Precedent and Doctrine in a Complicated World

Steven Callander*                  Tom S. Clark†

Abstract

Courts resolve individual disputes and create principles of law to justify their decisions and guide the resolution of future cases. Those tasks present informational challenges that affect the whole judicial process. Judges must simultaneously learn about (1) the particular facts and legal implications of any dispute; (2) discover the doctrine that appropriately resolves the dispute; and (3) attempt to articulate those rules in the context of a single case so that future courts may reason from past cases. We propose a model of judicial learning and rule-writing in which there is a complicated relationship between facts and legal outcomes. The model has implications for many of the important questions in the judicial process, including the dynamics of common law development, the path-dependent nature of the law, optimal case selection for rule-making, and analogical reasoning in the law.

*Graduate School of Business, Stanford University; sjc@stanford.edu.
†Department of Political Science, Emory University; tom.clark@emory.edu.
1 Introduction

Democratic governance usually entails a system of institutions working in tandem to construct policy, the sum total of which constitutes “the law.” Typically, legislatures initiate policy, translating public will into codified law. Executive bureaucracies develop expertise and administer the law. Courts handle disputes arising under the law and articulate principles, which we refer to as doctrine, to guide the resolution of cases. The micro-level processes that underlie the distinct functions of law-making institutions have been extensively studied, especially the particular forms of policy making that each institution undertakes. One feature that distinguishes judicial policy from legislative policy is that it is made through the resolution of discrete cases, one at a time, rather than through the articulation of globally-applicable policies. This observation has served as a starting point for many theories of rule construction and judicial politics (e.g., Kornhauser 1992a; b, Cameron 1993, Cameron, Segal and Songer 2000, Friedman 2006, Lax 2007, Carrubba and Clark 2012).

That judicial policy is made via the resolution of individual cases is important in part because the rules judges construct are themselves bound up with the factual scenarios giving rise to the cases. Indeed, the deep interconnection between factual scenarios and legal rules is a problem inherent in common law adjudication (Stein 1992), as it creates possible indeterminacy of the law in factually distinct cases and creates room for disputing the relevance of past decisions for resolving new cases. The principles of precedent and stare decisis, underlying common law adjudication, require existing rules be respected and applied in factually similar future cases (Levi 1949). Thus, judges face the unavoidable problem that the rules they make are described in terms of the specific cases they resolve and therefore leave unspecified, in principle, what the rules say about all other cases; their opinions can only try to articulate in a non-authoritative way what the court’s rules say about those other cases. The law therefore develops through a process of analogizing from past cases to new problems.

Making matters worse, the practical realities of the world mean that truly identical disputes rarely arise, and the fact-bound nature of common law judicial inquiry means that judges themselves cannot be sure exactly how certain fact patterns should be resolved without observing those facts in the context of a real-world case. Thus, facing a judge creating doctrine are two deeply connected,
and challenging, tasks: learning what their preferred doctrine implies for individual cases and articulating principles of law that can guide future cases, within the confines of a single case (see also Jordan N.d.). Indeed, the commonly-discussed idea that lower courts provide laboratories for experimentation in the law (e.g., Howard 1981) and venues for legal questions to percolate (e.g., Perry 1991, Clark and Kastellec 2013) reveals that judges are unsure of how a given legal standard would dispose of myriad potential cases and value the opportunity to experiment with different factual scenarios by resolving different cases to learn about how doctrine works.

The upshot is that judges on supervisory courts (e.g., the US Supreme Court) need to wade through a large set of cases to both learn about the optimal way to dispose of cases and communicate to lower courts what doctrine dictates about the cases those lower courts decide. These activities pose a strategic challenge for a court seeking to optimally select which cases to hear under discretionary jurisdiction. Unfortunately, most of the theoretical research on case selection has focused on non-doctrinal features of cases or features not directly related to these challenges (cf. Ulmer 1972, Caldeira and Wright 1988, Perry 1991, McGuire 1994). Knowing how courts use the resolution of individual cases to learn about and communicate authoritative doctrine, however, is critical to understanding the normative position of those courts in a democratic polity.

We develop a model of judicial learning and communication that speaks directly to the problem of legal rule writing and case selection in the world of uncertainty. The central tension we investigate concerns the court’s uncertainty about how different factual scenarios will manifest and the challenge that presents for a superior court in picking cases through which to articulate its doctrine. Our model relaxes a core assumption of many models of legal rule-making: that judges can easily order case facts and simply need to articulate doctrine that instructs lower courts and external actors how to resolve future cases (e.g., Kornhauser 1992b, Cameron 1993, Lax 2007; 2011). By contrast, we conceive of the mapping from case facts to legal outcomes as a stochastic mapping about which the courts are uncertain. To do so, we adopt the technology of Brownian motion previously used to study complex policy in other contexts (Callander 2011; 2008). In our model, case facts map in a nonlinear, and unknown, way to legal outcomes. A court seeks to establish doctrine by selecting individual cases and resolving them, thereby communicating to lower courts how particular factual scenarios should be resolved. The model yields direct implications for both the strategy a court should adopt when exercising discretionary review and the path dependence of
the law, as well as insights about the utility and function of analogical reasoning through precedent. First, why do courts rely on precedent in order to construct doctrine—i.e., why do we construct law through a series of cases? Second, what are the features of optimal doctrine—i.e., given that courts accumulate precedents dynamically to construct a line of doctrine, what are the characteristics of optimally-constructed doctrine? Third, how can a court optimally select cases to build doctrine—i.e., what makes a case more or less useful for the court when building a body of law? Fourth, how does the particular set of disputes or cases that arise affect the path of law—i.e., in what ways are precedent and doctrine path dependent?

2 The Problems of Uncertainty in Doctrine

Policy making is an informationally-demanding undertaking (e.g., Huber and Shipan 2002, Callander 2011), and judicial policy-making particularly so. Focusing on courts’ generalist nature, limited institutional capacities, and the massive amounts of disputes that arise as individual cases and require resolution, studies of the judicial hierarchy have focused on two types of problems: enforcement and compliance problems (e.g., Perry 1991, Cameron, Segal and Songer 2000, Carrubba and Clark 2012) and judicial learning about legal rules (e.g., Beim 2013, Clark and Kastellec 2013). We argue judicial learning is complicated by the challenge of articulating through a series of individual cases to lower courts and external actors what its doctrine says about all other cases. Oftentimes, these challenges are connected as in the legal notion about cases that constitute “good vehicles”—“strategic opportunities to advance and clarify the law” (Estreicher and Sexton 1984, 734).

Indeed, as theorists and empiricists have often noted, which cases a court hears, and in which order, can affect the rules a court adopts, the structure of the law, and even the way in which individual cases are resolved (e.g., Kornhauser 1992b;a, Kastellec and Lax 2008). However, while the peculiarities of case selection and doctrinal path dependence are something of which scholars are aware, there has been scant theoretical development of optimal case selection and doctrinal articulation in a world of legal path dependence. This limitation is particularly acute, because the judiciary’s method of rule-making—analogue reasoning—places questions about case selection, doctrinal complexity, and path dependence in the fore.
2.1 The problem of legal uncertainty

Judicial opinions generally serve three functions: reporting the facts of the case; offering legal principles (and supporting rationales) for resolving the case; and explaining how the particular facts relate to the adopted legal rule or standard. The first of these three tasks is largely straightforward. The second is more politically interesting, and scholars have invested considerable time studying the dynamics of rule adoption and justification. The (usually) collegial nature of rule-making courts means judges must negotiate with and convince their colleagues in order to gain support for their preferred rule (Maltzman, Spriggs and Wahlbeck 2000, Lax 2007, Cameron and Kornhauser N.d.).

The third task, while often overlooked in the literature, is critically important. Many times, judges can agree on a legal standard but may disagree about how the agreed-upon legal standard resolves a particular dispute. The source of that disagreement is that the judges have a different view of how a particular set of facts maps into the legal world. Consider, for example, equal protection claims. There generally exist three (competing) standards by which equal protection claims are evaluated. The most permissive standard is the rational basis standard, under which governmental discrimination survives an equal protection claim if there exists a rational basis for the discriminatory policy. The most exacting standard is strict scrutiny and subjects governmental discrimination to a three-prong test, which laws and governmental actions are unlikely to survive. This level of review is usually applied to race-based forms of discrimination or discrimination against other “suspect classes.” A mid-level standard is known as intermediate scrutiny or heightened scrutiny. This level of review is often applied to sex-based forms of discrimination.

Equal protection claims are among the most litigated areas of the law. Given there exist (essentially) only three levels of review and most forms of discrimination are unanimously subjected to one particular level of review, the question arises—what is being litigated, and why are so many cases being reviewed and painstakingly adjudicated by high-level courts? Reading opinions from well-known equal protection cases reveals judges are often struggling to determine and communicate how particular factual scenarios map onto the chosen level of review. In other words, judges are uncertain about how discriminatory a particular set of facts is and use their opinions to compare

---

1It bears mentioning, though, that there exists a legitimate question of whether courts strategically characterize case facts, rather than adopting some kind of objective account of the issues at hand.
and contrast a particular factual scenario to others in an effort to articulate how the law should treat the current and future factual scenarios.

2.2 Legal uncertainty in a case space

To see the theoretical implications of this observation, consider the case-space model of adjudication. This model assumes there exists a dimension (or dimensions) of legally-relevant facts in any given area of the law, cases are represented as locations on that dimension(s), and legal rules partition cases into a dichotomous disposition space (Kornhauser 1992a, Cameron 1993, Lax 2007). As a stylized example, consider search and seizure law. The factual dimension might represent the intrusiveness of a search, and legal rules are thresholds that determine what level of intrusiveness constitutes an unreasonable search. Many applications of case-space modeling are concerned with how collegial judges bargain over rules (e.g., Kornhauser 1992a, Lax 2007) or adopt rules or auditing strategies that include compliance with their doctrine in the lower courts (e.g., Cameron, Segal and Songer 2000, Carrubba and Clark 2012, Lax 2012). A critical assumption underlying these models is that judges know how to map facts into outcomes, the only challenges are either learning the facts of a case or writing rules that are agreeable or enforceable. Unfortunately, often what is meant by the “fact” dimension in case-space models is the legal consequence of a particular bundle of facts. Consider our stylized example of search and seizure law. A given search’s intrusiveness is not simply a case fact. Instead, its a qualitative assessment of the interaction of many possible facts (Friedman 2006). The relevant dimension for case-space models is a legal outcome—the aggregation of many facts—and the challenge is that judges may not be able to ex ante flawlessly order case facts into legal outcomes.

We assume, instead, that judges have limited knowledge of how particular facts manifest into legal outcomes. This motivates the laborious process by which judges work in their opinions to explain how particular facts constitute given outcomes that a chosen legal rule handles one way or another. In our model, case resolution is not simply about correcting errors or controlling the law. Rather, judges must work to learn about which case facts distinguish between different factual scenarios and, only then, form judgments and articulate doctrine.
2.3 Legal uncertainty in models of judicial learning

Focusing on how judges learn about rules, several models have examined the incentive to allow legal questions and issues to percolate in the lower courts and how a higher court can use related lower court decisions to learn about optimal legal rules (e.g., Cooter and Kornhauser 1980, Clark and Kastellec 2013, Beim 2013). Closely related are models in which courts hear cases in a sequence, developing legal rules through what they learn (e.g., Parameswaran 2012) or building up the strength of precedent by repeatedly reinforcing a policy position (e.g., Bueno de Mesquita and Stephenson 2002). In Bueno de Mesquita and Stephenson (2002) the key tension is created by a communication constraint across the judicial hierarchy. Their focus is then on how the higher court trades-off an increase in uncertainty when a stream of precedent is eliminated against overcoming a systematic bias in judgments. Closer to our work is Ellison and Holden (2014) that allows for richer types of uncertainty and broader classes of doctrine. Yet their focus is also on overcoming communication constraints (as well as cognitive limitations of the lower courts). They show how the courts develop codes and language to facilitate doctrinal communication when perfect communication is not possible.

We depart from these papers in allowing perfect communication and endowing the lower courts with the ability to perfectly apply doctrine. We focus instead on the complementary problem posed by the pure difficulty of learning how case facts map into legal outcomes and what that implies for judicial reasoning, case selection, and the construction and communication of doctrine. Along these lines is the model of Baker and Mezetti (2012), although they model only a single court and, thus, do not analyze doctrinal communication. In their model the court searches for a unique cut-point in the space of case facts that distinguishes between judgments (guilty or innocent), and therefore restrict attention to legal environments that conform perfectly to proper rules (Lax 2007). Moreover, they impose constraints on what the court can learn about an outcome, restricting the information gleaned to the appropriate judgment and suppressing whether the case was an easy call or only marginal.

Most of these studies conclude judges face informationally-demanding tasks requiring specialized knowledge not just about the mappings from case to outcomes (i.e., the law) but also about the particulars in any given case. Indeed, the informational challenges inherent in the judicial effort to
uncover their preferred legal rules and articulate effective doctrine is one important component for the division of labor between fact-finding trial courts and law-finding appellate courts. Law-finding by multiple, disparate lower courts would be inefficient, as it would impose considerable resource and communication costs (e.g., Kornhauser 1995). The model we develop contemplates both the challenges of constructing and articulating doctrine (usually the purview of courts at the top of the hierarchy) but also the challenges associated with resolving individual cases using precedent and doctrine—the purview of all courts.

2.4 Analogical reasoning and legal uncertainty

The sources of uncertainty we describe have particular import when considered in the context of judicial reasoning. Judges reason about new cases analogically, and many legal theorists contend that analogical reasoning is the hallmark of adjudication in common law systems as well as in many civil law systems\(^2\) (e.g., Levi 1949, Sunstein 1993). The essential feature of analogical reasoning in the courts is that individual cases are decided and the doctrinal implications of those decisions for other cases are determined by the similarities among cases (Sunstein 1999).

Despite the acceptance of analogical reasoning among practitioners and legal scholars, a formal basis that motivates its use has proven elusive. Our model closes this gap and captures the essential features of analogical reasoning. In our model legal rules and doctrine emerge endogenously from a series of related cases. Moreover, the development of legal rules makes use of collective amounts of information from many cases, and takes advantage of the collective wisdom of multiple judges deciding similar cases, as described by Sherwin (1999). An important consequence of this process is that adjudication leads to doctrine that entails exceptions, or carve-outs, to what might seem otherwise predictable and proper rules (e.g., Schauer 1991, Lax 2007). With this foundation of judicial reasoning in place, we then use the model to generate novel insights about the development of law through discrete cases.

\(^2\)It bears noting, though, there is debate in the legal literature about the normative implications of analogical reasoning (Alexander 1996; Cross 2003).
3 A Modeling Structure: Brownian Motion

We first describe the legal landscape in which learning and decision making occurs. We then turn to a series of implications derived from the model that provide micro foundations and prescriptions about judicial reasoning and optimal behavior in rule writing and case selection.

3.1 The Legal Landscape

We consider a single area of law, wherein the set of possible case facts is infinite and given by the real line, \( \mathbb{R} \). Each set of case facts is associated with a legal outcome, to which we often refer as the level of intrusiveness. Case facts are mapped into outcomes by the function \( \psi \), such that the outcome for case facts \( p \) is given by \( \psi(p) \), with \( \psi(p) \in \mathbb{R} \). To focus on learning over time, the same mapping is in effect throughout time. The challenge of judicial decision making is that the true path of \( \psi \) is unobserved by the courts. Thus, the courts do not necessarily know the correct legal outcome for a given set of facts \textit{ex ante}. In our search and seizure example, the courts might know a threshold of intrusiveness they find unacceptable (in the outcome space), but they do not know how intrusive any given search is until they review the case.

We adopt the perspective that the mapping \( \psi(p) \) is the realized path of a Brownian motion with drift \( \mu \) and variance \( \sigma^2 \). Figure 1 depicts one example. The drift parameter represents the expected rate of change in outcome as the case facts change. The variance parameter measures the volatility, or “noisiness,” of the process. An area of the law with higher variance, therefore, is more unpredictable and harder to grasp. We interpret \( \sigma^2 \) as the complexity of the legal environment.

The mapping \( \psi \) is highly non-linear and non-monotonic. The non-monotonicities represent the complicated and surprising interactions of particular sets of case facts (and provide the basis for legal carve-outs and exceptions). The case fact dimension, \( p \), can be seen as describing a single dimensional variable with \( \sigma^2 \) measuring the complementarities and interaction effects of different case facts. Alternatively, it can be interpreted as a projection from a rich, high-dimensional, fact space onto a rough ordering of case facts, where the complexity of the dimensions and their relationships to each other are captured by \( \sigma^2 \). We suppose that the courts know the rough ordering, \( p \), but they do not know how case facts map into outcomes until a case is heard.
Figure 1: *Example Brownian path.* The x-axis depicts the fact dimension \((p)\) and the y-axis depicts the legal outcome. The solid line shows the realization of the Brownian path \(\psi(p)\). The dashed line shows the courts’ *ex ante* belief about \(\psi(p)\), given knowledge of one single precedent and the drift parameter, \(\mu\).

At the beginning of play, the courts know the outcome associated with one set of case facts, what may be viewed as the “case of first instance.” We normalize this set of case facts to \(p = 0\). Substantively, this case anchors beliefs and knowledge in this area of the law; practically it means courts know that the mapping from case facts to outcomes passes through point \((0, \psi(0))\).

How courts reason and construct judgments and doctrine depends on what they know and what they do not, which, in turn, depends on the structure underlying the legal environment. In the following sections we show constructing this foundation with the Brownian motion leads to behavior that matches conventional wisdom on judicial reasoning (indeed, we will show that the Brownian motion is necessary to explain this behavior) and that it generates new predictions and insights into judicial practice. First, however, we describe the judicial environment.

### 3.2 Judicial Institutions

The judiciary consists of a single Higher Court and a set of Lower Courts. For each case that appears, the court hearing the case must issue a judgment, where judgments are binary. The judgment may be to permit or exclude, as in our intrusiveness example, or it may be guilty or innocent, liable or not, and so on.
To focus on the complexity of the legal environment, we assume common preferences among all courts—they want to get decisions correct and share a common standard regarding what the correct decision is. We normalize this standard to zero, such that cases facts corresponding to a positive outcome should be judged one way and those corresponding to a negative outcome the other way. Continuing the example of intrusiveness, we say a set of case facts is adjudicated as either Permit or Exclude, where the courts prefer Exclude if $\psi (p) > 0$ and Permit otherwise.\(^3\) We therefore abstract away strategic concerns about collegial bargaining (e.g., Kornhauser 1992b, Lax 2007) and compliance (e.g., Staton and Vanberg 2008, Lax 2012).

The value in a correct ruling—and the costliness of an error—depends on the outcome associated with the case facts. For a realization of $\psi (p)$ that is close to the threshold of zero, a correct or incorrect decision does not bear much consequence. Getting the judgment correct is better than getting it wrong, but making a mistake is not a significant error. On the other hand, for values of $\psi (p)$ that are from the threshold, an incorrect decision is a significant injustice and of much greater significance. For example, the more obviously intrusive a case, the greater disutility the Court suffers should the search be permitted. In this way, we follow the intuition described in Beim, Hirsch and Kastellec (2014). Specifically, we model the benefits and costs of ruling as strictly increasing in the distance of $\psi (p)$ from zero, and for simplicity we adopt the linear functional form.\(^4\) For case facts $p \in \mathbb{R}$, utility is given by:

$$u(p) = \begin{cases} 
|\psi(p)| & \text{if correct decision made.} \\
-|\psi(p)| & \text{if incorrect decision made.} 
\end{cases}$$

While the courts share preferences they differ in their capabilities, with lower courts experts in “fact-finding” and appellate courts experts in “law-finding” (e.g., Kornhauser 1995, Cameron and Kornhauser 2006). We adopt a particularly stark reduced form. The Lower Courts are precise discerners of facts but they lack the analytic power to translate facts into outcomes. Thus, given an appropriate doctrine with which to rule by, Lower Courts can be reliable implementers. The Lower Courts cannot, however, identify when doctrine is inappropriate or wrong. As is the case in

\(^3\)We set aside the zero probability event of $\psi (p) = 0$.

\(^4\)Switching applications, the outcome may measure the level of guilt; we are thus happier to be sending to jail for life someone who we are 95% sure is guilty versus someone for whom we are 70% sure.
practically, the Lower Courts handle the overwhelming majority of cases: Specifically, in each period
the Lower Courts hear a set of cases drawn according to the distribution \( f(\cdot) \), where \( f \) is uniform
over the \([-\tau_l, \tau_r]\) in the space of case facts, where \( \tau_l, \tau_r \in (0, \infty] \).

The Higher Court has the analytic capacity to hear cases and identify appropriate outcomes—to
learn the value of \( \psi(p) \) for particular case facts \( p \). Ideally, the Higher Court would hear every case
presented and issue appropriate rulings in each instance. This is obviously impractical in a large,
modern society. We operationalize this limitation, following standard models in the literature (e.g.,
Cameron, Segal and Songer 2000) by limiting the court to hearing a single case in each period and
paying a positive cost when it does so.

The courts also differ in their knowledge of the environment in which they operate. The more
capable Higher Court knows the value of the drift and variance parameters, \( \mu \) and \( \sigma^2 \), whereas the
Lower Courts do not. The Lower Courts know the outcome mapping is generated by a Brownian
motion and begin with some prior belief over these parameters, updating in each period via Bayes
rule. To save on additional notation, we simply describe these beliefs at any time \( t \) by \( \hat{\mu}_t \) and \( \hat{\sigma}^2_t \).

### 3.3 Timing and the Strategy of Judicial Behavior

Our interest is in the dynamic evolution of doctrine and judicial decision making. In each period,
\( t = 1, 2, 3, ... \), cases are heard and doctrine updated. Specifically, the sequence of actions in each
period is as follows:

1. The Higher Court decides whether to hear a case and, if so, selects the case facts \( p \in \mathbb{R} \),
   learns \( \psi(p) \), and issues a judgment \( J(p) \in \{P, E\} \), for Permit or Exclude.

2. The Higher Court issues an updated doctrine, \( \lambda_t(p) \subseteq \{\mathbb{R} \cup \emptyset\} \), for each set of possible case
   facts, \( p \in \mathbb{R} \).

3. The Lower Courts face cases according to the distribution \( f \) and issue judgments \( j(p) \in \{P, R\} \) for each case \( p \in \mathbb{R} \) that is heard.

The difference between what is reported by the court in Step 1 versus Step 2 can be thought
of as the difference between what is required and what is voluntary. The judgment in each case
describes the case facts and the binary judgment. This is the minimum that the court is obligated
to report.\textsuperscript{5} The information in Step 2, on the other hand, is conveyed voluntarily by the Court. It is the information in the opinions written by the justice(s), describing, amongst other things, the arguments and reasons for the decision. The Court may, in effect, leave this blank, such as when a case is decided without opinion. Or the Court may offer voluminous argument, logic, and opinion, and this need not be restricted to the current case but can make claims about other potential cases that could appear on the docket of the Lower Courts (i.e., an opinion may contain \textit{dictum}). We interpret this additional information as \textit{doctrine}. It contains non-binding information and advice for how the Higher Court wishes Lower Courts to treat case facts that appear.

Our interest lay in the question of when doctrine is necessary for efficient decision making by the Lower Courts, and what type of doctrine is sufficient to ensure this. Of particular interest are three types of doctrine. First, is when doctrine is \textit{empty}, or non-existent. In this case, the Higher Court merely issues judgments on the cases it hears and nothing more—i.e., \( \lambda_t(p) = \emptyset \) for all \( p \). Second is the \textit{minimalist} doctrine, in which doctrine is non-empty yet does not extend beyond the cases that have been heard by the Higher Court (e.g., Sunstein 1999). Of particular interest is the \textit{truthful minimalist} doctrine that requires \( \lambda_t(p) = \psi(p) \) if the case \( p \) has been heard and \( \lambda_t(p') = \emptyset \), for all \( p' \) otherwise. By describing only the outcomes it observes the court is merely conveying its knowledge through doctrine and is not extending this knowledge to recommendations or requirements on other sets of case facts that may emerge. The third doctrinal type extends beyond minimalist doctrine and is what we refer to as a \textit{rule}. A rule prescribes behaviors for sets of case facts that the Higher Court has not heard—i.e., \( \lambda_t(p) \in \mathbb{R} \forall p \). A rule is \textit{complete} if advice is offered on all possible case facts, otherwise it is an \textit{incomplete} or \textit{partial} rule.

As time passes, the set of cases heard by the Higher Court potentially grows. This creates an expanding history of cases and judgments from which courts at all levels can draw. We refer to the record of cases heard as the “case history”, denoted \( P_t = \{0, p_1, p_2, \ldots, p_{t'}\} \), where \( t' \leq t - 1 \). Each pair \( (p, J(p)) \) of case facts and the corresponding judgment constitutes a \textit{precedent}. Thus, over time, the body of precedent in an area of the law can grow. We collect precedents in the “judgment history”, denoted \( J_t = \{(0, J(0)), (p_1, J(p_1)), (p_2, J(p_2)), \ldots, (p_{t'}, J(p_{t'}))\} \). The Higher Court also possesses knowledge of the precise outcomes of case facts. The out-

\textsuperscript{5}Thus, we suppose that if the Higher Court accedes to hear a case, it necessarily renders a judgment. Allowing the court to defer would not impact our results (as, after all, the court could randomize in judgment if it didn’t wish to communicate, and, through doctrine in Step 2, can convey any partial signal it wishes.)
come history collects the pairs of cases heard and their corresponding outcomes, denoted $H_t = \{[0, \psi(0)], [p_1, \psi(p_1)], [p_2, \psi(p_2)], \ldots, [p_t, \psi(p_t)]\}$.

### 3.4 A Typology of Cases

As precedent accumulates, the model generates a rich set of cases to be heard. Among this variety, two divides are of substantive importance. First is the distinction between cases existing within the boundaries of established precedent versus those exploring new terrain outside the confines of precedent. In line with our emphasis on innovation in legal knowledge, we borrow terminology from the innovation literature (Schumpeter 1934) and refer to cases in new terrain as *exploratory* and cases within previously established regions as *exploitative* (of previous knowledge).

Formally, a case is exploitative if the case facts are in between two previously heard cases—i.e., $p_t \in (\min \{P_t\}, \max \{P_t\})$. Otherwise a case is exploratory and falls to the right of the right-most case already heard or to the left of the left-most case in precedent. In what follows it will be useful to refer to the “nearest precedent” for a case that is being heard. For exploitative case $p$ the nearest precedents are $p_l, p_r \in P_t$ where $p_l < p < p_r$ and $p' \notin (p_l, p_r), \forall p' \in P_t$. For exploratory case $p$ there is only a single nearest precedent given by $\tilde{p}_l = \min \{P_t\}$ if $p < \tilde{p}_l$ and $\tilde{p}_r = \max \{P_t\}$ if $p > \tilde{p}_r$.

A second distinction among cases concerns judgments. It is often argued that an important role of new precedent is to tackle ambiguities in the previous case history. These ambiguities emerge in exploitative cases when the nearest precedents offer conflicting guidance (opposing judgments). We say that such cases are *standard*. If, on the other hand, the nearest precedents offer consistent judgments then this ambiguity is not present. We say that such a case is *non-standard* should the Higher Court choose to hear it. Formally, for an exploitative case $p$, and nearest precedents $p_l, p_r \in P_t$, the case is standard when $J(p_l) \neq J(p_r)$ and it is nonstandard otherwise. An exploratory case is standard on the right flank if $J(\tilde{p}_r) > 0$ and $\mu < 0$ (or $J(\tilde{p}_r) < 0$ and $\mu > 0$), and non-standard otherwise (the requirement for cases on the left flank is analogous).
4 Legal Reasoning in a Complicated World

Our model both captures past empirical observations and generates new predictions that can be verified by new observations (e.g., Kuhn 1962). In this section, we show how legal reasoning within our model matches conventional wisdom about how judges use precedent and analogy. We then turn to new predictions about legal reasoning and, in subsequent sections, the selection of cases and the path of law.

The problem “fact-finding” Lower Courts face is how to render judgments on case facts without knowing the correct outcome. This is a hard problem, yet the Lower Courts have guidance from previously decided cases from the Higher Court. The adherence to precedent – the principle of stare decisis – is perhaps the dominant characteristic of judicial decision making. Our courts are no different and follow precedent diligently and precisely, matching this empirical regularity. We formalize the use of precedent as follows.

**Definition 1** Precedent is followed if a lower court faces case facts \( p_t \in P_t \) and rules \( j(p) \) such that \( (p, j(p)) \in J_t \).

In the application of precedent, two questions are prominent. How much information do the lower courts need to faithfully adhere to precedent? And, second, how useful is precedent in a complicated world? Lemma 1 provides the answer to the first question. To follow precedent without error the lower court need only possess the judgment history of the Higher Court.

**Lemma 1** The judgment history is necessary and sufficient for the application of precedent by Lower Courts.

Lemma 1 is important for what is not required: Lower Courts do not require doctrinal guidance or even information about outcomes to apply precedent successfully. Doctrine can be empty without interfering with the application of precedent. This result highlights the fundamental simplicity of stare decisis. On any case that matches a precedent, the Lower Courts merely need to know the Higher Court’s judgment and imitate that judgment.

The simplicity of stare decisis should, at the same time, reveal its limitations, the focus of the second question above. It is readily apparent that precedent, as defined, applies only to a trivially

---

6This follows immediately from our assumption of common preferences across the courts. Allowing for differing preferences will, as in other models of judicial hierarchies, weaken but not destroy adherence to precedent as the informational value remains. We return to this point in the discussion section.
small set of cases—those with case facts that are \textit{identical} to a case previously heard. This definition is sufficiently narrow as to verge on irrelevance. An accepted truth held by legal academics, judges, and lawyers alike, is that every case is unique. Each new case differs—sometimes substantially, sometimes minimally—from all cases that have come before. This uniqueness is a second empirical regularity that is captured in our model. In our model the courts face an infinite set of possible case facts, where each set of case facts is essentially unique. In fact, in the model as in reality, the probability that a case facts match a precedent exactly—down to the final detail—is exactly zero.

Thus a conundrum: given no case ever exactly matches precedent, how should lower courts decide new cases? And, of what use is precedent if it applies only to zero-probability events? The answers to these questions are intertwined. A corollary to the axiom that all cases are unique is that no case exists in a vacuum. Cases possess similarities to each other, and the practice of law is to exploit these similarities to inform judgment. The leading method in practice to compare and make use of factually distinct cases is to reason by analogy. Our model matches this third empirical regularity as the lower courts use precedent to reason via analogy on factually distinct cases. Although often considered the defining feature of what lawyers do (Sunstein 1993), what exactly analogical reasoning entails and its logical foundations, have proven frustratingly elusive. We provide here a rigorous foundation for the use of analogical reasoning in judicial decision making.

To develop the logic, we begin with the following question: How do the Lower Courts optimally formulate judgments should they possess the complete outcome history? We then subsequently return to what this means for doctrine.

We begin with exploitative cases (those between precedents). Proposition 1 shows optimal judgment requires reasoning by analogy from the nearest precedents. The proposition is striking for what is omitted—to reason by analogy, the Lower Courts do not require any theoretical knowledge of the underlying legal area (the drift and variance parameters). Rather, they formulate judgments with only the outcomes from past cases and by simple interpolation between them.

\textbf{Proposition 1} For an exploitative case, \( p \), the optimal Lower Court judgment balances proportionally the nearest precedents \( p_l \) and \( p_r \). Formally, the expected outcome is:

\[
\mathbb{E} [\psi (p)] = \psi (p_l) + \frac{p - p_l}{p_r - p_l} (\psi (p_r) - \psi (p_l)).
\]

and the court rules \( j (p) = P \) if \( \mathbb{E} [\psi (p)] > 0 \) and \( j (p) = E \) otherwise.
Figure 2: *Example Brownian bridges*. The x-axis depicts the fact dimension \((p)\) and the y-axis depicts the legal outcome. The points illustrate known precedents. The dashed lines depict expected outcomes for cases not heard by the Higher Court. The lower courts interpolate between known points, known as Brownian bridges, and they extrapolate on the flanks using the drift term, \(\mu\).

This situation is depicted in Figure 2. The dashed line between the two precedents at \(p_l\) and \(p_r\) represents the expected outcome for the exploitative cases between them (mathematically, this is known as a *Brownian bridge*). The optimal decision rule for the Lower Courts then is to adjudicate according to this expected outcome.

When analogizing, the Lower Courts utilize only the nearest precedents, discarding all other precedent. This matches another empirical regularity as in practice judges and lawyers seek out guiding (controlling or on point) precedents to inform their decisions. Our result refines this insight by showing why distance is not the sole factor in selecting guiding precedent as the *direction* of precedent also matters. Specifically, the proposition imposes no requirement that the “nearest precedents” are the two closest precedents in an absolute sense. Moreover, the result yields the novel insight that analogical reasoning is not merely effective but is, in fact, the optimal decision making procedure. Critics of analogical reasoning argue it is without basis, and proponents often weaken their claim to it merely being a reasonable procedure. We provide not only a rigorous basis but show that in a complicated world it can indeed be optimal.

The optimality of analogical reasoning does not imply, of course, that it is infallible. Analogical reasoning provides the Lower Courts with the best guess as to the true outcome and mistakes
are inevitable. The true outcome is distributed normally around this expected outcome, and for exploitative cases the variance is given by the following:

\[ \text{var} (\psi (p)) = \frac{(p - p_l)(p_r - p)}{p_r - p_l} \sigma^2. \]  

(1)

Uncertainty reaches a maximum at the exact center of the bridge and approaches zero toward the ends. Intuitively, uncertainty is greater the further a new case is from what is known from precedent. To the best of our knowledge, this result provides the first measure of the accuracy of analogical reasoning.

Proposition 1 implies that to reason optimally via analogy, the Lower Courts require information about the outcomes of precedent and not simply the judgments. Without outcome information, the courts’ interpolation would be crude and inefficient. At the same time, the courts do not require any additional information to decide optimally. For exploitative cases, therefore, optimal doctrine cannot be empty but it need not extend beyond minimalist.

**Proposition 2** The minimalist truthful doctrine is necessary and sufficient for optimal analogical reasoning on exploitative cases.

The implication that the Judgment History alone is insufficient for optimal legal reasoning in complicated environments provides a micro-foundation for the general reluctance of supervisory courts to decide cases without opinion. Without additional information, the Lower Courts cannot weight precedent—they can only average them—and this leads to suboptimal decision making. Proposition 2 also provides a micro-foundation for why supervisory courts are not more expansive in their doctrine. Doing so is without cost in our model, yet the proposition shows additional doctrinal guidance is not necessary. This fact will bite in richer settings, such as when preferences across the judicial hierarchy are not aligned. We return to this possibility in the discussion section.\(^7\)

Exploratory cases, by contrast, provide a challenge to analogical reasoning. The difficulty is that interpolation is not possible for cases on the flanks of an existing body of precedent. Nevertheless, we show that analogy still guides the decision making of lower courts. However, without the ability

\(^7\)These results highlight a critical distinction between our model and past models of precedential accumulation (e.g., Bueno de Mesquita and Stephenson 2002). Because of the infinite variety of cases, courts cannot completely articulate a legal rule for every possible case, yet communication of doctrine that does exist is precise and complete and does not require reinforcement. Instead, the value of precedent here is in gathering further information to make doctrine more complete so as to provide guidance on the infinite variety of cases faced by the lower courts.
to interpolate, the courts must complement analogical reasoning with theoretical knowledge of the underlying legal environment.

**Proposition 3** For an exploratory case \( p \), the optimal Lower Court judgment extrapolates from the nearest precedent, \( \bar{p} \in \{\bar{p}_l, \bar{p}_r\} \). Formally, the expected outcome is:

\[
E[\psi(p)] = \psi(\bar{p}) + (p - \bar{p}) \mu_t.
\]

and the court rules \( j(p) = P \) if \( E[\psi(p)] > 0 \) and \( j(p) = E \) otherwise.

Proposition 3 refines the idea of analogical reasoning, showing a balancing test across precedent is not always sufficient for optimal judgment. For exploratory cases the Lower Courts have only one precedent from which to extrapolate, yet that one precedent does not offer guidance on how to perform this extrapolation. The solution is for the court to combine two types of knowledge: the experience of precedent and theoretical knowledge of the underlying environment. The theoretical knowledge guides the extrapolation from precedent, as depicted on the flanks in Figure 2.

Lower Court judgments on exploratory cases are, of course, fallible. The drift lines in Figure 2 represent expected outcomes, with the true outcomes distributed normally. Variance increases linearly in the distance from precedent, reflecting again the intuitive property that uncertainty is greater the further a set of case facts is from what is known. When the nearest precedent is \( \bar{p} \in \{\bar{p}_l, \bar{p}_r\} \), variance for case facts \( p \) is given by:

\[
var(\psi(p)) = |p - \bar{p}| \sigma^2. \tag{2}
\]

The accuracy of analogical reasoning here, as was the case with exploitative cases, decreases in the difference between a new case and existing precedent, although for exploratory cases the level of uncertainty increases without bound.

With Lower Court reasoning in place, we turn again to what this implies for doctrine. The informational requirements of Proposition 3 imply that outcome information alone is insufficient for optimal analogical reasoning. The Lower Courts lack knowledge of \( \mu \), possessing only a noisy belief encapsulated by the estimate \( \hat{\mu}_t \). Consequently, on exploratory cases, optimal decision making by the Lower Courts requires that doctrine convey theoretical knowledge.
Proposition 4 The minimalist truthful doctrine is insufficient for optimal analogical reasoning on exploratory cases. For optimal analogical reasoning, a sufficient partial rule is the doctrine \( \lambda_t(p) = \mathbb{E}[\psi(p)] \) for all exploratory case facts \( p \), and minimalist otherwise.

That is, in exploratory cases, the Higher Court must offer a partial rule, \( \lambda_t(p) \), which contains theoretical knowledge for all cases outside the range of what has previously been considered. This rule, or guidance, represents what is known as *dicta* in the law and is necessary for optimal analogical reasoning. The partial rule in the proposition provides specific guidance for all cases that face this constraint and is, therefore, sufficient for optimal adjudication by the Lower Courts. As the doctrinal rule is linear, an alternative is for doctrine to provide only a scattering of points, or even a single point, in this line. Substantively, this would take the form of the Higher Court describing the judgment and outcome it would expect on some hypothetical case. To work this approach requires the Lower Courts to interpret this hypothetical case in a special way. The Lower Courts must know to interpolate between the nearest precedent and this hypothetical case, and to continue extrapolating from the hypothetical case at the same rate (i.e., to recreate the doctrinal line in the proposition by tracing from precedent through the hypothetical case). Whichever of these approaches is the most realistic in practice, what they and other all sufficient partial rules share is that doctrine must convey information beyond that contained in the minimalist truthful doctrine.

Proposition 4 contrasts sharply with Proposition 2. On exploitative cases, it is possible for non-specialist judges—those who can reason from analogy but possess no specialized knowledge—to adjudicate efficiently with only a minimalist doctrine. In contrast, these same judges are at a loss on exploratory cases if additional guidance is not forthcoming. Exploratory cases require specialist knowledge of an area of law, knowledge that is either possessed innately or is communicated via doctrine. The contrast between Propositions 2 and 4 exposes the subtlety of doctrinal communication. By characterizing exactly what analogical reasoning entails, we are able to draw deeper insights into analogical reasoning and what it implies for the construction of doctrine.

The analogical reasoning that emerges from our model is intuitive and captures prominent stylized facts about judicial decision making in practice. As accepted as is the intuition for analogical reasoning in practice, formal models have not been able to adequately capture its logic and structure. The final result of this section demonstrates why this has been the case. Not only does the Brownian motion framework lead to analogical reasoning by the court. The causality runs the other
way. That is, if analogical reasoning is optimal for the courts then the mapping from case facts to outcomes must be generated by a Brownian motion.\footnote{Formally, if discontinuous mappings were allowed then this statement would require Levy processes, which are a simple sum of a Brownian motion and a jump diffusion process.}

Theorem 1 Analogical reasoning by linear interpolation and extrapolation from the nearest precedents is optimal if and only if the legal mapping is generated by a Brownian motion.

In a sense, our results imply the formalization that we adopt. Were the mapping generated by some other process—say a deterministic quadratic function—analogical reasoning would not be optimal, and, indeed, the question of its effectiveness, even as a rule-of-thumb, would remain an open question. Thus, the belief that analogical reasoning is the basic tool of lawyers and judges is, in turn, a belief that the legal environment is complicated, difficult to infer, and best approximated formally by a Brownian motion.

5 The Logic of Certiorari

The Higher Court’s problem is different from that of the Lower Courts. Although it is a “law-finding” court with no need itself for analogical reasoning, analogical reasoning nevertheless drives the logic of its actions. The Higher Court must decide whether and which case to hear. To make this determination the Higher Court anticipates the use of analogical reasoning by the Lower Courts and must incorporate this logic into its calculations.

A first step in this process is to understand how well analogical reasoning works. As remarked previously, analogical reasoning is optimal yet fallible. The realized utility from a case is the distance between the outcome and the decision threshold at zero. With incomplete information, this calculation is not so simple as we must work with expected utility and account for all possible realizations of the outcome. Nevertheless, we show that this calculation takes a surprisingly subtle yet simple form.

For case facts $p \notin P_t$, the judgment to Exclude delivers expected utility:

$$EU(\text{Exclude}|p) = \int_0^\infty |\psi(z)| \phi(z) \, dz - \int_{-\infty}^0 |\psi(z)| \phi(z) \, dz,$$
where \( \phi(z) \) is the density of the normal distribution of mean \( \mathbb{E} [\psi(p)] \) and variance \( \text{var} (\psi(p)) \), with variance given by Equation 1 or 2 depending on the type of case. The first term on the right-hand side is the set of outcomes for which the decision to Exclude is correct, whereas the second term is the outcomes for which this decision is incorrect. Rearranging, we have:

\[
EU (\text{Exclude}|p) = \int_{-\infty}^{\infty} \psi(z) \phi(z) \, dz - \int_{-\infty}^{0} \psi(z) \phi(z) \, dz
\]

\[= \int_{0}^{\infty} \psi(z) \phi(z) \, dz + \int_{-\infty}^{0} \psi(z) \phi(z) \, dz\]

\[= \int_{-\infty}^{\infty} \psi(z) \phi(z) \, dz = \mathbb{E} [\psi(p)] \] (3)

Thus, expected utility is simply the expected value of the outcome mapping. Thus, the Lower Courts should rule Exclude when \( \mathbb{E} [\psi(p)] > 0 \) and Permit when \( \mathbb{E} [\psi(p)] < 0 \). (This property confirms the conclusion in Propositions 1 and 3.) The more interesting, and surprising, aspect of Equation 3 is that variance plays no role.\(^9\) Expected utility is entirely independent of the complexity of the legal area.

**Proposition 5** Expected utility at time \( t \) depends only on the outcome history, \( H_t \), and the drift term, \( \mu \). In particular, it is otherwise independent of the complexity of the area of the law, \( \sigma^2 \).

The logic for this result is easiest to see geometrically. Consider case facts \( p \) in the left panel of Figure 3, with nearest precedents \( p_l \) and \( p_r \) and expected outcome \( \mathbb{E} [\psi(p)] \). Expected utility is given by this expected outcome, yet true realized utility will only equal this in the unlikely event that the true outcome matches this point precisely. The range of possible outcomes is distributed normally around this point, and the symmetry of this distribution can be exploited to simplify the problem.

To see this, consider the pair of possible outcomes, \( l_1 \) and \( u_1 \), that are equally distant from the expected outcome. If \( l_1 \) is the realized outcome then utility is lower than expected, whereas utility is higher than expected if the realized outcome is \( u_1 \). Conveniently, the decrease at \( l_1 \) exactly matches the increase at \( u_1 \), such that in expectation the net impact on expected utility is zero. As this pair of outcomes is arbitrary, the same logic holds for any pair of symmetric outcomes. As the

\(^9\)Recall the expressions for expected outcome in Propositions 1 and 3, respectively.
pairs all net out in utility terms, this implies that expected utility is exactly given by the expected outcome $E[\psi(p)]$.

The simplicity of expected utility is elegant, yet it leaves two puzzling questions: What—if any—role does uncertainty and legal complexity play in the making of law? And, how does this role influence the granting of certiorari and the logic of case selection? Return again to the left panel of Figure 3 and the pair of outcomes $l_2$ and $u_2$. Because the expected outcome is positive, Lower Courts rule Exclude, yet if the true outcome is $l_2$ this judgment will have been in error. This possibility did not impact the logic of Proposition 5 as the mistake at $l_1$ is exactly counterbalanced by the extra utility should $u_2$ be the true outcome.\footnote{Note that the outcome $l_2$ enters expected utility as a negative value as the judgment is in error.}

Yet the possibility of a mistake matters to the Higher Court in deciding whether to hear a new case. By hearing case facts $p$ the Higher Court can learn the true outcome and update doctrine, guiding the Lower Courts to the correct judgment should the true outcome be $l_2$ (or any outcome below zero). By accumulating precedent, the Higher Court can slowly but surely remove error from the Lower Courts’ judgments. The benefit of this corrective is directly proportional to how often judgments are wrong and how serious are the mistakes made. Both of these factors strictly increase...
in the complexity of the legal area ($\sigma^2$). All else equal, therefore, for case facts $p$, the value to the Higher Court of issuing certiorari is strictly increasing in complexity.

This fact is central to the logic of certiorari, yet a complete account requires a further step. Hearing case facts $p$ removes all future uncertainty about the correct judgment for $p$, yet, as noted earlier, the direct application of the precedent at $p$ is vanishingly unlikely. To appreciate the full value of granting certiorari to case facts $p$, the Higher Court must take a broader perspective. Consider again Figure 3 and suppose the realized outcome is $l_2$. Observing the outcome of $p$ splits into two the Brownian bridge between $p_l$ and $p_r$, one bridge is between $p_l$ and $p$ and the other from $p$ to $p_r$. Both of the new bridges now cross zero, at $n_1$ and $n_2$, respectively. This is important as now the Lower Courts, applying analogical reasoning to precedent, rule Permit for cases within the interval $(n_1, n_2)$. Granting certiorari to case facts $p$, therefore, changes lower court behavior not only at $p$ but over a possibly broad range of case facts. In this event, the precedent $p$ creates a carve-out in the doctrinal space.\(^{11}\)

In deciding which case to hear, therefore, the Higher Court must determine how a single case can affect the quality of judgment across all possible cases. As law-finding courts do in practice, the Higher Court in our model selects cases with a broader view toward shaping the law beyond the particular case at hand. Proposition 6 describes how this broader perspective interacts with complexity. The full impact of hearing a new case accords with the logic for a single case: The greater the complexity of the legal area, the more value exists in hearing new cases, for exploratory and exploitative cases alike.

**Proposition 6** The expected utility from hearing case facts $p \notin P_t$ is strictly increasing in the complexity of the area of the law. Namely, for the outcome history $H_t$, if it is profitable to hear a case with variance $\bar{\sigma}^2$, it is profitable to hear a case whenever $\sigma^2 > \bar{\sigma}^2$.

This result comports with intuition: The more uncertain is an area of the law, the more value there is in hearing additional cases. The subtlety of Propositions 5 and 6 is that uncertainty does not play into utility directly. Instead, uncertainty represents the potential value that can be unlocked by hearing new cases and getting doctrine right.

\(^{11}\)For cases within this interval, with the exception of $p$ itself, the possibility of error is not removed completely, but the quality of the ‘best guess’ as to the correct judgment has been improved.
6 Optimal Case Selection

In choosing a case to hear, the Higher Court faces a number of considerations. First, and most prominent, is the probability that the Lower Court judgment on a particular case is wrong. That is, the probability the Lower Courts rule Exclude when the correct judgment is Permit, and vice error. We refer to this probability as the level of error uncertainty. In our model there always exists at least one case for which the level of error uncertainty is exactly $\frac{1}{2}$. On these cases, the Lower Courts are as likely to be wrong as they are right. A natural intuition is that it is these cases the Higher Court should choose to hear as doing so will correct, for the case heard, the most mistakes in adjudication. Despite the appeal of this logic, it is mistaken.\footnote{For simplicity, we hereafter assume that the Higher Court chooses cases to maximize the efficiency of the law in each period. A far-sighted court may sacrifice current utility if it improves long-run performance. This would complicate the proof of Proposition 7 considerably without changing the underlying logic. This assumption need not apply to the Lower Courts, which optimally adjudicate via analogical reasoning regardless of their planning horizon.}

**Proposition 7** With probability one the Higher Court chooses a case each period that has less than maximum error uncertainty (i.e., less than $\frac{1}{2}$).

This result reflects the broader perspective the Higher Court takes. The Court weighs how a judgment affects the body of law, not just the particular case in front of it, leading to a view of utility that is broader than mere error uncertainty. This result captures the conventional wisdom that lower level courts focus on the individual cases before them whereas higher courts use the review of select cases to affect the “big picture” of the law (e.g., Kornhauser 1995). While the rationales the Higher and Lower Courts pursue are different, the underlying logic of their behavior is very similar. That cases are unique yet interdependent inspires the Lower Courts to reason by analogy to form their judgments. Surprisingly, it is the same force that inspires the Higher Court to account for the broader impact when choosing which cases to hear.

This leaves the question of which cases the Higher Court chooses to hear. Unfortunately, this question does not permit a tight answer. The interdependence of cases within the model captures nicely the reality of judicial decision making, yet it also captures the fact that reality is messy. Nevertheless, the incentives the Higher Court must trade-off in selecting cases are intuitive. The basic divide is between different types of uncertainty. The Court cares about both how frequently the Lower Courts make mistakes in their judgments (the error uncertainty) as well as how serious are the mistakes that are made.
Figure 4: *Cases the Higher Court will hear on a spanning Brownian bridge (exploitative and standard).*

To see the distinction, consider Figure 4 and exploitative cases between $p_l$ and $p_r$. Case facts $p$ have an expected outcome of 0, and thus an error uncertainty of $\frac{1}{2}$ as, regardless of the judgment rendered, the court will be wrong half of the time as exactly half of the possible outcomes are above zero and the other half below. Yet the mistakes that are made may not be wrong by too much. That is to say, if the uncertainty at case facts $p$ has low variance, then the likely outcomes are clustered around 0 and any mistake made is likely to impose little cost. In contrast, the case facts at $\frac{p_l + p_r}{2}$ have lower error uncertainty but higher overall uncertainty over the true outcome (as the variance of beliefs is maximized at the mid-point between precedents). Consequently, in ruling Permit on these cases the Lower Courts are right more than they are wrong, but when they are wrong they are more likely to be spectacularly wrong.

In choosing a case to hear the Court must balance these incentives. For a single case this calculation is straightforward (as for that case the true outcome is revealed and all future errors are avoided). Accounting for the broader impact across all cases is much more involved. What is clear is that in balancing these incentives the Higher Court will choose a case strictly between the cases that maximize error and total uncertainty, respectively. Proposition 8 formalizes this result for exploitative cases.
Proposition 8 For the set of exploitative cases \( p \in (p_l, p_r), \) where \( \psi (p_l) = t > 0 \) and \( \psi (p_r) = b < 0, \) the optimal case facts for the Higher Court to hear, \( p^* \), satisfy the following requirements:

(i) \( t < |b| : \) Then \( |p^* - p_l| < |p_r - p^*| \) and \( \mathbb{E} [\psi (p^*)] < 0; \)
(ii) \( t > |b| : \) Then \( |p^* - p_l| > |p_r - p^*| \) and \( \mathbb{E} [\psi (p^*)] > 0; \)
(iii) \( t = |b| : \) Then \( |p^* - p_l| = |p_r - p^*| \) and \( \mathbb{E} [\psi (p^*)] = 0. \)

Figure 4 depicts the region of case facts from which the Higher Court selects for case (i) in the proposition (when \( t < |b| \)). The selected case maximizes error uncertainty or total uncertainty only if \( t = |b|, \) in which it maximizes both as they coincide at the same set of case facts. This is a zero probability even and occurs only when the ends of the bridge are perfectly symmetric.

This result matches the empirical regularity that higher courts often choose cases for which the outcome is already “known,” or at least thought to yield one judgment with high probability. While seemingly a puzzling practice, it emerges naturally in our model in which the outcomes of cases are interdependent. For such a case it may very well be that the correct judgment is in little doubt, it remains possible that the true outcome is highly uncertain, and thus confirming the correct judgment while learning and communicating the true outcome can provide benefit to a broad swath of other cases, cases that the Higher Court does not have the time to hear.

It is striking that the logic for this behavior by the Higher Court is the exact same logic that drives analogical reasoning by lower court justices. For both levels in the judicial hierarchy, behavior is driven by the interdependency of case outcomes. The Lower Courts reason by analogy as outcomes are interdependent, and this same interdependence motivates the Higher Court to consider the broader impact of the cases it hears (as it knows the Lower Courts will reason by analogy from the precedent handed down). The power of our theorizing is to explain these twin behaviors and to expose the connection between them, providing a parsimonious theory of broad judicial behavior.

7 The Path of Law

A common view is that the evolution of the law proceeds in a thoughtful, conservative trajectory, with knowledge steadily accumulating and ultimately converging, possibly without end, on a complete understanding of the legal environment. This view often clashes with the reality of judicial decision making that is more haphazard, irregular, and apparently without reason. In this section
we turn to aggregate properties of the law, reconciling these views, showing how a seemingly chaotic trajectory can nevertheless emerge from rational, deliberate calculation.

Consider the first period. The Higher Court faces a single precedent with the choice of only exploratory cases, either to the right or the left.

**Proposition 9** If the Higher Court hears a case in the first period then $p_1^* > 0$, where $\mathbb{E}\left[\psi(p_1^*)\right] < 0$ for $\tau_r$ sufficiently large. Moreover, if the court hears a case for $\hat{\sigma}^2$ and $\hat{\psi}(0)$, then it hears a case whenever $\sigma^2 \geq \hat{\sigma}^2$ and $\psi(0) \geq \hat{\psi}(0)$.

The Higher Court chooses to hear cases only when there is sufficient value to extract from expanding the set of precedent. This requirement is satisfied when the area of law is sufficiently complex ($\sigma^2$ is large enough), or if the existing precedent is sufficiently unhelpful ($\psi(p)$ is sufficiently far from zero).

Should the Higher Court choose to hear a case, it will choose a case to the right of zero as these cases have greater error uncertainty (given $\psi(p) > 0$ and $\mu < 0$). Less obvious is that the Court chooses a case sufficiently distant from the existing precedent that the expected outcome is on the other side of zero (less than 0). Nevertheless, the logic is the same as in Proposition 8. Cases with expected outcome below zero offer higher combinations of error and outcome uncertainty than do the corresponding cases with expected outcome greater than zero.$^{13}$ Thus, when the Higher Court decides to explore an area of the law it does so boldly.

Hearing a case in the first period establishes a second precedent, and with this the environment complicates further and the choice of the Court is less straightforward. The Court now has a choice over exploitative and exploratory cases, and, depending on the outcome of the case heard in period 1, standard or non-standard cases.$^{14}$ The choice of case in the second period depends on the precise case facts and outcomes of existing precedent as well as the complexity of the legal area. The richness of this dependence only grows over time as more precedents are added and the possibilities multiply. Calculations reveal that no pattern appears in the sequence of cases, and it is possible that any type of case follows any other type (e.g., a standard exploitative case may be followed by another exploitative case or an exploratory case, by a standard case or a non-standard

$^{13}$The requirement on $\tau_r$, the right-side boundary of case facts, ensures there are enough cases in the neighborhood that will benefit from hearing $p_1^*$.

$^{14}$The exploratory case to the right of $p_1^*$ is non-standard if $\psi(p_1^*) < 0$ and standard otherwise. The reverse holds for an exploitative case between the two precedents.
case). The path dependence of the law, therefore, is not limited to the precise case facts chosen but extends even to the type of cases chosen.

Although the cases chosen exhibit no predictable pattern, this need not translate into complicated or chaotic doctrine. The behavior of the Lower Courts frequently reduces to simple rules, even as the set of precedent grows large. To see this, observe that it is possible to translate precedent and doctrine—via analogical reasoning and partial rules—into a set of cut-points in the space of case facts, where the cut-points demarcate between judgments of Permit and Exclude. We define doctrinal complexity as the number of cut-points required to fully describe doctrine. This concept captures how simple, or straightforward, it is for Lower Courts to follow the Higher Court’s advice.

For instance, if only a single precedent is in place, the Higher Court issues a doctrinal rule that extrapolates from the precedent according to the drift term, $\mu$. This induces simple threshold behavior by the Lower Courts. Namely, there exists some $q$ in the space of case facts, such that the Lower Courts rule Exclude for $p < q$ and Permit for $p > q$. This $q$ can be seen in Figure 1. Even when more precedents are available, it is still possible that a single threshold is sufficient to describe doctrine and lower court behavior. The number of thresholds required to describe doctrine depends on the particular outcomes realized. We say that a doctrine is maximally simple if it can permit a single threshold test for the Lower Courts, and that doctrine is more complicated the more thresholds that are required. We are interested in how doctrinal complexity evolves as more and more precedents are established.

While, as the law itself, doctrinal complexity evolves irregularly, it follows a more systematic path. In particular, doctrinal complexity can only move in one direction: upward. Doctrinal complexity cannot decrease, regardless of the type of case heard or the outcome observed.

**Proposition 10** *Doctrinal complexity weakly increases over time.*

The proposition explains why legal procedure is often so complicated. Whenever complexity increases in doctrine, it can never be reversed. Consequently, legal precedent faces a relentless pressure toward greater complexity. Thus, in our model, it is not possible for a case to be heard that ‘cuts’ through the complexity of doctrine.\(^{15}\)

\(^{15}\)To the extent that such cases are an empirical phenomenon, our model demonstrates their presence is not a fundamental property. Extending the model to isolate what might allow such overarching precedent is a worthy direction of investigation.
Proposition 10 establishes only the direction of change in doctrinal complexity, leaving open the cause of change and the rate of potential increase. The following results demonstrate how change in doctrinal complexity follows systematically from the type of cases that are heard. To see the possibilities, return to the non-standard case, \( p \), depicted in the left-hand panel of Figure 3. As described previously, should the Higher Court hear case \( p \) and observe outcome \( l_2 \), it would update doctrine and create a carve-out in the doctrinal space. Specifically, the Lower Courts now rule Permit for case facts in the interval \((n_1, n_2)\) and Exclude for case facts immediately to either side. The carve-out that is created makes following doctrine more involved as the Lower Courts must now retain two additional thresholds to faithfully execute doctrine. Thus, the carve-out increases doctrinal complexity, always by steps of two (and complexity, therefore, always remains on odd number). On the other hand, should the realization at \( p \) not create a carve-out, then doctrinal complexity is unaffected. These possibilities are described in the following.

**Proposition 11** When the Higher Court hears a non-standard case doctrinal complexity increases by two with probability in \((0, \frac{1}{2})\), otherwise it remains unchanged.

The proposition addresses only doctrinal complexity. For all possible outcomes the doctrine itself is updated. However, it is only if an outcome is on the opposite side of zero to what is expected that the complexity of doctrine increases, and this surprising outcome occurs less than half of the time. The impact of non-standard cases is, therefore, non-standard: In expectation there will be no substantive impact on the Lower Courts’ judgments, but when an impact does occur, the effect is striking and important.

Consider next the effect of standard cases on doctrinal complexity. Take, for example, case facts \( p \) on the right hand side of Figure 3.\(^{16}\) Hearing case \( p \) reveals outcome \( \psi(p) \), causing the bridge between case facts \( x \) and \( y \) to break into two. Of the new bridges formed, one must cross zero whereas the other must not. The significance in this fact is that hearing a new case cannot change the number of times that the doctrinal path crosses zero. Hence, the degree of doctrinal complexity must remain unchanged.

Although standard cases cannot affect doctrinal complexity, they always impact doctrine in a meaningful way. Regardless of the outcome realized, one bridge will cross zero and this crossing point will be different to what came before (with probability one). Thus, standard cases will

\(^{16}\)The logic described here does not depend on whether the cases are exploratory or exploitative.
inevitably change doctrine in a way that affects the Lower Courts’ behavior, even if this never manifests as an increase in doctrinal complexity.

**Proposition 12** Doctrinal complexity is unchanged when the Higher Court hears a standard case. Nevertheless, the doctrinal cut-point changes with probability 1.

In contrast with non-standard cases, the impacts of standard cases are, therefore, monotonously standard. Standard cases fine tune the law, so to speak, each and every time they are heard, yet they never fundamentally change the nature of doctrine. Standard cases are, in this sense, the worker-bees of doctrinal evolution: constantly working away at refining doctrinal precision but never inducing a paradigm shift in legal practice.

Finally, we consider whether the evolution of law ever comes to an end. On this question we can be definitive: the Higher Court will eventually cease hearing new cases (in this area of the law). It is possible that the addition of one precedent makes hearing another case more attractive; for instance, if the realized outcome increases the error uncertainty for a large set of cases. Nevertheless, it is always the case that outcome uncertainty decreases. It decreases for the case heard (uncertainty goes to precisely zero) and for every other case in the same region of the case facts space (i.e., between the nearest precedents). For no case facts does outcome uncertainty ever increase. Consequently, an inexorable pressure towards lower outcome uncertainty is applied as new cases are heard and the set of precedent grows. Inevitably, this comes to dominate any variation in error uncertainty, and the benefit of hearing new cases converges sufficiently close to zero to make hearing them not worthwhile (for even an arbitrarily small cost of hearing cases).\(^{17}\)

**Proposition 13** With probability one, the Higher Court stops hearing new cases in finite time.

The end of precedent implies that the court never arrives at complete knowledge and, more importantly, that judicial errors persist without end. Yet our result implies that the cost of these errors—at least in expectation—are not of sufficient concern to justify the cost in removing them. Moreover, the end of precedent implies that doctrinal complexity can increase no more. Thus, legal doctrine never reaches infinite complexity, always stopping at some finite level.

**Corollary 1** With probability one, doctrinal complexity stops increasing in finite time.

\(^{17}\)This result holds for any degree of forward-looking behavior by the court.
When the expansion of precedent does come to a halt, the end is sudden and final. Once the Higher Court chooses to not hear a new case, it never again begins hearing cases. This follows from the stationarity of the Court’s problem: If it chooses to not hear a case in one period, nothing is learned about the environment and the same problem—and thus the same choice—occur in the next period and every subsequent period.\textsuperscript{18}

8 Empirical Implications

Our model provides a novel framework for understanding the complex nature of judicial law-making through individual case resolution. It also yields a series of empirical implications that can inform future work, both as predictions that can be evaluated and as lessons for what can be inferred from observed patterns of judicial behavior. We consider two in particular.

The nature of doctrine. First, our model yields empirical implications that speak to debates about the breadth of judicial doctrine. In particular, there exists a debate about whether courts ought to create broad legal rules that apply to wide classes of cases or instead engage in minimalism, making decisions that are narrowly circumscribed to the particulars of the individual cases they decide. Some scholars argue that broad, deep judicial decisions can be made with positive normative and practical implications (e.g., Dworkin 1986). An example of such a case might be \textit{Brown v. Board of Education}, invalidating all racial segregation in public schools, or \textit{Miranda v. Arizona}, which held that individuals’ statements made while in police custody are only admissible in court if the defendant had been informed of his right not to speak. Other scholars argue, by contrast, that judicial decisions are best when minimalist (e.g., Sunstein 1999). Examples of minimalist decisions include \textit{US v Lopez}, a case that invalidated a national law that prohibited guns in school zones. What is more, beyond the normative literature, the question of whether and when courts are better off issuing broad or narrow decisions has been the subject of recent attention in the positive theoretical literature (Staton and Vanberg 2008, Lax 2012, Fox and Vanberg 2014, Clark Forthcoming). Our model, specifically Propositions 2 and 4, predicts that the breadth of doctrine will be a function of the factual relationship between the cases in which they are made and past

\textsuperscript{18}Identifying why an area of the law may reignite is worth exploring. Our results suggest that an exogenous change to the environment may be necessary. Another possibility is that the Higher Court toggles between different areas of the law as the value to hearing new cases in any particular area oscillates as precedent accumulates.
precedents. When working its way into new areas of the law, courts should be less reluctant to take a minimalist approach.

**Fact patterns and the path of law.** Second, our theoretical framework helps make sense of previously intractable theoretical dilemmas concerning what we can infer from the pattern of cases high appellate courts decide. Most specifically, scholars of the US judiciary have debated what can be learned from studying the select cases the US Supreme Court chooses to decide (e.g., Cross 1997, Friedman 2006). However, scholars have developed consequential theories of judicial decision-making that extrapolate from the relationship between case facts and judicial choice to over-arching conjectures about how courts construct law (e.g., Kritzer and Richards 2002).\(^\text{19}\) However, as Kastellec and Lax (2008) demonstrate, what we might infer about judicial preferences and rule-making from relationships between facts and voting patterns depends crucially on the underlying theory of case selection. When does the Court refine law by selecting factually similar cases? When does the Court reach out to new factual scenarios not previously considered? While our model cannot predict what the precise path will be through a set of case facts, it does provide a theoretical framework for understanding the incentives a court faces and, as a consequence, how a court will work its way through the path of law.

9 **Discussion and Conclusion**

Beyond the model’s empirical implications, we conclude with a number of normative issues implicated by our analysis and further questions one might extend our analysis to consider.

**Analogical reasoning.** As noted earlier in the paper, there exists controversy over the normative implications of analogical reasoning on the courts—namely, about whether it produces efficient and predictable outcomes. Analogical reasoning in our model takes a particular form, exhibiting several characteristics that resonate closely with experience. The first is which precedents are used and, more surprisingly, which are not used. Lower court judges in our model discard all precedents that are not the nearest. Practicing lawyers also do not invoke every precedent, instead limiting

\(^{19}\)Notably, much of this work has been criticized recently for limitations in its empirical strategy (Lax and Rader 2010).
themselves, to what are known as the *controlling* or *on point* precedents. These precedents—in line with the nearest precedents in our model—are those most similar to the case under consideration. To the standard intuition, we add the nuance that *directional* information also matters. Thus, the courts look for the nearest precedent in either direction, even if these cases are not necessarily the closest in an absolute sense. A second property that resonates with practice is how judges balance precedent. Judges in our model do not weight precedent equally. Rather, consistent with practice, the judges balance according to how similar—or relevant—are the particular case facts to the case at hand.

A third, and perhaps most important property, is that analogical reasoning in our model matches descriptions of the process as it works in practice. Legal formalists have long critiqued analogical reasoning (and legal realism more broadly) as lacking a coherent vision. This is a point that Sunstein, the leading proponent of analogical reasoning, concedes. Indeed, he goes further, observing that judges are often unable to even articulate the logic of their judgments. (Sunstein 1993, 747) writes:

> But it is characteristic of reasoning by analogy, as I understand it here, that lawyers are not able to explain the basis for these beliefs in much depth or detail, or with full specification of the beliefs that account for that theory. Lawyers (and almost all other people) typically lack any large scale theory.

Our model offers a potential resolution of this ambiguity. The lower court judges in our model similarly cannot articulate a broad vision or theory of the legal environment they operate in. Indeed, it is not even necessary that they possess any theoretical knowledge of the underlying environment. Yet we show that these justices are nevertheless able to operate efficiently by simply applying the power of analogy across existing precedent.

**Case and doctrinal reversals.** The results of our analysis rationalize two empirical regularities in judicial decision making. In particular, the expansion of precedent over time explains why courts might decide identical cases differently at different points in time. As precedent expands, Lower Courts may rule differently on the exact same set of case facts to what they ruled with a smaller precedential history. It is also possible for the Higher Court to overrule Lower Courts and reverse
judgments. However, it is not possible in our model for the Higher Court to reverse itself on particular case facts. Nevertheless, even in the unchanging environment we analyze, it is possible for the Higher Court to reverse doctrine. When hearing exploratory cases—cases outside of the range of factual scenarios previously considered—the Higher Court has the opportunity to alter the content of its doctrine (as well as to create carve-outs). Substantively, we might consider these instances as examples where new dimensions relevant to the Court’s evaluation of cases may present themselves (e.g., Gennaioli and Shleifer 2007). Consider our example of search and seizure. During the late-18th century, when the Fourth Amendment was written, the idea of wiretaps was beyond imagination. Later, as technology evolved, so too did Fourth Amendment jurisprudence. However, today we see new challenges arising that present factual scenarios outside of the realm of possibility during the mid-20th century, when doctrine developed to suit to types of searches and privacy questions that were then the universe of things to be considered. Today, cell phones and electronic communication present unique opportunities to distinguish questions that would have, before their time, been otherwise indistinguishable. (Alternatively, one might interpret such an example as an instance in which the assumed fixed interval of case facts changes.) These types of changes in the world could necessitate a reconsideration of precedent by the Higher Court in order to maintain a workable, effective doctrine. Such dynamics are also likely to have particular import in the context of another extension—divergence of preferences among courts—which we take up next.

Non-aligned preferences. By assuming judges share preferences over how to dispose of cases—i.e., the judges are a “team” (Kornhauser 1995)—we have isolated one particular tension for the Higher Court. Namely, we focus on the informational challenges in choosing cases to learn about and communicate doctrine under conditions of complex policy and uncertainty. Were the Lower Courts to diverge from the Higher Court with respect to the threshold for distinguishing between judgments (for example, the Lower Courts preferred \( j(p) = E \) if and only if \( \psi(p) > \epsilon > 0 \)), there may arise instances of agency loss or shirking on the part of the Lower Courts. As a consequence, we may encounter incentives to articulate doctrine that is not in line with the Higher Court’s own preferred threshold. This may involve endowing Lower Courts with authority either by entrusting them with doctrine-making capacity (e.g., Gailmard and Patty 2013) or crafting doctrine that contains room for discretion in rule-application (e.g., Staton and Vanberg 2008, Lax 2012). It may
also involve modifying the precise rule communicated (e.g., Carrubba and Clark 2012) or, more interestingly, deliberate obfuscation in the doctrine communicated. These actions may sway the Lower Courts toward the preference of the Higher Court, but themselves impose costs as the Lower Courts are restricted in their ability to reason by analogy. We expect the magnitude of the agency problem—in terms of the volume of cases, the degree of preference divergence, or the degree of informational asymmetry across the levels of the hierarchy—will condition how a court might use tools of discretion and insincere doctrinal articulation.

**Endogenous emergence of *stare decisis***. Finally, our model and potential extensions of the framework point provide analytic insight not just into why courts might sometimes reverse themselves but also into why courts strongly avoid reversing themselves. Indeed, the framework we propose suggests a rationale for the practice of *stare decisis*, the doctrine which holds that courts should not reconsider questions they have already answered. Stare decisis in practice is a stronger norm at higher levels of the judicial hierarchy than it is at lower levels. In the model, we note that we do not impose any exogenous requirement that the Lower Court follow precedent or adjudicate consistently with doctrine, although respect for precedent follows immediately from the assumption of common interest across the courts; nevertheless, the observed pattern emerges endogenously. Given the prominence of stare decisis in legal systems around the world, models of the judicial process that can integrate the complexity of law, the method of legal reasoning, and the endogenous support for multiple norms of institutional behavior mark a step towards more comprehensive theories of the law.

**References**


Kornhauser, Lewis A. 1995. “Adjudication by a Resource-Constrained Team: Hierarchy and Prece-

Law Cases: Does Chevron Matter?” Paper prepared for the annual meeting of the American
Political Science Association, Boston, MA.


Lax, Jeffrey R. and Kelly T. Rader. 2010. “Legal Constraints on Supreme Court Decision Making:


Princeton University working paper.

Cambridge, MA: Harvard University Press.


Proofs of Formal Results

This section contains the proofs of the formal statements. Most of the results follow directly from the arguments and mathematics developed in the text, and for these results we merely point to the relevant properties. Our principal focus here is on the remaining substantive results.

Proof of Lemma 1: The Lower Courts require the judgment history to follow precedent (see Definition 1), and as precedent applies only to case facts in the judgment history, it is immediate that the judgment history is sufficient. □

Proof of Proposition 1: The expression for $E[\psi(p)]$ is the standard expression for the expected outcome on a Brownian bridge. The decision rule follows from the independent calculations in Section 5 of the text. □

Proof of Proposition 2: The necessity and sufficiency are evident from inspection of the expression for $E[\psi(p)]$ in Proposition 1. □

Proof of Proposition 3: This is the analogue to Proposition 1 for case facts on the flanks of known points. The expression for $E[\psi(p)]$ is the standard expression for the expected outcome of a Brownian motion. The decision rule follows from the calculations in Section 5 of the text. □

Proof of Proposition 4: The sufficiency of the partial rule follows from inspection of $E[\psi(p)]$ in Proposition 3. The inclusion of the drift term, $\mu$, in the expression for $E[\psi(p)]$ implies that the minimalist truthful doctrine is insufficient. □

Proof of Theorem 1: The “if” direction follows from Propositions 1-4. For the “only if” direction, we require that analogical reasoning implies that the generating process is the Brownian motion. That only the nearest precedents in either direction are used implies that the generating process possesses the Markov property. Linear interpolation and extrapolation implies that each “innovation” in the mapping is independent and identically distributed. As the mapping is continuous, these conditions collectively define the Brownian motion. □

Proof of Proposition 5: The result follows from the calculations in Equation (3) and the surrounding discussion. □
Proof of Proposition 6: Begin by considering case facts on the bridge between \((x, t)\) and \((y, b)\), where \(t > 0 > b\) and \(x < y\). The analysis for exploratory cases is analogous. Without loss of generality, normalize the end points of the bridge to \((0, 1)\) and \((1, b)\), with \(b \leq -1\) (so that the right-end of the bridge is lower than the left-end is higher), and the 0-threshold is crossed at some \(p^* \leq \frac{1}{2}\). We proceed point-by-point. Suppose case facts \(p \in (x, y)\) is heard and consider the impact on some \(q \in (x, y)\), where most frequently we have \(q \neq p\). We demonstrate that the marginal gain in utility at \(q\) is strictly increasing in \(\sigma\), and thus it is in expectation across all \(q\) and the result follows. Note that Proposition xx implies the benchmark utility is independent of \(\sigma\) and for this exercise it is sufficient to compare the level of expected utility in \(\sigma\).

Upon hearing case facts \(p\), the bridge between \(x\) and \(y\) is split into two bridges, one between 0 and \(p\) and the other between \(p\) and 1. The expected outcome of \(p\) is the realization of a normally distributed random variable of variance \(p \cdot (1 - p) \sigma^2\) (and the outcome is observed precisely). The expected outcome for all other case facts are now given by the points on the newly formed bridges. As the bridges are linear, the expected outcome for all other case facts change according to a random variable with standard deviation proportional to how far along the bridge the point is. (The key part of the logic here is that although the value of \(\psi(q)\) is not observed precisely for \(q \neq p\), the expected value is updated with the revelation of \(\psi(p)\) in such a way as if it had, albeit with a signal of lower precision.)

Formally, for case facts \(q \in (0, p)\) the change in the value of \(E[\psi(q)]\) is given by a normal distribution of mean 0 and standard deviation: \(\sigma_q^p = \frac{q}{p} \sigma \sqrt{(1 - p)}\). And for case facts \(q' \in (p, 1)\) the change in the value of \(E[\psi(q')]\) is given by a normal distribution of mean 0 and standard deviation: \(\sigma_{q'}^p = \frac{1 - q'}{1 - p} \sigma \sqrt{(1 - p)}\). These expressions are both strictly increasing in \(\sigma\).

The analysis of points on both bridges is identical, so without loss of generality, consider case facts \(q \in (0, p)\). Proposition 1 established that the decision rule follows the realization of \(E[\psi(q)]\). Following the expressions and logic in Section 5, utility is gained when a realization is observed on the opposing side of the zero-threshold. With linear utility, the expected gain from receiving a signal is the expected value of outcomes on the opposing side of zero multiplied by two (as the judgment at \(q\) is changed and the payoff—which is of constant amount—goes from a negative to a positive). The expected outcome at \(q\) is \(\mu^q = 1 - q (1 + b)\), and the standard deviation is \(\sigma_q^p\), from above. Supposing without loss of generality that \(\mu^q \geq 0\), the marginal utility at \(q\) of hearing \(p\) is
given by:

\[ V_q^p = 2 \int_{-\infty}^{0} -z \phi(z|\mu_q, \sigma_q^p) \, dz = 2 \int_{-\infty}^{0} -z \frac{1}{\sigma_q^p \sqrt{2\pi}} e^{-\frac{(z-\mu_q)^2}{2(\sigma_q^p)^2}} \, dz. \]  

(4)

Where \( \phi(z|\mu_q, \sigma_q^p) \) is the pdf of the normal distribution with mean \( \mu_q \) and standard deviation \( \sigma_q^p \).

Renormalizing around zero for clarity, and integrating, this becomes:

\[ V_q^p = 2 \int_{\mu_q}^{\infty} (z - \mu_q) \frac{1}{\sigma_q^p \sqrt{2\pi}} e^{-\frac{z^2}{2(\sigma_q^p)^2}} \, dz = 2 \sigma \sqrt{\frac{2}{\pi}} e^{z^2 \sigma^2} (1 - \Phi(z|0, \sigma_q^p)) \]

(5)

Differentiating with respect to \( \sigma_q^p \) and simplifying:

\[ \frac{dV_q^p}{d\sigma_q^p} = \frac{2e^{-\frac{\mu_q^2}{2(\sigma_q^p)^2}}}{\sigma_q^p \sqrt{2\pi}} \left( 1 + \frac{\mu_q^2}{(\sigma_q^p)^2} - \frac{\mu_q^2}{(\sigma_q^p)^2} \right) = \frac{2e^{-\frac{\mu_q^2}{2(\sigma_q^p)^2}}}{\sigma_q^p \sqrt{2\pi}} > 0. \]  

(6)

As \( \sigma_q^p \) is increasing in \( \sigma \), we thus have: \( \frac{dV_q^p}{d\sigma} > 0 \), completing the argument. \( \Box \)

**Proof of Proposition 7:** We continue the approach in the proof of Proposition 6, this time we compare points pairwise. Recalling from the previous proof the values of \( \sigma_q^p \) and \( \sigma_q^p' \), we have:

\[ \frac{d}{dp} \left( \sigma_q^p \right) = \frac{d}{dp} \left( \sigma_q \sqrt{\frac{1-p}{p}} \right) = \frac{d}{dp} \left( \sigma_q \sqrt{\frac{1-p}{p}} \right) = -\frac{q\sigma^2}{2} \sqrt{\frac{1-p}{p}} \cdot \frac{1}{p} < 0 \]

\[ \frac{d}{dp} \left( \sigma_q^p \right) = \frac{d}{dp} \left( \sigma \frac{1-q}{1-p} \sqrt{\frac{1-p}{p}} \right) = \frac{d}{dp} \left( \sigma \frac{1-q}{1-p} \sqrt{\frac{1-p}{p}} \right) = \sigma \frac{1-q}{2} \sqrt{\frac{1-p}{p}} \cdot \frac{1}{(1-p)^2} = \sigma \frac{1-q}{2} \sqrt{\frac{1-p}{p}} \cdot \frac{1}{(1-p)^2} > 0 \]

Now compare two points: \( x_1 = \hat{p} - \Delta \) and \( x_2 = \hat{p} + \Delta \), for \( \Delta \leq \hat{p} \), where \( x_1 \) is on the left bridge and \( x_2 \) on the right-side bridge. The above expressions then become:

\[ \frac{d}{dp} \left( \sigma_q^p \right) = -\frac{x_1}{2} \sqrt{\frac{1}{p(1-p)} \cdot \frac{1}{p}}, \quad \text{and,} \quad \frac{d}{dp} \left( \sigma_q^p \right) = \sigma \frac{x_2}{2} \sqrt{\frac{1}{p(1-p)} \cdot \frac{1}{(1-p)}} \]

42
Taking the ratio:

\[- \frac{d}{dp}(\sigma^p_{x_1}) \frac{d}{dp}(\sigma^p_{x_2}) = \frac{x_1}{1-x_2} \frac{1-p}{p} = \frac{\hat{p} - \Delta}{1 - \hat{p} - \Delta} \frac{1-p}{p} \tag{7}\]

At \( p = \hat{p} \), we have:

\[- \frac{d}{dp}(\sigma^p_{x_1}) \frac{d}{dp}(\sigma^p_{x_2}) = \frac{\hat{p} - \Delta}{1 - \hat{p} - \Delta} \frac{1-\hat{p}}{\hat{p}} = \frac{1}{1-\hat{p}} \leq 1 \text{ if } \hat{p} \leq \frac{1}{2}. \tag{8}\]

With the inequality strict if \( \hat{p} < \frac{1}{2} \). This implies that the marginal decrease in standard deviation at \( x_1 \) is dominated by the marginal increase in standard deviation at \( x_2 \). At \( \hat{p} = p^* < \frac{1}{2} \), the distributions at \( x_1 \) and \( x_2 \) have means that are equally distant from zero but with differing standard deviations, specifically \( \sigma^p_{x_2} > \sigma^p_{x_1} \) as \( p^* < \frac{1}{2} \). Differentiating Equation (6):

\[ \frac{d^2 V^p_{q}}{d (\sigma^p_{q})^2} = \frac{2e^{-(\mu_q)^2/2(\sigma_q)^2} (\mu_q)^2}{\sqrt{2\pi}} (\sigma_q)^3 > 0. \tag{9}\]

This says that a unit increase in standard deviation is more valuable at \( x_2 \) than at \( x_1 \). Combined with Equation (8), we have that at \( p^* \), an increase in \( p \) increases utility at \( x_2 \) by more than it decreases utility at \( x_1 \). Aggregating across all matched pairs, we conclude that expected utility is increasing in \( p \) for all case facts in \((0, 2p^*)\). The remaining case facts, \((2p^*, 1)\), are on the right-side bridge and have increasing variance in \( p \), which by Proposition 6 delivers increasing expected utility in \( p \). Thus, \( p^* \) cannot be the optimal \( p \) to hear. The reverse logic holds when \( p^* > \frac{1}{2} \). This leaves only \( p^* = \frac{1}{2} \), which requires \( b = -1 \), a zero probability event. The proposition follows. \( \square \)

**Proof of Proposition 8:** Part (iii) follows from the proof of Proposition 7 and the observation that at \( t = |b| = 1 \), the ratio in Equation (7) is exactly 1 and the matched points \( x_1 \) and \( x_2 \) have equal standard deviation and collectively span the entirety of the bridge \([0, 1]\). Cases (i) and (ii) are perfectly analogous, so without loss of generality consider case (i), the case analyzed in the preceding proof, and suppose \( \hat{p} < p^* \). The inequality in Equation (8) continues to hold and is strict. The comparison of \( x_1 \) and \( x_2 \) now involves different standard deviations and expected values that
are not equally distant from the zero threshold; specifically, \( \mu_{x_1} > |\mu_{x_2}| \). Differentiating Equation (6) with respect to \( \mu_q \), however, gives:

\[
\frac{d^2V^p_q}{d\sigma^p_q d\mu_q} = 2e^{-\frac{(\mu_q)^2}{2(\sigma^p_q)^2}} \frac{-\mu_q}{\sqrt{2\pi}(\sigma^p_q)^2} < 0
\]

(10)

This, combined with Equation (9), implies that the marginal value of an increase in standard deviation is greater at \( x_2 \) than at \( x_1 \). Thus, the logic of the previous proof at \( p = p^* \) generalizes to all \( \hat{p} < p^* \). A rearrangement of terms shows that a similar dominance relationship exists for all \( \hat{p} > \frac{1}{2} \), and, therefore, the optimal \( p \) must lie in the interval \(( p^*, \frac{1}{2} )\), as required. □

**Proof of Proposition 9:** The result on \( p^*_1 \) follows from the proof of Proposition 8. The result on \( \sigma^2 \) follows from the proof to Proposition 6. For \( \psi(0) \), compare case facts \( p > 0 \) for \( \bar{\psi}(0) \) to case facts \( p' = p + \frac{\omega}{\mu} \) for \( \bar{\psi}(0) + \omega \) when \( \omega > 0 \) (and recalling that \( \mu < 0 \)). \( p \) and \( p' \) have the same expected outcome, yet the variance at \( p' \) is \( \frac{\omega}{\mu} \) larger than at \( p \). As a \( p' \) exists for every \( p \), the proof of Proposition 6 implies that hearing a case at \( \bar{\psi}(0) + \omega \) is strictly more profitable than at \( \bar{\psi}(0) \).

**Proof of Proposition 10:** The path of \( \mathbb{E} [\psi(p)] \) is continuous, from Propositions 1 and 3. Consider a Brownian bridge that crosses the zero threshold (i.e., cases on the bridge are standard). Linearity of the bridge implies the crossing is unique. Hearing a case on the bridge splits the bridge into two. Linearity and continuity implies that, for both bridges combined, a unique crossing exists and doctrinal complexity is unchanged. For non-standard cases the bridge does not cross zero. Obviously, doctrinal complexity cannot decrease and a realization as in Figure 3 demonstrates the possibility of an increase. These arguments also hold for exploratory cases by applying the fact that \( \mathbb{E} [\psi(p)] \to \pm\infty \) as \( p \to \pm\infty \). □

**Proof of Proposition 11:** The result is by construction following the argument in the text. □

**Proof of Proposition 12:** That doctrinal complexity is unchanged for standard cases is proven in the proof to Proposition 10. For the doctrinal cut-point to remain unchanged, we require \( \psi(p) = \mathbb{E} [\psi(p)] \), a zero probability event. □
Proof of Proposition 13: Suppose the Higher Court hears cases until a bridge is formed that crosses zero (i.e., until a standard exploitative case exists); by Proposition 9 this happens with probability greater than \( \frac{1}{2} \) each period and, therefore, with probability one in finite time. Consider now experimenting on the bridge. It is straightforward to show that the expected utility from a bridge is linear in the width of the bridge. As the width of the bridge approaches zero, therefore, expected utility approaches zero and the variance of any case facts on the bridge also approach 0. It follows that the value of hearing new cases on the bridge approaches 0 as more cases are heard, and, for fixed cost of hearing cases, the Higher Court eventually stops in finite time. The same logic applies to any bridge that does not span zero (and contain non-standard cases). The Court can then return to hearing cases on the flanks and the process just described iterates. By construction, the right-most precedent has \( \psi(\bar{p}_r) < 0 \) and for hearing cases to be worthwhile, \( p - \bar{p}_r \) is bounded away from zero. With probability one, therefore, a \( \psi(p) \) is eventually realized that is sufficiently distant from 0 to make hearing further cases unprofitable. The same logic applies to cases \( p < \bar{p}_l \) and with probability one the Court stops hearing new cases in finite time. Finally, note that this argument does not require the clean separation of stages (the Court can switch between exploratory and exploitative cases and the logic still applies). □

Proof of Corollary 1: Immediate from Proposition 13. □