) \\
&< (c+k)\frac{(1-\pi)c + k}{c+k}\exp\Big(-\frac{\theta - \hat{\theta}}{c+k}\Big) \\
&= ((1-\pi)c + k)\exp\Big(-\frac{\theta - \hat{\theta}}{c+k}\Big),
\end{aligned}$$

where the inequality follows from

$$\exp\Big(-\frac{\hat{\theta} - \underline{\underline{\theta}}}{c+k}\Big) < \frac{c}{c+k}\exp\Big(-\frac{\hat{\theta} - \underline{\underline{\theta}}}{c}\Big) + \frac{k}{c+k}\exp(0) = \frac{c}{c+k}(1-\pi) + \frac{k}{c+k} = \frac{(1-\pi)c + k}{c+k}.$$

Thus, cost savings with discovery are higher than with no discovery for strong types, too. In sum, fixed trial costs reinforce the cost reductions associated with positive discovery.

Finally, compare limited positive discovery ($\pi < \bar{\pi}$) with high discovery ($\pi > \bar{\pi}$). The expected cost saving under positive discovery (relative to going to trial) is:

$$((1-\pi)c+k)[\pi\exp(0)+(1-\pi)\exp\Big(-\frac{\theta-\hat{\theta}}{(1-\pi)c+k}\Big)].$$

The cost saving under high discovery can be written as:

$$((1-\pi)c+k)[\frac{k}{(1-\pi)c+k}\exp(0)+\frac{c(1-\pi)}{(1-\pi)c+k}\exp\Big(-\frac{\theta-\hat{\theta}}{c}\Big)].$$

Note that $(1-\pi)\Big(-\frac{\theta-\hat{\theta}}{(1-\pi)c+k}\Big)=\frac{c(1-\pi)}{(1-\pi)c+k}\Big(-\frac{\theta-\hat{\theta}}{c}\Big)$. Thus,

- If $k < \pi c$ then $c > (1-\pi)c+k \Rightarrow \frac{\theta-\hat{\theta}}{c} < \frac{\theta-\hat{\theta}}{(1-\pi)c+k}$. Therefore, the convexity of $\exp(\cdot)$ implies that cost savings are higher with limited discovery, $\pi < \bar{\pi}$.
- If, instead, $k > \pi c$ then $c < (1-\pi)c+k \Rightarrow \frac{\theta-\hat{\theta}}{c} > \frac{\theta-\hat{\theta}}{(1-\pi)c+k}$. Therefore, the convexity of $\exp(\cdot)$ implies that the cost savings are higher with high discovery, $\pi > \bar{\pi}$.

Since cost savings for weak types $\theta < \hat{\theta}$ do not vary with π when $\rho(\pi)=\pi$, expected cost savings reflect those for strong types. Thus, cost savings are strictly higher with limited discovery $\pi < \bar{\pi}$ than with high discovery, $\pi > \bar{\pi}$ if and only if $k < \bar{\pi}c$.

We now investigate how the optimal extent of discovery π^* varies with k. Consider $k < \pi c$ so that the optimal extent of discovery is positive, but less than $\bar{\pi}$. This implies that π^* solves the first-order condition:

$$\frac{\partial}{\partial\pi}[\int_{\underline{\theta}}^{\hat{\theta}(\pi)}\Delta^w(\theta,\pi)dF(\theta)+\int_{\hat{\theta}(\pi)}^{\bar{\theta}}\Delta^s(\theta,\pi)dF(\theta)]=0,$$

where,

$$\Delta^w(\theta,\pi)=(1-\exp\Big(-\frac{\theta-\underline{\theta}}{c}\Big))(c+k)-\frac{\rho(\pi)}{\pi}(1-\exp\Big(-\frac{\theta-\underline{\theta}}{\frac{\rho(\pi)}{\pi}(c)}\Big))c$$

and

$$\Delta^s(\theta,\pi)=((1-\pi)c+k)[\pi\exp(0)+(1-\pi)\exp\Big(-\frac{\theta-\hat{\theta}}{(1-\pi)c+k}\Big)]-(c+k)\exp\Big(-\frac{\theta-\underline{\theta}}{c+k}\Big).$$

Since $\Delta^w(\hat{\theta}, \pi) = \Delta^s(\hat{\theta}, \pi)$, the first-order condition implies:

$$f(\pi, k) \equiv \int_{\underline{\theta}}^{\hat{\theta}(\pi)} \Delta^w_\pi(\theta, \pi) dF(\theta) + \int_{\hat{\theta}(\pi)}^{\bar{\theta}} \Delta^s_\pi(\theta, \pi) dF(\theta) \Rightarrow f(\pi^*(k), k) = 0$$

Note that $\hat{\theta}$ is independent of k and $\Delta^w_{\pi,k}(\theta, \pi) = 0$,

$$f_k(\pi, k) = \int_{\hat{\theta}(\pi)}^{\bar{\theta}} \Delta^s_{\pi,k}(\theta, \pi) dF(\theta).$$

Algebra gives:

$$\Delta^s_{\pi,k}(\theta, \pi) = 1 - \exp(-\delta) - \left(1 + \frac{(1-\pi)c}{(1-\pi)c + k}\delta\right) \exp(-\delta)\delta$$

where $\delta = \frac{\theta - \hat{\theta}}{(1-\pi)c+k}$.

Note that $g(\delta) \equiv 1 - \exp(-\delta) - \left(1 + \frac{(1-\pi)c}{(1-\pi)c+k}\delta\right) \exp(-\delta)\delta$ has a first order derivative of $g'(\delta) = (1 - \frac{(1-\pi)c}{(1-\pi)c+k})\delta + \frac{(1-\pi)c}{(1-\pi)c+k}\delta^2 > 0$ for $\delta > 0$ and $g(0) = 0$. This proves that $g(\delta) > 0$ for any $\delta > 0$. That is, for any $\theta > \hat{\theta}$,

$$\Delta^s_{\pi,k}(\theta, \pi) > 0.$$

In other words, $f_k(\pi, k) > 0$. Since $f(\pi^*(k), k) = 0$, by the Implicit Function Theorem and concavity of the objective function, $\pi^*(k)$ is increasing in k. This implies that the optimal extent of discovery is higher with positive fixed trial costs ($0 < k < \bar{\pi}c$) than under no fixed cost ($k = 0$). If $k > \bar{\pi}c$, then the optimal extent of discovery exceeds $\bar{\pi}$, which, in turn exceeds the optimal extent of discovery when there are no fixed costs. □

Proof of Proposition ??. By assumption, $\pi_1^2 + \pi_2^2 = \pi$. From the definitions of the cutoffs $\hat{\theta}_1^2$ and $\hat{\theta}_2^2$,

$$\exp\Big(-\frac{\hat{\theta}_1^2 - \underline{\theta}}{c_p + c_d}\Big) = 1 - \pi_1^2, \quad \text{and} \quad \exp\Big(-\frac{\hat{\theta}_2^2 - \hat{\theta}_1^2}{(1-\pi_1^2)(c_p + c_d)}\Big) = 1 - \frac{\pi_2^2}{1-\pi_1^2} = \frac{1-\pi}{1-\pi_1^2}.$$

Therefore,

$$\exp\Big(-\frac{\hat{\theta}_2^2 - \underline{\theta}}{c_p + c_d}\Big) = \exp\Big(-\frac{\hat{\theta}_2^2 - \hat{\theta}_1^2}{c_p + c_d}\Big) \exp\Big(-\frac{\hat{\theta}_1^2 - \underline{\theta}}{c_p + c_d}\Big) = \Big(\frac{1-\pi}{1-\pi_1^2}\Big)^{1-\pi_1^2}(1 - \pi_1^2).$$

8

[8] The expression $\left(\frac{1-\pi}{1-\pi_1^2}\right)^{1-\pi_1^2}(1-\pi_1^2)$ is increasing in π_1^2 for small π_1^2 and increasing in π_1^2 for big π_1^2. It reaches its minimum at $\pi_1^2 = 0$ or $\pi_1^2 = \pi$. Therefore, $\left(\frac{1-\pi}{1-\pi_1^2}\right)^{1-\pi_1^2}(1-\pi_1^2) \geq 1 - \pi > 1 - \bar{\pi}$. This implies that $\hat{\theta}_2^2 < \overline{\theta}$.

From the definition of $\hat{\theta}_1^1$,

$$\exp\Big(-\frac{\hat{\theta}_1^1-\underline{\theta}}{c_p+c_d}\Big)=1-\pi.$$

Since $\frac{1-\pi}{1-\pi_1^2}<1$ and $1-\pi_1^2<1$, we have

$$\Big(\frac{1-\pi}{1-\pi_1^2}\Big)^{1-\pi_1^2}>\frac{1-\pi}{1-\pi_1^2}\Rightarrow\Big(\frac{1-\pi}{1-\pi_1^2}\Big)^{1-\pi_1^2}(1-\pi_1^2)>1-\pi\Rightarrow\exp\Big(-\frac{\hat{\theta}_2^2-\underline{\theta}}{c_p+c_d}\Big)>\exp\Big(-\frac{\hat{\theta}_1^1-\underline{\theta}}{c_p+c_d}\Big).$$

Thus, $\hat{\theta}_1^1>\hat{\theta}_2^2$.

We decompose the impact of the number of rounds of discovery for θ in four regions. Let $C^N(\theta)$ denote the expected total investigation costs (as a share of c_p+c_d) for type θ with N rounds of discovery. To ease notation, we omit the argument θ of the probability of rejection, for example, writing r_2^2 rather than $r_2^2(\theta)$.

(1) $\theta<\hat{\theta}_1^2$: With two rounds of discovery, the expected investigation cost (as a share of c_p+c_d) is $C^2(\theta)=r_1^2\pi_1^2=1-\exp\Big(-\frac{\theta-\underline{\theta}}{c_p+c_d}\Big)$ because there is only a cost if his offer is rejected. Under one round of discovery, the expected investigation cost is $C^1(\theta)=r_1^1\pi=1-\exp\Big(-\frac{\theta-\underline{\theta}}{c_p+c_d}\Big)$. Therefore, the cost for $\theta<\hat{\theta}_1^2$ is the same regardless of how many rounds of discovery.

(2) $\theta\in[\hat{\theta}_1^2,\hat{\theta}_2^2)$: Under two discoveries, this type always incur a cost share of π_1^2 for the first discovery. If the first discovery does not work and the offer gets rejected, he incur a further cost of the second discovery π_2^2. Therefore, the expected cost is $C^2(\theta)=\pi_1^2+(1-\pi_1^2)r_2^2\pi_2^2$.

With one round of discovery, $\theta\in[\hat{\theta}_1^2,\hat{\theta}_2^2)$ makes an offer before discovery, and incurs the sole discovery cost only if his offer is rejected, so the expected cost is $C^1(\theta)=r_1^1\pi$.

Subtracting, we find that for $\theta\in[\hat{\theta}_1^2,\hat{\theta}_2^2)$,

$$\begin{aligned}
&C^1(\theta)-C^2(\theta)\\
&=r_1^1\pi-[\pi_1^2+(1-\pi_1^2)r_2^2\pi_2^2]\\
&=r_1^1\pi-[\pi_1^2+(1-\pi_1^2)\frac{1-\pi_1^2}{\pi_2^2}(1-\exp\Big(-\frac{\theta-\hat{\theta}_1^2}{(1-\pi_1^2)(c_p+c_d)}\Big))\pi_2^2]\\
&=\pi_1^2(1-\pi_1^2)+(1-\pi_1^2)\exp\Big(-\frac{\theta-\hat{\theta}_1^2}{(1-\pi_1^2)(c_p+c_d)}\Big)(1-\pi_1^2)-\exp\Big(-\frac{\theta-\hat{\theta}_1^2}{c_p+c_d}\Big)(1-\pi_1^2)\\
&=(1-\pi_1^2)[\pi_1^2+(1-\pi_1^2)\exp\Big(-\frac{\theta-\hat{\theta}_1^2}{(1-\pi_1^2)(c_p+c_d)}\Big)-\exp\Big(-\frac{\theta-\hat{\theta}_1^2}{c_p+c_d}\Big)]\geq 0.
\end{aligned}$$

The equality only holds at $\theta = \hat{\theta}_1^2$. Therefore, two rounds of discovery yield strictly greater investigation cost savings for all $\theta \in (\hat{\theta}_1^2, \hat{\theta}_2^2)$.

(3) $\theta \in [\hat{\theta}_2^2, \hat{\theta}_1^1)$. With two rounds of discovery, the expected total investigation costs share is:

$$
\begin{aligned}
C^2(\theta) &= \pi_1^2 + (1-\pi_1^2)\pi_2^2 + (1-\pi_1^2)(1-\frac{\pi_2^2}{1-\pi_1^2})(1-\exp\Big(-\frac{\theta-\hat{\theta}_2^2}{(1-\pi_1^2-\pi_2^2)(c_p+c_d)}\Big))(1-\pi_1^2-\pi_2^2) \\
&= \pi_1^2 + \pi_2^2 - \pi_1^2\pi_2^2 + (1-\pi_1^2)\frac{1-\pi}{1-\pi_1^2}(1-\exp\Big(-\frac{\theta-\hat{\theta}_2^2}{(1-\pi_1^2-\pi_2^2)(c_p+c_d)}\Big))(1-\pi_1^2-\pi_2^2). \\
&= \pi - \pi_1^2(\pi-\pi_1^2) + (1-\pi)^2(1-\exp\Big(-\frac{\theta-\hat{\theta}_2^2}{(1-\pi)(c_p+c_d)}\Big)). \qquad (9)
\end{aligned}
$$

With one round of discovery, the expected investigation cost share is $C^1(\theta) = r_1^1\pi = 1 - \exp\Big(-\frac{\theta-\underline{\theta}}{c_p+c_d}\Big)$.

The part of the difference $C^1(\theta) - C^2(\theta)$ that involves θ is:

$$
(1-\pi)^2\exp\Big(-\frac{\theta-\hat{\theta}_2^2}{(1-\pi)(c_p+c_d)}\Big) - \exp\Big(-\frac{\theta-\underline{\theta}}{c_p+c_d}\Big).
$$

Differentiating this difference with respect to θ at $\theta \in (\hat{\theta}_2^2, \hat{\theta}_1^1)$ yields:

$$
\exp\Big(-\frac{\theta-\hat{\theta}_2^2}{c_p+c_d}\Big)\exp\Big(-\frac{\hat{\theta}_2^2-\underline{\theta}}{c_p+c_d}\Big) - \exp\Big(-\frac{\theta-\hat{\theta}_2^2}{(1-\pi)(c_p+c_d)}\Big)(1-\pi) > 0,
$$

where the inequality follows from $\exp\Big(-\frac{\theta-\hat{\theta}_2^2}{c_p+c_d}\Big) > \exp\Big(-\frac{\theta-\hat{\theta}_2^2}{(1-\pi)(c_p+c_d)}\Big)$ and $\exp\Big(-\frac{\hat{\theta}_2^2-\underline{\theta}}{c_p+c_d}\Big) > 1-\pi$.

Therefore, $C^1(\theta) - C^2(\theta)$ increases in θ. Thus, if we can show that $C^1(\theta) - C^2(\theta)$ is positive for $\theta = \hat{\theta}_2^2$, then it is positive for every $\theta \in [\hat{\theta}_2^2, \hat{\theta}_1^1)$. Note that $\hat{\theta}_2^2$ is indifferent between making his first offer before the second discovery and after, so it is also characterized by Case (2), and we have already established that $\hat{\theta}_2^2$ gains from two rounds of discovery in Case (2). Therefore, $C^1(\theta) - C^2(\theta) > 0$ for all $\theta \in [\hat{\theta}_2^2, \hat{\theta}_1^1)$.

(4) $\theta \geq \hat{\theta}_1^1 (> \hat{\theta}_2^2)$: With two rounds of discovery, equation (**??**) reveals that the expected cost is:

$$
C^2(\theta) = \pi - \pi_1^2(\pi-\pi_1^2) + (1-\pi)^2(1-\exp\Big(-\frac{\theta-\hat{\theta}_2^2}{(1-\pi)(c_p+c_d)}\Big)).
$$

With one round of discovery, the expected cost is:

$$
C^1(\theta) = \pi + (1-\pi)(1-\exp\Big(-\frac{\theta-\hat{\theta}_1^1}{(1-\pi)(c_p+c_d)}\Big))(1-\pi).
$$

Subtracting yields

$$C^1(\theta) - C^2(\theta) = \frac{1}{(1-\pi)^2}[\pi_1^2(\pi - \pi_1^2) - \exp\Big(-\frac{\theta - \hat{\theta}_1^1}{(1-\pi)(c_p + c_d)}\Big) + \exp\Big(-\frac{\theta - \hat{\theta}_2^2}{(1-\pi)(c_p + c_d)}\Big)].$$

The derivative of this cost difference with respect to θ has the same sign as:

$$\exp\Big(-\frac{\theta - \hat{\theta}_1^1}{(1-\pi)(c_p + c_d)}\Big) - \exp\Big(-\frac{\theta - \hat{\theta}_2^2}{(1-\pi)(c_p + c_d)}\Big) > 0,$$

since $\hat{\theta}_1^1 > \hat{\theta}_2^2$. Therefore, $C^1(\theta) - C^2(\theta)$ increases in θ. Thus, if we can prove that it is strictly positive for $\theta = \hat{\theta}_1^1$, then it follows for all $\theta > \theta_1^1$. Note that $\hat{\theta}_1^1$ is indifferent between making his first offer before and after discovery, so it is a subcase of Case (3), where we have established the lower investigation costs of two rounds of discovery. Therefore, $C^1(\theta_1^1) - C^2(\theta_1^1) > 0$.

Combining all four cases of θ, we conclude that, holding total discovery constant and limited, expected total investigation costs are reduced when discovery is divided into two rounds. $\square$