Explaining Confusion in the Law

JBrandon Duck-Mayr*

Abstract

Judges, scholars, and commentators decry confusing areas of judicially created policy. This could hurt courts' policy making efficacy, so why do judges allow it to happen? Existing accounts explain inconsistent decisions, but not inconsistent rule formulation. I show policy can become confusing when judges explain rules in more abstract terms than they decide cases. To do so, I expand standard case-space models of judicial decision making to account for relationships between specific facts and broader doctrinal dimensions. This model of judicial decision making as a process of multi-step reasoning reveals that preference aggregation in such a context can lead to inconsistent collegial rules. I also outline a class of preference configurations on collegial courts in which this problem cannot arise.

Word count: 7,042

*PhD Candidate, Department of Political Science, Washington University in St. Louis. I would like to thank Randall Calvert, Keith Schnakenberg, James F. Spriggs, II, Morgan Hazelton, Jordan Carr Peterson, and participants at Washington University in St. Louis' Formal Theory Workshop and at the 2019 Midwest Political Science Association Annual Meeting.

A wide range of observers have noted particularly confusing rules being produced by courts across several areas of the law. For example, legal scholars complain the U.S. Supreme "Court's numerous [federal] preemption cases follow no predictable jurisprudential or analytical pattern" (Dinh 2000).[1] Political commentators criticize the Court's "Establishment Clause decisions that have been, in the words of Alice in Wonderland, curiouser and curiouser," and hope the Court will "leaven with clarity the confusion it has sown" (Will 2019). Supreme Court Justice Clarence Thomas bemoans "an Establishment Clause jurisprudence in shambles," claiming the Court's "jurisprudence has confounded the lower courts and rendered the constitutionality of displays of religious imagery on government property anyone's guess..." (*Utah Highway Patrol Assoc. v. American Atheists Inc.*, 565 U.S. 994 (2011) at 994, Thomas, J., dissenting).

Courts' policies are implemented by others, from lower courts applying appellate court rules, to outside actors enforcing judicially created policies (Maltzman, Spriggs and Wahlbeck 2000, 5). When courts' rulings are unpredictable, and their rules are confusing, it impedes these actors' ability to implement judicial policies. Moreover, inconsistency in legal doctrine reduces judicial legitimacy (Landa and Lax 2009, 959). Why would courts create confusing policies that endanger judicial legitimacy and their efficacy as policymakers? Perhaps judges are free to act relatively unconstrained (e.g. Segal and Spaeth 2002), and current court members simply prefer outcomes inconsistent with prior cases. Or perhaps courts' decisions are well explained by pronounced rules, even when scholars and commentators believe an area of the law is in disarray (Segal 1984). Maltzman, Spriggs and Wahlbeck (2000) explain that bargaining over opinion content among justices may produce results inconsistent with what we might otherwise expect. However, none of these accounts explain why courts' *descriptions* of their decision rules do not provide clear guidance for lower court judges and other policy enforcers.

I use a social choice theoretic model to show preference aggregation on collegial courts can result in confusing rules when judges communicate policy in terms of subjective criteria

[1] Drahozal (2004) gives a book-length review of the confusion rampant in federalism cases.

that depend on objective facts.[2] That is, judges often explain rules using a low number of abstract determinations that in turn are derived from specific facts of cases. I show this kind of multi-step reasoning in appellate review can result in inconsistent collegial rules.

For example, in Fourth Amendment search and seizure cases, the constitutionality of police conduct can depend on (1) the intrusiveness of the search or severity of the seizure, and (2) whether the police had the requisite level of suspicion (e.g. probable cause) required to support such conduct. The court must determine how intrusiveness and police suspicion translate into outcomes, and further use the specific facts of cases to determine the level of police suspicion: "As the Court recognizes, determinations of probable cause and reasonable suspicion involve a two-step process. First, a court must identify all of the relevant historical facts . . . and second, it must decide whether . . . those facts would give rise to a reasonable suspicion justifying a stop or probable cause to search" (*Ornelas v. United States*, 517 U.S. 690 (1996) at 700–701, Scalia, J., dissenting).

When courts engage in such multi-step reasoning, opportunity for inconsistency in the resulting collegial rules arises, even when all the judges possess well-behaved preferences. The problem arises because with multiple levels of judgment or preference aggregation, judges can agree on outcomes while disagreeing on the proper justification for that outcome, so that applying the reasoning relied on by a majority coalition in any one case can be inconsistent with collegial outcomes in other cases. This source of inconsistency in the law has so far gone unnoticed despite related results in the literature (e.g. Kornhauser 1992; Landa and Lax 2009) because models have left unexplored the interaction between disagreements over doctrine and disagreements over intermediate legal determinations.[3]

After a short survey of the substantive literature, I provide a brief overview of related models before detailing the setup of a model that allows for courts' multi-step reasoning. I then show why confusion in the law can result when appellate courts communicate policy

[2]For formal statements of results and proofs of propositions, see Appendix A.

[3]That is, while Kornhauser (1992) allows disagreement over intermediate determinations, and Lax (2007) and Landa and Lax (2009) allow disagreement over doctrine, neither allow disagreement over both.

this way, as well as when they can safely do so while maintaining clear policy; I illustrate these results with a simple Fourth Amendment example.

1 Causes and Consequences of Confusion in the Law

If an appellate court's "jurisprudence [confounds] the lower courts" and makes the proper decision in future cases "anyone's guess" (*Utah Highway Patrol Assoc.*, 565 U.S. at 994, Thomas, J., dissenting), the court will be less effective as a policy maker. Such confusion also raises normative concerns—crafting a confusing doctrine leaves citizens potentially less empowered to assert their rights (since they can't tell when they apply). Nevertheless, legal scholars highlight time and time again various doctrines that have grown inconsistent, from death penalty jurisprudence (Robinson and Simon 2006) to First Amendment jurisprudence (Post 1995) to federalism jurisprudence (Drahozal 2004). Unfortunately, a theory of why doctrine so often becomes inconsistent has so far been lacking.

Empirical work has well documented the effects of unclear doctrine on courts' policy-making efficacy. Spriggs (1996) argues administrative agencies will be more likely to follow Supreme Court opinions that offer clearer guidance, and finds evidence that agencies more closely follow opinions that were more specific and explicit. Westerland et al. (2010) hypothesize that unclear signals from the U.S. Supreme Court will lead to lower compliance by the appellate courts, finding an increased number of concurrences indeed reliably correlated with lower compliance.

Empirical work has also uncovered some causes of inconsistency or complexity in judicial behavior. Collins (2008) finds individual justices' choices are more variable in complex cases. Owens and Wedeking (2011) use text analysis methods to measure the cognitive complexity of court decisions,[4] finding, for example, that some justices provide clearer guidance in their opinion than others on average, and that majority opinions are less clear than dissents,

[4]Though they acknowledge that *doctrinal complexity*—the topic of this article—is another aspect of clarity of Supreme Court opinions (Owens and Wedeking 2011, 1038).

perhaps due to the bargaining entailed in crafting a binding precedent (1032–1033; Maltzman, Spriggs and Wahlbeck 2000).

Related theoretical work includes the discovery of the "doctrinal paradox" (Kornhauser 1992) and its extension (Landa and Lax 2009),[5] as well as work on rules vs. standards (e.g. Clark 2016; Lax 2012). The doctrinal paradox shows that outcomes depend on whether judges on collegial courts decide cases by majority vote over outcomes or by majority vote over intermediate determinations, such as whether police had probable cause. Interestingly, Kornhauser (1992, 447) explicitly envisions the cases as coming from a fact space that the judges must then map to these intermediate conclusions, but does not model *how* the judges make these intermediate determinations at all; accounting for this step in judicial reasoning is one of the principal technical contributions of this article.

The rules vs. standards literature tackles a separate but related issue to the doctrinal inconsistency I study. These studies seek to explain when judges will issue specific policies and when they will use vague policy. For example, Staton and Vanberg (2008) shows courts may use vague rules to prevent observed noncompliance with rulings by ideologically divergent governments or to allow leeway to governments that are ideologically aligned with the court.

Most on point for the present article in this vein are Clark (2016) and Lax (2012). Clark studies the trade-off between an opinion that clearly disposes of cases closely related to the present case and an opinion that is less precise but has more impact on dissimilar cases. Clark finds judges will be more precise when the instant case is most representative of potential disputes and when they anticipate being able to issue additional clarifying rulings in the future. This analysis starts from the important point that judges generally cannot specify a complete mapping from cases to outcomes in a single opinion. The import of Proposition 2 below, detailing the general susceptibility of doctrine to inconsistency, involves this issue; inconsistency has real bite precisely when judges cannot perfectly map every potential future

[5]Study of the phenomenon identified by Kornhauser (1992) as the doctrinal paradox spread throughout legal theory and social choice theory and became known also as the "discursive dilemma" (e.g., List 2012; List and Pettit 2002; Nehring and Puppe 2006, 2010).

dispute to an outcome.

Lax (2012) considers the ability of an appellate court to promulgate a bright-line rule that depends only on an objective fact, or a standard based also on a subjective dimension such as severity of the weather. In this context, we may say bright-line rules are specific or precise, whereas standards based on a subjective dimension are less precise, either because the Court cannot perfectly observe the subjective dimension or because it is difficult to specify doctrinal requirements on that dimension. In the first case, standards are preferred despite their vagueness when the ability to observe the subjective dimension is relatively higher, or there is lower risk of ideologically opposed lower courts. In the second, standards can be attractive despite imprecision if the weight placed on the subjective dimension in the Court's preferences is high enough, or if the cost of writing more precise opinions is low enough. This provides a nuanced account of incentives to rely on potentially vague doctrine, but again, does not wrestle with inconsistency in doctrine.

In sum, a curious hole is left in the literature relating to legal confusion. Evidence exists that courts' policy-making efficacy depends on legal clarity, and normatively we may expect courts to consistently interpret legal rights. Empirical work has uncovered some correlates of lack of clarity in the law, and theoretical work has shown conditions under which judges may choose vagueness over precision and clarity. However, the literature currently lacks a theory of doctrinal inconsistency. I extend models of case-based adjudication (Kornhauser 1992) and rulemaking (Lax 2007) to show an explanation for inconsistent doctrine embedded in legal reasoning: Judges generally engage in multiple steps of judgment aggregation, and this multi-step reasoning provides more opportunity for inconsistency in aggregation than previous models have accounted for.

2 Rule Making on Collegial Courts

I use a case space model to study rule making on collegial courts (Lax 2011). A case space model considers the set of all possible cases, or factual scenarios, a court could be presented with, and represents judicial policy as dividing that space into outcomes. That is, the set of possible cases is divided into two sets: the set of cases where plaintiffs win and the set of cases where defendants win; or, the set of cases where government activity is permissible, and the set where it is unconstitutional.

In a traditional case space model, the court is presented with a case $x \in X \subseteq \mathbb{R}^n$, the set of all possible cases the court could hear.[6] Each judge j then has a preferred *rule* mapping cases to outcomes $\rho_j : X \to \{-1, 1\}$.[7] The dimensions of X are interpreted as "whatever facts might be considered relevant to the judges" (Landa and Lax 2009, 593). Often models consider these facts to be high-level doctrinal concerns, such as the intrusiveness of a police search (Clark and Carrubba 2012), or sometimes specific "historical" facts, such as the speed at which a car is travelling (Lax 2012).

I will use as a running example the constitutionality of a seizure of a person—an investigatory stop or an arrest—under the Fourth Amendment.[8] The Fourth Amendment to the U.S. Constitution provides the "right of the people to be secure ... against unreasonable searches and seizures, shall not be violated..." (U.S. Const. Amend. IV). However, courts "must evaluate the reasonableness of a particular search or seizure in light of the particular circumstances" (*Terry v. Ohio*, 392 U.S. 1 (1968) at 21). For example, while arrests require probable cause, investigatory stops are less intrusive seizures that require only "reasonable suspicion" (*Terry*).

So, we might think of the case space dimensions as the doctrinal concerns of the level of

[6]The dimensions of the case space could be the set of real numbers (e.g. Clark and Carrubba 2012), real intervals such as $[0, 1]$ (e.g. Lax 2007), or discrete sets such as $\{0, 1\}$ (e.g. Landa and Lax 2009); the space may be unidimensional (e.g. Hübert 2019) or multidimensional (e.g. Badawi and Baker 2015).

[7]The dichotomous outcomes are sometimes presented with other labels, such as Y and N (e.g. Landa and Lax 2009).

[8]Fourth Amendment doctrine is a familiar example both to empirical (e.g. Segal 1984) and theoretical (e.g. Clark and Carrubba 2012) studies of case-based judicial decision making.

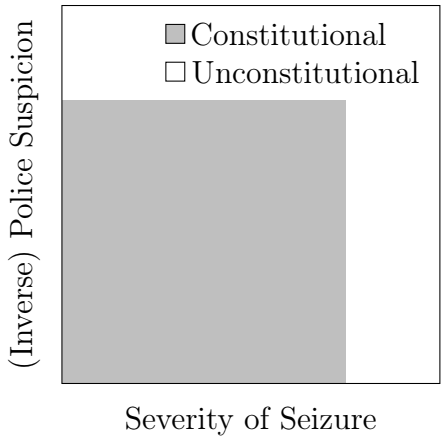
police suspicion and severity of the seizure; an example of a rule in such a space is depicted in Figure 1a. In this example, there are some seizures so severe they could never be found constitutional, some circumstances under which there is so little evidence of criminality that no seizure could be constitutional, but as long as the seizure is sufficiently not severe and the police have sufficient certainty that criminal conduct has occurred, the judge will find the seizure was constitutional.

Judges on collegial courts decide cases by majority rule over dispositions. The *implicit collegial rule*, or ICR, is the mapping between cases and outcomes that results from these majority votes over outcomes (Lax 2007, 595). In other words, the ICR represents "the law."[9] An example of a three judge panel's individual preferences and the resulting ICR is depicted in Figure 1b. In this case the judges' preferences aggregate to an ICR in which for the lowest range of police suspicion, no seizure is warranted, for a moderate range of police suspicion low levels of seizure are permissible, and at the highest range of police suspicion a much broader range of seizures are found constitutional.

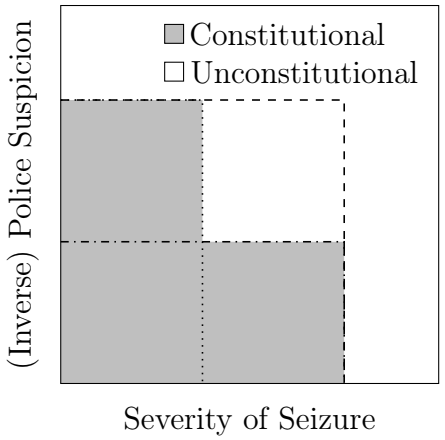
3 Model

Case space dimensions that capture high level doctrinal concerns are generated from historical facts, as Justice Scalia discusses in the *Ornelas* exerpt quoted in the introduction. As Lax (2007) explains, "in equal protection cases ... the dimensions might include (1) how 'suspect' the class invoked is ... (2) how compelling the state interest is ... and (3) how necessary the classification is .... Or, these dimensions could be broken down further" (594). While the technology of traditional case space models can be used to model decisions based on historical facts, doctrinal concerns, or both, it lacks the ability to model the relationship *between* doctrinal concerns and the dimension of historical facts they are derived from. Abstracting away from this relationship is useful for analyzing other aspects of judicial decision making.

[9] I join Justice Oliver Wendell Holmes, Jr. in claiming, "The prophecies of what the courts will do in fact, and nothing more pretentious, are what I mean by the law" (Holmes 1897, 461).



(a) Judge 1's rule. The shaded region is the set of cases the judge finds the police seizure to be constitutional and in the unshaded region the judge finds the police conduct unconstitutional.



(b) The ICR. Dashed and dotted lines mark the set of cases in which each judge rules the police conduct constitutional. The set of cases the court as a whole rules the police conduct was constitutional is shaded in gray.

Figure 1: An example individual rule and ICR for Fourth Amendment police seizure cases. The case space is comprised of two dimensions: severity of the police seizure, where larger values indicate a more intrusive seizure, and inverse police suspicion, where larger values indicate less certainty that criminal conduct has occurred.

However, to understand why outside observers are confused by judicial doctrine, it will be useful to separately represent the high dimensional space of all possible historical facts and the lower dimensional doctrinal space, and the relationship between these spaces.

A legal case presented to a court can be uniquely identified by its historical facts, such as whether a person seized by the police was placed in handcuffs or not, or how long a person was detained. We will say there are N potentially relevant dimensions of historical facts, so that $H \subseteq \mathbb{R}^N$ is the set of all possible combinations of historical facts.

A set of judges J (with $|J|$ odd) must decide cases presented to it from H, and assign them one of two outcomes $\{-1, 1\}$. So, as in other case space models, we will discuss policy as a partition of cases into outcomes. However, judges (and the public they communicate policies to) do not think about policy by considering every possible combination of historical facts, even if they could. They think about and communicate policy in more abstract terms informed by the historical facts, such as the severity of a police seizure or the degree of

police certainty of criminality that supports the seizure. So we also need to define a lower dimensional doctrinal space, $D \subseteq \mathbb{R}^n$, with $1 < n < N$.[10] Then each judge j has a preferred doctrine δ_j mapping D to $\{-1, 1\}$. A doctrine is *monotonic* if for any two points $d, d' \in D$, $d_i \geq d_i{}' \; \forall \; i$ implies $\delta(d) \geq \delta(d')$. We will assume the judges (and other relevant actors such as the public or lower court judges attempting to comply with the collegial appellate court's rulings) can "consistently label" the dimensions of H and D such that higher values of any h_k or d_i should lead to a weakly higher outcome, all else equal.

Unfortunately, as we will see, judges can disagree not only over doctrine, but how historical facts map onto doctrinal facts.[11] Not only could judges disagree whether a particular type of police seizure needs to be supported by probable cause or only by reasonable suspicion, but they could disagree about whether the historical facts support a finding of probable cause or not. So, we add the last moving part to the model: each judge j maps historical facts on to D; I will call this mapping a "fact finding function" $f_j : H \to D$.[12] For convenience, for a case h and a fact finding function f_j, we will write d_{ij} to mean the ith element of $f_j(h)$. A fact finding function is *monotonic* if for any two points $h, h' \in H$, $h_k \geq h'_k \; \forall \; k$ implies $d_{ij} \geq d'_{ij} \; \forall \; i$. For the remainder of the article, I assume all f_j and δ_j are monotonic.

In sum, each judge's preferred disposition is thus determined by $\delta_j(f_j(h))$; the judge is

[10] A one dimensional doctrinal space could be possible but would be rare. Some case space models present a simplified unidimensional case even when the lower dimensional abstract doctrine space is still multidimensional. For example, Clark and Carrubba (2012) discuss a unidimensional case space for police search cases, where the dimension is the intrusiveness of the search, when in fact such a formulation must (at least) be some combination of the search's intrusiveness and police certainty of criminality. (Consider that a search at intrusiveness level x that would be constitutional given probable cause may still be unconstitutional in the absence of probable cause). Such an abstraction is useful for studying some questions but not for examining why judicial policy can appear inconsistent to the citizens and other judges charged with following them.

[11] Note that Landa and Lax (2008) also discuss the various types of disagreements which judges on a collegial court may face. They make connections from related models to a couple of aspects of disagreement over the mapping between historical facts to doctrinal facts, though none directly address the difficulty discussed in this article.

[12] In practice, the judges would be presented with not only the lower court's determination of the case's placement in H given the evidence presented at trial, but also D, but we will ignore for now the lower court's determination of a case's placement in doctrinal space. I discuss appellate courts' deferenece to trial courts' findings as to a case's placement in H and D in Appendix B. In short, appellate judges generally defer to trial courts regarding historical facts, though exceptions can apply in constitutional cases (Hoffman 2001; Redish and Gohl 2017), but are less deferential regarding doctrinal concerns. Readers interested in this procedural issue should consult Appendix B.

presented with the historical facts, they determine how those facts relate to the doctrinal dimensions they find relevant, and thus how the case should be decided according to their preferred doctrine. Thus, a judge's preferred *rule*, or mapping from unique cases to outcomes, is a pair $\rho_j = (f_j, \delta_j)$. This process is depicted in Figure 2.

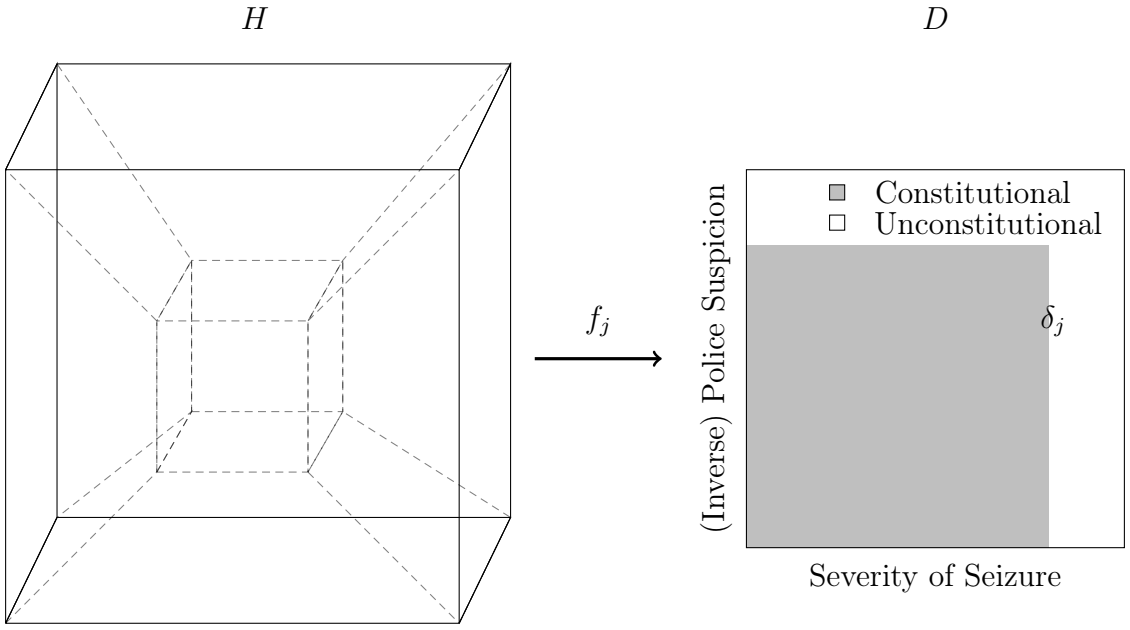


Figure 2: Assigning outcomes by translating a fact space to a doctrine space. A judge j is presented with a set of historical facts, a point in a pontentially high dimensional space H. Cases in this issue area are discussed using broader doctrinal terms—the lower dimensional space D. So, the judge uses the function f_j to translate the case from a point in H to a point in D, the space in which she describes her preferred partition (δ_j) of cases into -1 outcomes and 1 outcomes.

Judges decide cases by majority vote over outcomes. Similarly to Landa and Lax (2009), define an *outcome set* as specifying the outcome (-1 or 1) with each case $h \in H$, and the *collegial outcome set* as the outcome set formed by majority voting among J over the outcome in each case h. A *consistent rule* is a rule $\rho = (f, \delta)$ such that f is monotonic and δ is monotonic in f. The *implicit collegial rule* (ICR) is the rule $\rho_m = (f_m, \delta_m)$ constructed as follows: f_m takes the (dimension by dimension) median value of the f_j for every j in the majority coalition for every $h \in H$; and δ_m maps D_m to $\{-1, 1\}$ using the collegial outcome set. A summary of notation used is presented in Table 1.

Table 1: Notation Used

j	A judge on the collegial appellate court.
J	The set of judges on the collegial appellate court.
H	The set of all possible combinations of historical facts.
H_k	One of the N dimensions of H.
D	The set of all possible combinations of doctrinal determinations.
D_i	One of the n dimensions of D.
f_j	The mapping from historical facts to doctrinal dimensions as seen by judge j.
δ_j	The mapping from doctrinal determinations to outcomes preferred by judge j.
ρ	A pair (f, δ) mapping H to outcomes through D such that the outcome in case h is $\delta(f(h))$.
ρ_m	The implicit collegial rule (f_m, δ_m), where f_m takes the (dimension by dimension) median value of the f_j for every j in the majority coalition for every $h \in H$ and δ_m maps D_m to $\{-1, 1\}$ using the collegial outcome set.

4 Multi-step Reasoning Can Create Confusing Rules

Let us start with the simplest case, where the judges happen to agree on doctrine; that is, δ_j is the same for all j.[13] For example, suppose the judges agree that some seizures of a person are never justified, probable cause is needed to justify others, and that some seizures can be justified merely by reasonable suspicion, but that the judges disagree on the set of historical facts that support a finding of probable cause or reasonable suspicion.

Three types of doctrines in particular will be of interest, both because they are common types of legal doctrines and because of their aggregation properties. Call a doctrine δ such that

$$\delta(d) = \begin{cases} 1 & \text{if } d \cdot w \geq \tau \\ -1 & \text{otherwise,} \end{cases}$$

where τ is a scalar threshold and w is a vector of weights on the dimensions of D a *balancing*

[13] This is the case examined by, for example, Kornhauser (1992).

test.[14] A doctrine δ such that

$$\delta(d) = \begin{cases} 1 & \text{if } d_i \geq \tau_i \ \forall \ i \\ -1 & \text{otherwise,} \end{cases}$$

where τ is a vector of thresholds of length n, shall be called a *conjunctive test.*[15] Finally, define a *disjunctive test* as a doctrine δ such that

$$\delta(d) = \begin{cases} 1 & \text{if } \exists \ i : d_i \geq \tau_i \\ -1 & \text{otherwise,} \end{cases}$$

where τ is a vector of thresholds of length n.

Then we can state the following:

Proposition 1. *If all $\delta_j = \delta^*$, and δ^* is a balancing test, then ρ_m is a consistent rule. If δ^* is a conjunctive or disjunctive test, ρ_m need not be a consistent rule.*

Call the situation in the first sentence of Proposition 1 a "shared balancing test." Then let $\boldsymbol{\delta} = \{\delta_j\}$ (and similarly for $\mathbf{f}$), and let $\mathcal{F}(\boldsymbol{\delta})$ be the set of combinations of monotonic fact finding functions for the judges such that ρ_m is not consistent given $\boldsymbol{\delta}$ and $\mathbf{f}$. Now we will deal with the more general case where judges may disagree on doctrine and state a more ominous result, which is a more general form of the second sentence in Proposition 1:

Proposition 2. *If $\boldsymbol{\delta}$ is not a shared balancing test, $\mathcal{F}(\boldsymbol{\delta})$ is nonempty.*

The implications of Proposition 2 explain a structural reason embedded in our common law system for inconsistent doctrine. Because the judges are engaging in multi-step reasoning to determine case outcomes, in general the court's opinions taken as a whole can be inconsistent in the sense that doctrine is not monotonic in the findings of legal facts. To understand why

[14] One may note the similarity between such a doctrine and what Landa and Lax (2009) call "base rules."
[15] Note that balancing, conjunctive, and disjunctive tests are all monotonic doctrines.

such monotonicity is crucial, consider a situation in which we have not observed the court's rulings in all of (the infinite number of) the potential cases, nor has the court completely revealed in its opinions ρ_m (which is of course the situation we find ourselves in at all times).[16] Then what we can say about the law, or "the prophecies of what the courts will do in fact" (Holmes 1897), becomes very limited. If δ_m is guaranteed to be monotonic in f_m, we could deduce outcomes in some regions of D_m, and we will have some information about the set of fact finding functions that could be f_m. However, if δ_m is *not* guaranteed to be monotonic in f_m, much less could be said about the outcomes we should expect in cases not observed.

Moreover, when the revealed outcomes show δ_m to be non-monotonic in f_m, the collegial doctrine is revealed to be "perverse" (Lax 2007, 594; Landa and Lax 2009, 952).[17] In other words, a person observing two different cases may believe in case one, the police had probable cause and conducted a seizure of a person amounting to an arrest, and in the second case, the police arrested a suspect with *more* evidence of criminality than in the first case, but find the court rules the police conduct constitutional in the first case but unconstitutional in the second. The ICR may even assign different outcomes to cases at the same location in the doctrine space. For example, a person may view two different set of historical facts and determine that in both cases police had probable cause and conducted a seizure of a person amounting to an arrest, and therefore acted in accordance with the Fourth Amendment, but observe the court rule the actions as constitutional in one case and unconstitutional in the other.

An example of inconsistent doctrine in the context of the Fourth Amendment is depicted

[16] See, for example, the discussion of the imperfect ability of judges to communicate their preferences in Clark (2016).

[17] Callander and Clark (2017) consider not expecting monotonicity of legal rules, and discuss expecting only a reliance on the dictate that like cases be treated alike (187). This makes sense in the context explored there, where the outcome is not binary, but a latent legal outcome in $\mathbb{R}$. Then similarity across a unidimensional case space that is non-monotonic is easy to understand. However, when we allow ourselves to describe doctrine in multiple dimensions, "wrinkles" or "cut outs" that Callander and Clark (2017) account for with non-monotonicity can often be accounted for by a monotonic rule in a richer historical fact and/or doctrine space. It's also harder to say what similarity in points "close" to each other in the case space is when we are dealing only with the dichotomous outcomes and not a latent legal outcome, absent monotonicity. Additionally, in at least some areas of the law, like the Fourth Amendment example discussed here, monotonicity *is* a normative expectation.

in Figure 3. Each of the judges has monotonic doctrines (disjunctive tests) and monotonic fact-finding functions, but doctrine is decidedly inconsistent; in this example, $H = [0,1]^4$, $D = [0,1]^2$, and the judges' fact-finding functions and doctrines are given in Table 2.[18] Although difficult to depict, in the darkly shaded region where both outcomes occur, the density of cases receiving each outcome varies, and importantly sometimes in an alternating fashion. We see both types of problems mentioned in the previous paragraph: opposing outcomes occurring at the same point in D, and violations of strict monotonicity as well.

Table 2: Doctrines and fact-finding functions for the judges on the collegial court.

j	f	δ
1	$(0.5h_1 + 0.5h_2, 0.5h_3 + 0.5h_4)$	$1 \Leftrightarrow d_1 > 0.750 \wedge d_2 > 0.750$
2	$(0.6h_1 + 0.4h_2, 0.4h_3 + 0.6h_4)$	$1 \Leftrightarrow d_1 > 0.375 \wedge d_2 > 0.750$
3	$(0.4h_1 + 0.6h_2, 0.6h_3 + 0.4h_4)$	$1 \Leftrightarrow d_1 > 0.750 \wedge d_2 > 0.375$

For example, consider the cases listed in Table 3. In case 1, both judges 2 and 3 find the case satisfies one element of their disjunctive test (though different ones), so both vote for outcome 1, while in case 2, both judges 1 and 3 find the case satisfies neither element of their disjunctive test, and so both vote for outcome -1. So, while d_m places both cases at $(0.375, 0375)$, they receive opposing outcomes! This is so because while the judges' individual preferences at both levels of aggregation are assumed to be very well behaved, the different levels of aggregation do not always agree with each other. Then, in case 3, judges 1 and 2 find the case satisfies neither element of their disjunctive test, and so both vote for outcome -1, resulting in a case at $d_m = (0.375, 0.4)$ having an outcome of -1 even though a case at $d_m = (0.375, 0.375)$ has an outcome of 1.

Of course, in this example, the structure of H is relatively simple, and the f_j appear easy enough to communicate. Even if lower court judges and members of the public have not had a chance to observe the full mapping from H to D_m to outcomes, assuming the judges had full knowledge of their preferences they could simply announce when deciding any case the

[18] This configuration of preferred doctrines is also used in Figure 1.

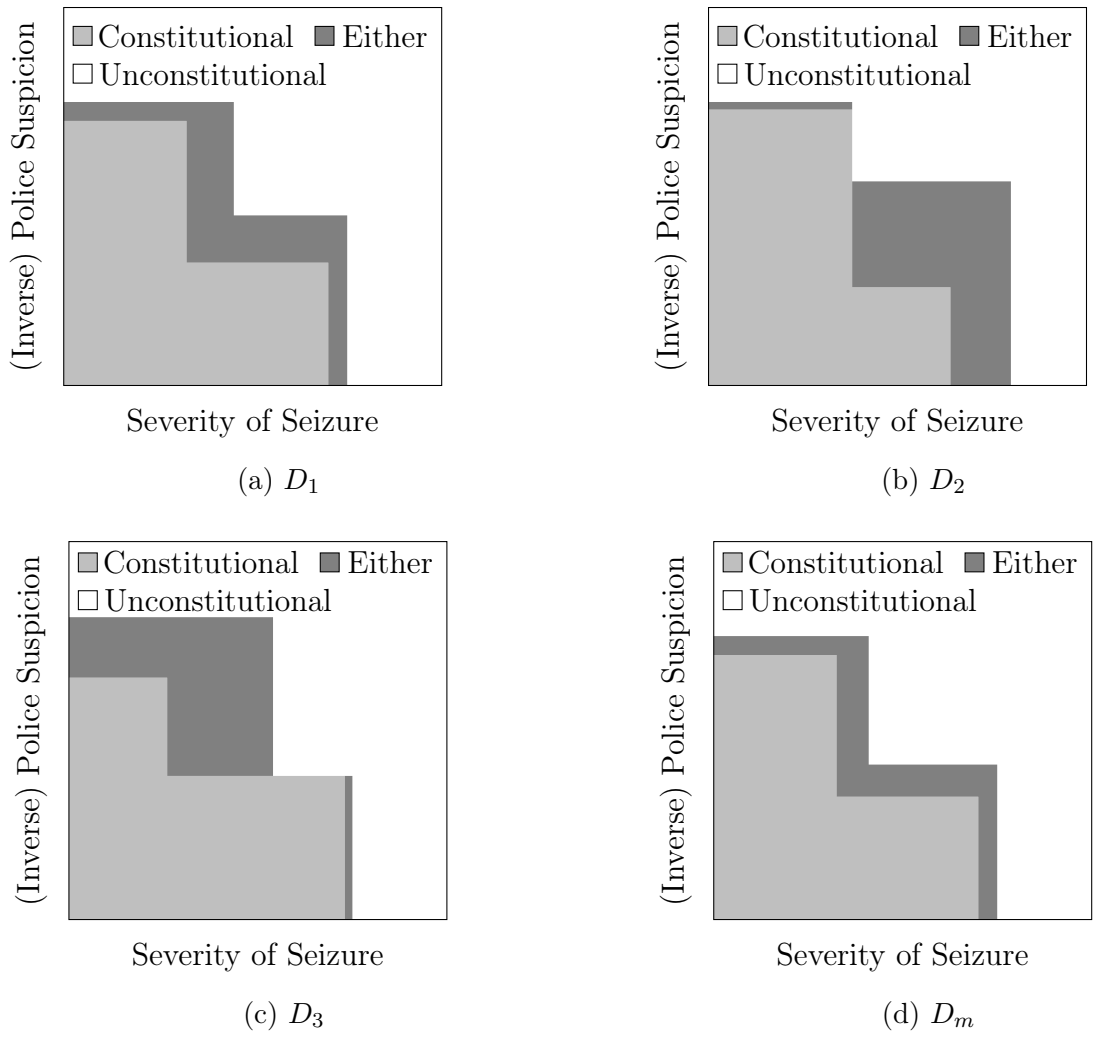


Figure 3: An example of an inconsistent doctrine. The doctrine space is comprised of two dimensions: severity of the police seizure, where larger values indicate a more intrusive seizure, and inverse police suspicion, where larger values indicate less certainty that criminal conduct has occurred. The judges all have preferred monotonic doctrines and monotonic fact-finding functions, but the collegial outcome set is not monotonic in any of the judges' projection of historical facts into doctrine space, or even the projection taking the dimension-by-dimension median placement of the majority coalition in every case.

association between H and the collegial outcome set. However, it is important to note the differences between an easily understood toy example like this and the even worse situation we generally find ourselves in. Generally H will be of a much higher dimensionality than 4; consider our Fourth Amendment example, where it is relevant whether the police restrained the suspect, the duration of the seizure, the credibility of information the police are acting on, what the supect was doing at the time of the seizure, etc. (many of which could be

Table 3: An example of an inconsistent collegial rule.

Case	h	j	d	Outcome
1	$(0.70, 0.05, 0.75, 0.00)$	1	$(0.375, 0.375)$	1
		2	$(0.440, 0.300)$	
		3	$(0.310, 0.450)$	
		m	$(0.375, 0.375)$	
2	$(0.15, 0.60, 0.15, 0.60)$	1	$(0.375, 0.375)$	-1
		2	$(0.330, 0.420)$	
		3	$(0.420, 0.330)$	
		m	$(0.375, 0.375)$	
3	$(0.15, 0.65, 0.40, 0.40)$	1	$(0.400, 0.400)$	-1
		2	$(0.350, 0.400)$	
		3	$(0.450, 0.400)$	
		m	$(0.375, 0.400)$	

further broken down into multiple historical fact dimensions, but were not for simplicity here). Moreover, *the judges are unlikely to know even their own full mapping* f_j. Judges are commonly presented with new historical factual dimensions that they have never considered before, and they often do not know how such facts affect where the judge will place the case in D until they have occasion to consider it.

In other words, judges often *cannot* create general doctrinal statements in terms of H; they must communicate their general decision principles in terms of D, and relate cases $h \in H$ to D as they come. In this setting, I have shown a very troublesome result; policies generated and communicated using such multi-step reasoning are generally subject to doctrinal inconsistency.

5 Discussion

Legal inconsistency is a problem, both for judges as policy makers, since agents and outside actors cannot follow or implement rules they do not understand, and for the public, who might normatively expect consistent application of legal rights. The prior literature offers explanations for inconsistency in individual judges' choices and their preferences (e.g. Collins 2008; Maltzman, Spriggs and Wahlbeck 2000) or for lack of precision in doctrine (e.g. Clark

2016; Fox and Vanberg 2014; Lax 2012; Staton and Vanberg 2008), but so far lacked a theory to explain *inconsistent* doctrine. By expanding on case-based models such as those from Kornhauser (1992) and Lax (2007) to account for judges' multi-step reasoning, I fill this gap. When we allow for disagreement over both how historical facts should be aggregated to doctrinal dimensions and how the doctrine space should be partitioned into outcomes, the resulting judgment and preference aggregation among judges displays inconsistency even under strict assumptions about how well-behaved the individual judgments and preferences are. The general presence of the danger of this inconsistency explains why so often, courts' doctrines become inconsistent (Drahozal 2004; Post 1995; Robinson and Simon 2006; Will 2019).

References

Badawi, Adam B. and Scott Baker. 2015. "Appellate Lawmaking in a Judicial Hierarchy." *The Journal of Law & Economics* 58(1):139–172.

Callander, Steven and Tom S. Clark. 2017. "Precedent and Doctrine in a Complicated World." *American Political Science Review* 111(1):184–203.

Clark, Tom S. 2016. "Scope and precedent: judicial rule-making under uncertainty." *Journal of Theoretical Politics* 28(3):353–384.

Clark, Tom S. and Clifford J. Carrubba. 2012. "A Theory of Opinion Writing in a Political Hierarchy." *Journal of Politics* 74(2):584–603.

Collins, Jr., Paul M. 2008. "The Consistency of Judicial Choice." *Journal of Politics* 70(3):861–873.

Dinh, Viet D. 2000. "Appellate Lawmaking in a Judicial Hierarchy." *Georgetown Law Journal* 88:2085–2118.

Drahozal, Christopher R. 2004. *The Supremacy Clause: A Reference Guide to the United States Constitution*. Westport, Connecticut: Praeger.

Fox, Justin and Georg Vanberg. 2014. "Narrow versus borad judicial decisions." *Journal of Theoretical Politics* 26(3):355–383.

Hoffman, Adam. 2001. "Corralling Constitutional Fact: De Novo Fact Review in the Federal Appellate Courts." *Duke Law Journal* 50:1427–1466.

Holmes, Jr., Oliver Wendell. 1897. "The Path of the Law." *Harvard Law Review* 10:457–478.

Kornhauser, Lewis A. 1992. "Modeling Collegial Courts II. Legal Doctrine." *Journal of Law, Economics, & Organization* 8(3):441–470.

Landa, Dimitri and Jeffrey R. Lax. 2008. "Disagreements on Collegial Courts: A Case-Space Approach." *Journal of Constitutional Law* 10(2):305–329.

Landa, Dimitri and Jeffrey R. Lax. 2009. "Legal Doctrine on Collegial Courts." *Journal of Politics* 71(3):946–963.

Lax, Jeffrey R. 2007. "Constructing Legal Rules on Appellate Courts." *American Political Science Review* 101(3):591–604.

Lax, Jeffrey R. 2012. "Political Constraints on Legal Doctrine: How Hierarchy Shapes the Law." *Journal of Politicse* 74(3):765–781.

List, Christian. 2012. "The Theory of Judgment Aggregation: An Introductory Review." *Synthese* 187(1):179–207.

List, Christian and Phillip Pettit. 2002. "Aggregating Sets of Judgments: An Impossibility Result." *Economics and Philosophy* 18:89–110.

Maltzman, Forrest, James F. Spriggs, II and Paul J. Wahlbeck. 2000. *Crafting Law on the Supreme Court: The Collegial Game*. Cambridge University Press.

Nehring, Klaus and Clemens Puppe. 2006. "Consistent Judgement Aggregation: The Truth-Functional Case." *Social Choice and Welfare* 31(1):41–57.

Nehring, Klaus and Clemens Puppe. 2010. "Justifiable Group Choice." *Journal of Economic Theory* 145(2):583–602.

Owens, Ryan J. and Justin P. Wedeking. 2011. "Justices and Legal Clarity: Analyzing the Complexity of U.S. Supreme Court Opinions." *Law & Society Review* 45(4):1027–1061.

Post, Robert. 1995. "Recuperating First Amendment Doctrine." *Stanford Law Review* 47(6):1249–1281.

Redish, Martin H. and William D. Gohl. 2017. "The Wandering Doctrine of Constitutional Fact." *Arizona Law Review* 59:289–338.

Robinson, Matthew B. and Kathleen M. Simon. 2006. "Logical and Consistent? An Analysis of Supreme Court Opinions Regarding the Death Penalty." *Justice Policy Journal* 3(1):1–59.

Segal, Jeffrey A. 1984. "Predicting Supreme Court Cases Probabilistically: The Search and Seizure Cases, 1962-1981." *American Political Science Review* 78(4):891–900.

Segal, Jeffrey A. and Harold J. Spaeth. 2002. *The Supreme Court and the Attitudinal Model Revisited*. Cambridge University Press.

Spriggs, II, James F. 1996. "The Supreme Court and Federal Administrative Agencies: A Resource-Based Theory and Analysis of Judicial Impact." *American Journal of Political Science* 40(4):1122–1151.

Staton, Jeffrey K. and Georg Vanberg. 2008. "The Value of Vagueness: Delegation, Defiance, and Judicial Opinions." *American Journal of Political Science* 52(3):504–519.

Westerland, Chad, Jeffrey A. Segal, Lee Epstein, Charles M. Cameron and Scott Comparato. 2010. "Strategic Defiance and Compliance in the U.S. Courts of Appeals." *American Journal of Political Science* 54(4):891–905.

Will, George F. 2019. "The Supreme Court can undo past confusion with its ruling on this WWI memorial.".
URL: *https://www.sltrib.com/opinion/commentary/2019/02/24/george-f-will-supreme/*

A Formal Results and Proofs

Proposition 1. *If all $\delta_j = \delta^*$, and δ^* is a balancing test, then ρ_m is a consistent rule. If δ^* is a conjunctive or disjunctive test, ρ_m need not be a consistent rule.*

Proof.

Balancing tests:

Suppose all $\delta_j = \delta^*$, a balancing test. Then when the collegial outcome is 1, d_j must lie on or beyond the hyperplane in D described by δ^* for every j in the majority coalition for any case h. Then the point in D that is a dimension-by-dimension median for the majority coalition at h must also lie on or beyond the hyperplane described by δ^*. Similarly, when the collegial outcome is -1, d_j must not lie as far as the hyperplane described by δ^* for any j in the majority coalition for any case h, and therefore the point in D that is a dimension-by-dimension median for the majority coalition at h must also not lie as far as the hyperplane described by δ^*. Therefore $\delta_m = \delta^*$ and is monotonic in f_m.

Conjunctive and disjunctive tests:

For conjunctive tests, when the collegial outcome is 1, $d_{ij} \geq \tau_i \ \forall \ i$ for every j in the majority coalition for any case h, and analogously for disjunctive tests and collegial outcomes

of -1. Then the point in D that is a dimension-by-dimension median for the majority coalition at h must also satisfy that condition. However, when the collegial outcome is -1 for a conjunctive test, we can only say $\exists\ i : d_{ij} \leq \tau_i$ rather than $d_{ij} \leq \tau_i\ \forall\ i$. Then for each dimension i, there may be a member of the majority coalition with a high enough placement of d_{ij} such that $d_{im} \geq \tau_i\ \forall\ i$ even though no member of the majority coalition would assign the outcome 1. (Again, an analogous argument applies for disjunctive tests). Therefore δ_m may not be monotonic in f_m. □

Lemma 1. *If δ_j is monotonic in f_j for every $j \in J$, then for every two cases $h, h' \in H$, $h_k \geq h'_k\ \forall\ k$ implies that the collegial outcome at h is weakly greater than the collegial outcome at h'.*

Proof. If f_j is monotonic and δ_j is monotonic in f_j, then $\delta_j(f_j)$ is monotonic in H. Then the collegial outcome set is monotonic in H by Proposition 3 from Lax (2007). □

Proposition 2. *If $\boldsymbol{\delta}$ is not a shared balancing test, $\mathcal{F}(\boldsymbol{\delta})$ is nonempty.*

Proof. If $\boldsymbol{\delta}$ is not a shared balancing test, but every δ_j is monotonic, then by Lemma 1, the collegial outcome set is monotonic in H. Then define two parallel hyperplanes in H, partitioning it into three regions, such that cases in the least extreme region contain only cases with -1 outcomes, cases in the most extreme region contain only cases with 1 outcomes, and cases in the "middle" region contain cases with both outcomes (this third region is guaranteed to exist since $\boldsymbol{\delta}$ is not a shared balancing test). Then since $N > n$, a monotonic fact finding function $\hat{f}$ can then be constructed such that corresponding hyperplanes exist in D, where δ_m is monotonic in the most and least extreme regions in $\hat{D}$, but non-monotonicity is induced in the middle region. □

Remark 1. The contrast between Lemma 1 and Proposition 2 may explain why legal scholars so roundly critiqued Fourth Amendment doctrine although Segal (1984) found decisions well explained by historical facts. Segal (1984) was right that if you have enough observations (possibly a lot depending on the structure of H), a clear relationship can be found between

H and outcomes. However, because humans cannot generally think in such high dimensional spaces, a clear relationship between D and outcomes is what's needed for clear doctrine; unfortunately Proposition 2 shows that even in general settings with well-behaved individual preferences, such a relationship may not obtain.

B Deference to Trial Court Findings

I refer to two types of facts, *historical facts* and *doctrinal facts*, the latter of which are sometimes what legal texts may refer to as mixed questions of law and fact. A historical fact is what a lay person may typically think of as a fact: whether or not an accused murderer's victim is deceased, whether or not a traffic light was green, etc. A doctrinal fact is a fact that requires some level of legal analysis to determine: whether or not there was probable cause for a search, whether or not a contract was formed, etc.

The model assumes the collegial appellate court determines for itself all the doctrinal facts. When the doctrinal facts are questions of law, this is appropriate. When the doctrinal facts are mixed questions, sometimes appellate courts give greater deference to trial courts' findings, and sometimes they review them *de novo*. What kinds of intermediate factual determinations do they subject to greater scrutiny? Extant legal reasoning suggests the answer to this question perhaps should be based on which "judicial actor is better positioned . . . to decide the issue in question." *Miller v. Fenton*, 474 U.S. 104 (1985), at 114. For example, appellate and trial judges are equally capable of examining the text of an unambiguous contract, while those present at trial are in a better position to determine the credibility of witness testimony, having been present to observe their demeanor. In *Ornelas v. United States*, 517 U.S. 690 (1996), quoted in the main text, the U.S. Supreme Court settled a circuit split on whether findings of probable cause should be reviewed *de novo* or with deference (in favor of *de novo* review).

In addition to standard of review choices regarding mixed questions, appellate courts

sometimes apply a "constitutional fact doctrine" that could allow for independent review even of historical facts (see, e.g., *Ohio Valley Water Co. v. Borough of Ben Avon*, 253 U.S. 287 (1920), allowing *de novo* review of property valuation on appeal from state agency determinations). Some legal scholars have criticized the evolution of this doctrine, worried it will be applied in inappropriate situations (see, e.g., Hoffman 2001; Redish and Gohl 2017). The results here provide a formal theoretic result to support such a concern.

Perhaps most troubling, heightened standards of review tend to be applied in cases regarding civil liberties such as First and Fourth Amendment rights (Hoffman 2001; Redish and Gohl 2017). For example, consider the determination in *Ornelas* to subject probable cause determinations to *de novo* review. Similarly, some circuits hold that whether speech constitutes a "true threat" unprotected by the First Amendment is a fact reviewable *de novo*, though other appellate courts disagree (Redish and Gohl 2017, 292). Heightened review of facts is applied in many First Amendment contexts, from cases involving freedom of speech to those implicating religious freedoms (Hoffman 2001, 1453–1455). The legal rules governing arguably our most important freedoms, for which there is often disagreement such that the first sentence of Proposition 1 may not apply, may be the areas in which courts most often apply increase scrutiny on findings of doctrinal facts.