A Bayesian View of the Result Model

Yoann Morello and Agata Ciabattoni

Vienna University of Technology (TU Wien)
yoann.morello@gmail.com, agata@logic.at

Abstract. Real-world datasets often contain inconsistencies that challenge traditional case-based reasoning models. Building upon the result model, a well-established formal representation of case base reasoning in law, we propose a Bayesian reinterpretation that effectively addresses such inconsistencies. Our Bayesian enhancement quantifies the reliability of precedents and encapsulates principled, explainable predictions even in the presence of conflict, representing a meaningful step forward in using these models to design AI agents.

Keywords: Case-based reasoning · Bayesian inference · Legal AI · Inconsistency · Explainability· Normative AI Agents

1 Introduction

A central challenge in machine ethics, the area of Artificial Intelligence (AI) focused on building systems capable of normative reasoning, lies in acquiring and representing normative information in a form that is both implementable and transparent. Symbolic methods, such as rule-based systems, provide transparency, but lack flexibility. Conversely, approaches based on black-box machine learning lack reliability and explainability. A hybrid approach inspired by common law reasoning has been introduced in [6] and is based on the *reason model* [11]; the approach aims to bridge this divide: norms are learned from particular cases (bottom-up) but represented symbolically (top-down). Indeed, in common law, norms arise not from statute but from precedents—that is, actions or decisions that serve as justification or support for future actions or decisions. As decisions accumulate, a body of normative expectations emerges, and constrain future decisions. The *reason model* formalizes this idea by requiring new decisions to be consistent with precedents and by integrating additional information provided by the judge in the form of explicit reasons justifying each decision. However, in many contexts, such as the DIAS (Drug-Interdiction Auto-Stop) dataset [9], such reasons are often unavailable. In these cases, the *result model*, a minimal symbolic framework introduced in [1] (see also [10]), enables normative inference directly from case outcomes and fact patterns. To connect these frameworks for formalizing common law to machine ethics, consider a hypothetical robot whose decision-making relies on questionnaires administered to users, each capturing a specific situation. The responses to these questionnaires serve as precedents, and the robot's future decisions in new situations should be constrained by them. Last but not least, the robot must provide explanations for its decisions.

Yet, when applying such frameworks to systems like our hypothetical robot, a key limitation becomes evident: most approaches rely on the assumption that the set of precedents is consistent. In other words, they assume that the dataset does not impose contradictory decisions for the same or stronger fact situations. However, this assumption is unrealistic.

Inconsistencies can indeed arise for several reasons. First, different users may hold divergent normative intuitions and provide conflicting judgments. Second, borderline cases may genuinely be ambiguous, leading even a single user to choose different sides in similar contexts. Third, the factors used to describe cases may fail to capture critical distinctions, resulting in substantively different situations being treated as equivalent in the dataset. These inconsistencies are not rare exceptions. They are, in fact, common features of real-world legal and ethical corpora, as illustrated by the DIAS dataset [8, 9].

Two distinct approaches to accommodating inconsistencies have recently been proposed in [5] and [15, 14]. The first, developed in the context of the reason model, provides a logically sound framework for reasoning under conflicting precedents. However, it does not leverage the statistical information that the case database reveals about the reliability of the constraints we aim to derive from precedent decisions. Our first contribution is to adapt this framework to the result model, yielding two predictive variants that we refer to as the *Strict Binary* and *Binary Majority* models. The second approach [15, 14], designed for models based on dimensions and a different notion of precedent, introduces the concept of *authoritativeness* to resolve conflicts. Our main contribution, the *Bayesian model*, extends this idea by learning a distribution over authoritativeness directly from data, thereby capturing both the strength of a precedent and the uncertainty surrounding its application.

Unlike previous approaches, the Bayesian model reinterprets the result model *probabilistically*, allowing it to handle inconsistencies in real case datasets while providing epistemic confidence in each prediction. It thus represents a promising *hybrid* between symbolic legal reasoning and data-driven inference: it retains the transparency and traceability of rule-based methods while making principled use of data to assess the reliability of conclusions.

The three predictive models presented in this paper: the *Strict Binary*, *Binary Majority*, and *Bayesian* models lay the groundwork for subsequent experimental evaluation.

2 Preliminaries

We provide essential contextual information for understanding the paper.

2.1 The Result Model in a Nutshell

The result model [1] is a highly conservative approach to precedential reasoning: it formalizes which types of inferences from prior cases are admissible — namely, that only weaker precedents can constrain new decisions. The result model is governed by the principle that:

> "To follow precedent, a constrained court must decide its case for the party analogous to the winner in the precedent case if the constrained case is as strong or stronger a case for that result than the precedent case was for its result".

To formalize this idea: according to [16] a factor is a consideration a decision maker must or may take into account to determine an outcome. Let F be the set of factors, partitioned into F^π (favoring the plaintiff) and F^δ (favoring the defendant), so that $F = F^\pi \cup F^\delta$.

Definition 1. *A case is a tuple $\langle X, s \rangle$, where $X \subseteq F$ is the set of present factors*[1] *and $s \in \{\pi, \delta\}$ is the side that won. We define $X^s = X \cap F^s$ as the factors in X that favor side s. We write $\overline{s}$ to denote the side opposite to s, so that $\overline{\pi} = \delta$ and $\overline{\delta} = \pi$.*

Definition 2 (Strength Ordering). *Given two fact situations X and Y, we say that X is* stronger for side s *than Y, written $X \succcurlyeq_s Y$, if*

$$X^s \supseteq Y^s \quad and \quad X^{\overline{s}} \subseteq Y^{\overline{s}}.$$

Example 1. Let $\{f_1^\pi, f_2^\pi\}$ the set F^π favoring the plaintiff and $F^\delta = \{f_1^\delta\}$, the set favoring the defendant. Consider two cases with the following sets of factors:

- $X = \{f_1^\pi, f_1^\delta\}$
- $Y = \{f_1^\pi, f_2^\pi\}$

The relevant sets are:

$$X^\delta = \{f_1^\delta\}, \quad Y^\delta = \emptyset \quad \Rightarrow \quad X^\delta \supseteq Y^\delta$$

$$X^\pi = \{f_1^\pi\}, \quad Y^\pi = \{f_1^\pi, f_2^\pi\} \quad \Rightarrow \quad X^\pi \subseteq Y^\pi$$

Therefore, $X \succcurlyeq_\delta Y$, that is case X is stronger for the defendant than Y, because it contains more factors supporting the defendant and fewer factors supporting the plaintiff.

Definition 3 (Case base). *A* case base *is a finite set of cases.*

Definition 4 (Consistency). *A case base is consistent if there do not exist $c = \langle X, \delta \rangle$ and $c' = \langle Y, \pi \rangle$ such that $X \succcurlyeq_\pi Y$.*

This condition can be equivalently stated in terms of the *a fortiori* constraint, which governs the admissibility of new decisions based on prior ones.

[1] Following Horty's factor-based account of precedent [11], we treat a *factor* as a legally salient pattern of facts that invariably favors one litigant. For illustration, the DIAS dataset contains pro-government factors such as Legal Indications of Drug Use and Vehicle Contents Suggest Drugs.

Definition 5 (A Fortiori Constraint). *Let $c = \langle X, s \rangle$ be a new case and let $c' = \langle Y, s \rangle$ be a past case in the case base. The* a fortiori *constraint requires that if $X \succcurlyeq_s Y$, then the decision in c must also be in favor of s.*

That is, no weaker case should be decided for a side if a stronger one was not.

Under this principle, consistency means that the case base contains no pair of cases violating the a fortiori constraint.

Example 2 (Violation of the a fortiori *constraint).* Consider the two cases:

$$c_1 = \langle X_1, \pi \rangle \quad \text{and} \quad c_2 = \langle X_2, \delta \rangle,$$

where the factor sets are:

$$X_1 = \{f_1^\pi, f_1^\delta\} \quad \text{and} \quad X_2 = \{f_1^\pi, f_2^\pi\}.$$

We observe that:

$$X_1^\pi = \{f_1^\pi\} \subseteq X_2^\pi = \{f_1^\pi, f_2^\pi\}, \quad X_1^\delta = \{f_1^\delta\} \supseteq X_2^\delta = \emptyset.$$

Hence, by definition, $X_2 \succcurlyeq_\pi X_1$, meaning c_2 is stronger than c_1 for the plaintiff. However, c_1 is decided for the plaintiff, while c_2 is decided for the defendant. This contradicts the *a fortiori* principle: if a weaker case is decided for the plaintiff, a stronger one should be as well.

We can identify several potential sources of such inconsistencies in the data, including:

1. **Suboptimal factor design.** In some cases, the factors used to describe legal situations may conflate distinct evidentiary strengths. For instance, in the DIAS dataset [8, 9], one can find cases[2] judged for the plaintiff with only the factor f_1^π =*"Vehicle contents suggest drugs"*, while other cases containing this same factor along with several additional plaintiff-favoring factors are judged for the defendant. Upon examining the textual descriptions, it becomes evident that in the cases judged for the plaintiff, f_1^π refers to strong indicators (e.g., a strong smell of marijuana, or visible white powder on the driver), whereas in cases judged for the defendant, it refers to weaker signs (e.g., faint odor, presence of syringes). Splitting this factor into two levels of evidentiary strength, or modeling it using a dimensional approach, would likely eliminate the observed inconsistency.
2. **Value pluralism across judges.** Different judges may apply the same legal principles in different ways due to divergent normative priorities.
3. **Indeterminacy in borderline cases.** When a case is near the boundary of legal ambiguity, even the same judge may decide in different directions on different occasions, essentially choosing between two well-balanced alternatives.

[2] Most of these inconsistencies might be attributable to divergent practices across different jurisdictions in categorizing cases.

Remark 1. Introduced by Horty [11], the *reason model* refines the result model by associating each decision with an explicit *reason*, that is, a subset of factors identified by the judge as justifying the outcome. The same principle of a fortiori reasoning can be recovered if we replace each real case by a counterfactual one, where the full set of winning-side factors is replaced by the stated reason. Under this transformation, analogical constraints can be applied as in the result model. The reason model improves on the result model by resolving certain inconsistencies and enabling a broader range of predictions. However, it is only applicable when judges explicitly provide reasons for their decisions.

2.2 Bayesian Framework

We briefly recall the core concepts of Bayesian statistics necessary to evaluate how reliably the *a fortiori* principle is supported by empirical data. Bayesian statistics models uncertainty about unknown quantities—such as the reliability of a legal constraint—by treating them as random variables described by probability distributions. As new data is observed, beliefs about these quantities are updated using *Bayes' theorem.* For a more comprehensive introduction, see [7]. Bayes' theorem relates our prior beliefs about a parameter θ (the *prior distribution*) to our updated beliefs after observing data (the *posterior distribution*):

$$p(\theta, \text{data}) = p(\theta \mid \text{data})\, p(\text{data}) = p(\text{data} \mid \theta)\, p(\theta).$$

Rewriting it in proportional form:

$$p(\theta \mid \text{data}) \propto p(\text{data} \mid \theta)\, p(\theta)$$

where

- $p(\theta)$ is the *prior distribution*, encoding our belief about θ before observing any data;
- $p(\text{data} \mid \theta)$ is the *likelihood*, measuring how likely the observed data is under parameter value θ, and
- $p(\theta \mid \text{data})$ is the *posterior distribution*, our updated belief incorporating the data.

The proportionality symbol ($\propto$) indicates that $p(\text{data})$ acts as a normalization constant independent of θ.

In the context of our work, we consider how often the *a fortiori* constraint holds between pairs of cases. Each such observation is a binary outcome: either the constraint is satisfied or violated. This suggests modeling constraint application as a Bernoulli trial.

Definition 6 (Bernoulli Distribution). *The Bernoulli distribution models binary outcomes and is parameterized by a probability $\theta \in [0, 1]$, which in our setting corresponds to the unknown reliability of the* a fortiori *constraint. For a given observation $x \in \{0, 1\}$, the probability mass function is:*

$$\text{Bern}(x; \theta) = \begin{cases} \theta & \textit{if } x = 1, \\ 1 - \theta & \textit{if } x = 0. \end{cases}$$

In our context, $x = 1$ indicates that the a fortiori *constraint is satisfied between a pair of cases, while $x = 0$ indicates a violation.*

Definition 7 (Uniform Distribution). *The uniform distribution on the interval $[0, 1]$ is defined by the constant probability density:*

$$\text{Uniform}(\theta) = 1 \quad \textit{for } \theta \in [0, 1].$$

It represents a non-informative prior*, expressing total ignorance: all values of θ are considered equally plausible before observing any data.*

Remark. The uniform distribution over $[0, 1]$ is a special case of the Beta distribution, specifically $\text{Beta}(1, 1)$. This provides a smooth transition into using the Beta family as a prior in our Bayesian analysis.

Bayesian inference yields not a single point estimate of θ but a full posterior distribution, reflecting the degree of certainty or uncertainty in our beliefs after observing data.

To perform Bayesian inference for a Bernoulli parameter θ, we choose a prior distribution over $[0, 1]$. An expressive family of priors is the Beta distribution:

Definition 8 (Beta Distribution). *The Beta distribution with parameters $\alpha > 0$ and $\beta > 0$ has density:*

$$\text{Beta}(\theta; \alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)},$$

where the Beta function $B(\alpha, \beta)$ is defined as:

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt.$$

Conjugacy. The Beta distribution is the *conjugate prior* to the Bernoulli distribution. This means that if we use a Beta prior and observe data from a Bernoulli process, the posterior distribution is also a Beta distribution, but with updated parameters. Conjugate priors are useful as they allow posterior distributions to be computed analytically, avoiding the need for numerical approximation.

More concretely, if $\theta \sim \text{Beta}(\alpha, \beta)$ and we observe n binary outcomes with k successes ($x_i = 1$) and $n - k$ failures ($x_i = 0$), the posterior distribution is:

$$\theta \mid \text{data} \sim \text{Beta}(\alpha + k, \beta + n - k).$$

Interpretation. The parameters α and β of the Beta distribution are often called *pseudo-counts* because they act like prior observations, even though no real data has been seen yet.

– For instance, a prior $\text{Beta}(1, 1)$ is uniform over $[0, 1]$, but it behaves as if we had seen one success and one failure before collecting any real data.

– When we later observe r actual successes and s actual failures, the posterior becomes $\text{Beta}(\alpha + r, \beta + s)$. So the prior contributes *as if* it had already added α successes and β failures to the total.
– This allows Bayesian inference to start with a prior belief and then update it smoothly as data is observed.

This property makes the Beta distribution especially suitable for our setting, where we wish to quantify how often the *a fortiori* constraint is upheld versus violated. A precedent that is frequently followed with few violations yields a Beta distribution sharply concentrated near 1. In contrast, if its constraint is often violated or rarely tested, the resulting posterior remains broad, reflecting greater uncertainty. Thus, each precedent can be associated with a specific Beta distribution that captures our empirical confidence in its normative force.

3 Predictive Models

Several works have attempted to model prediction using precedents; a comprehensive survey is given in [3]. Most of these approaches rely on analogical similarity measures between cases rather than enforcing normative reasoning rules such as those encoded in the result model. An application of a fortiori reasoning can be found in [19], though it applies to numerical (or many-valued) rather than binary data. It is worth noting that a fortiori reasoning is just another name for the result model, while its adaptation to continuous or multi-level attributes is typically called *the result model with dimensions*.

By contrast, our focus is on the binary setting, in which attributes/factors are either present or absent. We explicitly target inconsistency, and we describe below three predictive models. The first two models adapt established literature rules (strict abstention or majority tie-breaking), while our Bayesian model adds a principled, reliability-based treatment to make predictions under conflict; the latter employs Bayesian statistics to quantify the degree of confidence the data warrants in each prediction made via the result model.

The section is organized as follows: Sec. 3.1 recalls a strict binary model that only allows prediction in the absence of conflicting precedents; Sec. 3.2 presents the model minimizing inconsistency proposed by [5], which we reformulate as a voting mechanism inspired by social choice theory; Sec. 3.3 introduces our Bayesian model, which quantifies uncertainty and weights precedents by reliability.

3.1 Strict Binary Result Model.

This model predicts an outcome only if all applicable precedents agree. If any conflict is found among applicable constraints, the model abstains. This reflects the definition of constraints under inconsistencies defined by [5] for the reason model. To make this precise, we adapt below Canavotto's definitions (from the reason model) to the setting of the result model.

Definition 9 (Supporting and Conflicting Precedents). *Let $c = \langle X, s \rangle$ be a new case. Define*

$$\text{supporting}_s(c) \;=\; \{\, c' \in \mathcal{C} \mid \text{outcome}(c') = s \;\wedge\; c \succcurlyeq_s c' \},$$

$$\text{conflicting}_s(c) \;=\; \{\, c' \in \mathcal{C} \mid \text{outcome}(c') = \overline{s} \;\wedge\; c \succcurlyeq_{\overline{s}} c' \}.$$

Proposition 1 (Strict Decision Criterion). *Under the result-model adaptation, the strict binary result model decides c as follows:*

$$\textit{If } \text{supporting}_s(c) = \emptyset \textit{ and } \text{supporting}_{\overline{s}}(c) \neq \emptyset, \quad \textit{then decide } c \textit{ for } \overline{s}.$$

Otherwise, the model abstains.

Proof (Sketch). By definition, $\text{supporting}_{\overline{s}}(c) = \text{conflicting}_s(c)$. Thus

$$\text{supporting}_s(c) = \emptyset \quad \wedge \quad \text{conflicting}_s(c) \neq \emptyset$$

means there is no precedent enforcing s but at least one enforcing $\overline{s}$, forcing the decision $\overline{s}$, with abstention in all other cases. □

This principle of enforcing a decision only when precedents are one-sided was connected in [4] to standard deontic logic SDL, whose deontic operators of permission and obligation are interpreted, respectively, as existential and universal quantifiers over a set of possible worlds [20]. The name *Strict Binary* reflects the model's rigid enforcement criteria: it permits a binary prediction only in the absence of conflicting precedents, thereby adhering strictly to the normative force of the *a fortiori* constraint. The model abstains as soon as ambiguity arises, which contrasts with more permissive approaches that tolerate conflict by aggregating information from competing precedents.

Example 3. Consider the following three cases with factors drawn from $F^\pi = \{f_1^\pi\}$, and $F^\delta = \{f_1^\delta\}$:

- $c_1 = \langle \{f_1^\pi, f_1^\delta, f_2^\delta\}, \pi \rangle$ (a precedent for the plaintiff),
- $c_2 = \langle \{f_1^\pi, f_2^\pi, f_1^\delta\}, \delta \rangle$ (a precedent for the defendant),
- $c_3 = \langle \{f_1^\pi, f_1^\delta\}, ? \rangle$ (a new case to be decided).

We observe:

$$\{f_1^\pi, f_1^\delta\} \succcurlyeq_\pi \{f_1^\pi, f_1^\delta, f_2^\delta\}, \quad \text{and} \quad \{f_1^\pi, f_1^\delta\} \succcurlyeq_\delta \{f_1^\pi, f_2^\pi, f_1^\delta\}.$$

Thus, c_1 and c_2 both act as applicable precedents for c_3, but they enforce opposite outcomes. Since the strict binary model abstains whenever there are conflicting applicable precedents, it will refrain from making a prediction for c_3.

3.2 The Binary Majority Vote Model.

The Binary Strict model makes predictions only when there are precedents supporting a decision and none supporting the opposite. It abstains both when no applicable precedent exists and when precedents for both outcomes are found. As the number of inconsistencies in the case database increases, the number of such conflicts typically grows, leading the model to abstain on more and more test cases.

To address this issue,[5] proposed a refined framework (for the reason model): when faced with conflicting precedents, select the decision that conflicts with the smallest number of precedents in the database.

Inspired by social choice theory [12], we reformulate the minimal number of conflicts principle as a voting rule:

Proposition 2. *Minimizing the number of conflicts is equivalent to assigning each π-precedent of the new case a vote of $+1$, and each δ-precedent a vote of -1. The case is predicted as π if the total sum is strictly positive, as δ if strictly negative, and left unclassified in case of a tie.*

Proof. Choosing a decision s for a new case introduces #(precedents for $\overline{s}$) conflicts, as each such precedent supports the opposite side. Comparing the number of conflicts introduced by choosing s versus $\overline{s}$ amounts to checking the sign of:

$$\#\text{precedents for } \pi - \#\text{precedents for } \delta.$$

□

Example 4 (Majority Vote Resolution). The case c_3 from the example 3 would also not be predicted by the *Binary Majority* model, as there is an equal number of applicable precedents supporting each side: one for π (namely c_1) and one for δ (namely c_2). Now consider adding a new case:

$$c_4 = \langle \{f_1^\pi, f_1^\delta, f_2^\delta\}, \pi \rangle,$$

which is identical to c_1. Then both c_1 and c_4 support the plaintiff and are applicable to c_3, while only c_2 supports the defendant.

In this case, the *Binary Majority* model identifies a majority of applicable precedents supporting π, and therefore predicts c_3 for the plaintiff.

While this model improves coverage, it still assumes all precedents have equal authority in enforcing the *a fortiori* constraint. Yet not all precedents are equally reliable. Some cases may frequently violate their expected influence over stronger or weaker cases, suggesting they should be trusted less.

This observation invites us to revisit the foundational motivation for the constraint itself. The *a fortiori* principle expresses a form of normative reasoning from precedent: if a weaker case is decided for the plaintiff, then a stronger one should be as well—and symmetrically for the defendant. Inconsistencies in the dataset can then be interpreted as uncertainty about the applicability of this principle in particular contexts.

3.3 The Bayesian Result Model

The considerations at the end of the previous section leads to a more refined question: can we use the data to quantify our confidence in the *a fortiori* constraint? A natural starting point is to measure the proportion of case pairs that satisfy the constraint versus those that violate it. Intuitively, the more often the constraint is upheld, the stronger our belief in its normative force. This proportion-based idea was explicitly adopted by [14], who defines a precedent's *authoritativeness* as the fraction of times the expected constraint that one associates to it is respected in the dataset.

While our approach, like [14], aims to quantify the normative strength of precedents, it differs in two essential respects. First, we identify applicable precedents directly through the *a fortiori* constraint defined by the result model—that is, by comparing fact patterns via subset inclusion—whereas [14, 15] rely on a distance-based similarity metric over dimensional case representations. Their approach is more closely aligned with the HYPO framework [2], a pioneering and influential model of legal case-based reasoning that uses analogical arguments built from precedents estimated to be similar to the case at hand.

Second, whereas they assign each precedent a scalar *authoritativeness* score (which, under certain assumptions, corresponds to the mean of our Beta posterior), our Bayesian framework yields a full probability distribution. This richer representation encapsulates not only the expected strength of a precedent but also the uncertainty around that estimate, enabling principled reasoning under epistemic uncertainty in downstream decisions.

These considerations motivate the probabilistic framework we introduce below. We begin by observing that the *a fortiori* principle is not a single global rule, but rather a scheme of local constraints—one associated with each case c in the training set. Each c acts as a potential precedent and defines its own individual constraint:

Definition 10 (Individual *a fortiori* Constraint for a Case c). *Let c be a case in the training set.*

- *If $decision(c) = \pi$, then c enforces that any case c' with $c' \succcurlyeq c$ must also be decided for π.*
- *If $decision(c) = \delta$, then c enforces that any case c' with $c \succcurlyeq c'$ must also be decided for δ.*

This formulation provides the basis for our Bayesian treatment of precedent reliability. Rather than treating all precedents equally, we learn a distributional confidence for each case's individual constraint—allowing us to model varying degrees of normative strength among precedents.

Inconsistencies in the dataset imply that the *a fortiori* constraint for a case is only stochastically applied. Since the application of such a constraint is a binary (0/1) observation, it is natural to model this uncertainty using a Bernoulli distribution with an unknown success parameter θ_c. To estimate this parameter, we place a Beta prior over each θ_c, resulting in a posterior distribution:

$$\theta_c \sim \text{Beta}(r_c + 1, s_c + 1),$$

where r_c is the number of successful applications of the constraint in the training data (i.e., pairs where the constraint applies and is satisfied), and s_c is the number of observed violations. We use a uniform prior Beta$(1, 1)$ to model lack of prior knowledge, as discussed in Section 2.2. From the definition of the Beta distribution, this yields the posterior in closed form due to conjugacy. While we suggest here the uniform prior Beta$(1, 1)$, we note that alternative prior choices are possible and will return to this discussion in Section 3.3.

This use of Bayesian statistics differs significantly from prior work in rule learning. For instance, [13] and [18] apply Bayesian methods to discover symbolic rules or logical clauses from labeled data. Their goal is to induce rule structure. By contrast, we do not attempt to learn the constraint itself—it is fixed and justified on normative grounds. Instead, we use Bayesian inference to assess how confidently each instance of the constraint (i.e., each precedent) should be applied, based on empirical evidence from the case base.

This framework also aligns with the theoretical ideal of normative consistency: as the empirical evidence supporting a constraint grows, so too does our confidence in its validity. In particular, a high number of supporting instances r_c relative to violations s_c, combined with a large overall sample size, yields a posterior increasingly concentrated near 1. In the limiting case of a perfectly consistent and infinite dataset—where the constraint has never been violated—we recover full certainty: With $\theta_c \sim \text{Beta}(r_c + 1, 1)$, we have $\theta_c \to 1$ in probability as $r_c \to \infty$; thus the limiting law is the degenerate distribution at 1 (i.e., all probability mass is concentrated at a single point), representing an idealized belief that the variable equals 1 with certainty.

At prediction time, each precedent c casts a probabilistic vote for the side it supports via the *a fortiori* relation. This vote is drawn from a Bernoulli distribution with success parameter $\theta_c \sim \text{Beta}(r_c + 1, s_c + 1)$. Precedents with strong empirical support will more reliably contribute to the vote, while noisy or ambiguous ones may abstain.

Because the voting is stochastic, we simulate the prediction process using Monte Carlo sampling [17]. For each simulation run:

1. Sample θ_c from the learned Beta distribution for each precedent.
2. Sample a Bernoulli variable with parameter θ_c to decide whether the precedent casts a vote (1) or abstains (0).
3. Sum all votes to determine a predicted side.

Repeating this procedure many times yields an empirical probability mass function over vote sums. As the vote sum is integer-valued, we work with a probability mass function (pmf) rather than a density. We approximate this pmf by Monte Carlo sampling. We can then quantify prediction confidence by computing the proportion of simulations where the total vote aligns with the predicted decision. Let $f(v)$ denote the empirical probability mass function over total vote sums v. If the predicted decision corresponds to the sign of the sum (positive for π, negative for δ), these are:

$$\text{Conf} = \begin{cases} \sum_{v>0} f(v), & \text{if the predicted side is } \pi, \\ \sum_{v<0} f(v), & \text{if the predicted side is } \delta \;. \end{cases}$$

This confidence score refines binary predictions with epistemic insight and allows selective decision-making under uncertainty.

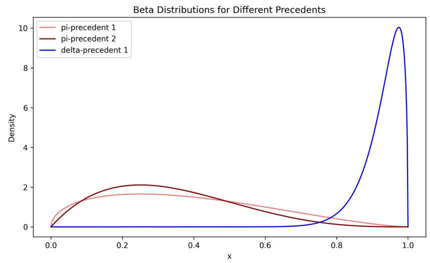
What an Explanation Looks Like in the Bayesian Model. Consider a synthetic case where two precedents support the plaintiff (π) and one supports the defendant (δ), yet judgment favors the defendant. We simulate this by assigning Beta posteriors Beta$(r_c{+}1, s_c{+}1)$ to each precedent—treating r_c as confirmations and s_c as violations of its *a fortiori* constraint—and sampling votes via Monte Carlo.

Fig. 1 displays two visualizations: the left shows the Beta distributions representing the authority (i.e., reliability of the *a fortiori* constraint) for each precedent; the right shows the results of a Monte Carlo simulation combining these votes. The figure explains a δ prediction *despite* a majority of π precedents; as panel (a) makes clear, the δ precedent is more reliable than the π precedents (peak farther right and narrower), which explains the outcome; (b) shows the *four* possible outcomes when combining the *three* opposing precedents: the left rectangle (-1) means the defendant (δ) wins, the middle rectangle (0) means tie/abstention, and the right rectangles (+1, +2) means the plaintiff (π) wins. The height of each rectangle is the relative frequency of that outcome across Monte Carlo runs. There are four rectangles because the δ precedent can contribute either 0 or -1, the π precedents can each contribute either 0 or +1, so their sum can only be -1, 0,+1 or +2.

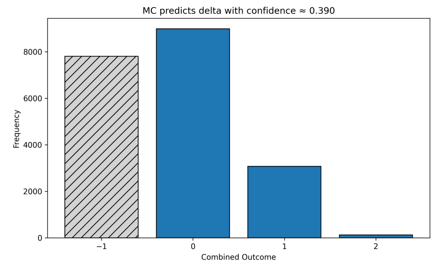
We observe that even when more precedents support one side, their influence is modulated by empirical reliability; by contrast, the Binary Majority model ignores reliability and counts all precedents equally. In this example, the δ-precedent, being more authoritative (with a Beta distribution sharply peaked near 1), overrides both π-precedents, which exhibit low mean and high variance. As a result, the model predicts a decision in favor of the defendant. However, the confidence in this prediction is relatively low, as shown by the relatively small area in the hatched region in Figure 1b.

Note that the *Binary Majority Vote Model*, which weights all precedents equally, would have favored the plaintiff, based solely on the count of precedents. This contrast shows how the Bayesian model considers not just the number but also the reliability of precedents.

Furthermore, the model's confidence score derived from the proportion of simulations favoring the predicted side can support downstream tasks such as aggregating predictions across models or deferring low-confidence cases, thereby enhancing reliability.



(a) Beta distributions: π-precedents have low expected authority and high variance (mass spread leftward), while the δ-precedent has a high mean and low variance (narrow peak near 1).



(b) Monte Carlo estimate of the vote pmf: The shaded bar marks the predicted side. Its height indicates the model's confidence in the predicted outcome.

Fig. 1: Illustrative output of the Bayesian Result Model for a synthetic case with two π-precedents and one δ-precedent.

Discussion on Prior choice. We have so far adopted the uniform prior ($\text{Beta}(1, 1)$) to model complete ignorance about precedent reliability. As in any Bayesian framework, this choice is somewhat arbitrary. Several alternatives are possible; here we outline two simple options; one is to compute *aggregate* parameters for the *a fortiori* constraint across the case base. Specifically, by considering all pairs of cases (c, c') with $c \succeq_\pi c'$, we may count

$$r_{\text{tot}} = \#\{\text{pairs where both decisions agree}\},$$
$$s_{\text{tot}} = \#\{\text{pairs where the weaker favors } \pi \text{ and the stronger } \delta\}.$$

A prior of the form $\text{Beta}(r_{\text{tot}}, s_{\text{tot}})$ then encodes a global tendency for the constraint to hold, leading to posteriors of the form

$$\theta_c \mid \text{data} \sim \text{Beta}(r_{\text{tot}} + r_c,\ s_{\text{tot}} + s_c).$$

An alternative is to maintain two separate priors, one for π-precedents and one for δ-precedents, computed in the same way.

Both approaches combine individual evidence about each case with pooled information from the entire case base, which may be advantageous when many cases are precedents for only a few others. However, they also dilute case-specific information. In a large case base where sufficient evidence is typically available for each precedent, the uninformative prior $\text{Beta}(1, 1)$ may then be preferable.

Caveat: many weak precedents. A limitation of all our models, only partly mitigated in the Bayesian one, is that many *weak* precedents can outweigh a single *strong* precedent. While the Bayesian model accounts for reliability (downweighting weak precedents), a substantial imbalance in their number can still dominate the outcome.

Future work should address this.

4 Conclusions and Future Work

We introduced a probabilistic extension of the result model that handles inconsistency while preserving the transparency of rule-based approaches. Our framework models the strength of precedents based on their observed reliability in the data. This approach offers several advantages:

1. **Uncertainty Quantification.** While classical models like the strict binary or majority variants appear to provide deterministic outputs, they in fact ignore uncertainty from data inconsistencies. The Bayesian model makes this uncertainty explicit, allowing predictions to be accompanied by meaningful confidence estimates derived from observed precedent behavior.
2. **Extension to Other Normative Modes.** Although we have focused on the *a fortiori* constraint, our approach can accommodate any admissible mode of reasoning over precedents. For example, one could assign learned weights to hypothetical analogies, rhetorical moves such as downplaying opposing factors, or domain-specific argument patterns. Each such normative mode could contribute a probabilistic "vote" toward a decision, encapsulating its strength and reliability as calibrated from data.
3. **Flexibility and Modularity.** The Bayesian framework is flexible and can incorporate application-specific beliefs about how precedents should be trusted. These priors can encode normative stances or policy constraints, enabling the system to adapt its reasoning to different institutional or legal settings.
4. **Extensibility to Richer Models.** Our framework can be naturally extended to the more expressive reason model, as well as to variants involving dimensional fact representations or explicit factor hierarchies. The probabilistic structure can incorporate rule-based justifications and reason-strength comparisons in a unified framework.

In summary, our Bayesian approach represents a promising hybrid between symbolic legal reasoning and data-driven inference. It offers a transparent and traceable decision procedure while making principled use of data to assess the reliability of conclusions. This positions it as a compelling foundation for building interpretable AI systems that reason under normative uncertainty.

We also plan to evaluate our model using the DIAS dataset ([9]), which contains court opinions addressing whether police officers in the U.S. possess the requisite constitutional suspicion to make a prolonged detention of a motorist suspected of trafficking or possessing drugs. The dataset includes 264 annotated vehicle inspection decisions. For benchmarking, we will implement the Strict Binary Result Model and Binary Majority Vote Model, and compare them against our approach and a finely tuned random forest classifier. Given the class imbalance, macro-averaged F1 will serve as our primary evaluation metric. Finally, we will conduct a qualitative analysis of conflict cases to assess how different models resolve inconsistencies in practice.

Acknowledgments: Work supported by WWTF [Grant ID: 10.47379/ICT23030].

Disclose or Interests: The authors have no disclosing interests to declare that are relevant to the content of this article.

References

1. Alexander, L.: Constrained by precedent'. Southern California Law Review **63**, 1 (1989)
2. Ashley, K.D.: Reasoning with cases and hypotheticals in hypo. International journal of man-machine studies **34**(6), 753–796 (1991)
3. Badra, F., Lesot, M.J.: Case-based prediction – a survey. International Journal of Approximate Reasoning **158**, 108920 (2023)
4. Canavotto, I.: Inconsistent precedents and deontic logic. In: Maranhão, J., Peterson, C., Straßer, C., van der Torre, L. (eds.) 16th International Conference, DEON 2023. pp. 317–333. College Publications (2023)
5. Canavotto, I.: Reasoning with inconsistent precedents. Artificial Intelligence and Law pp. 1–30 (2023)
6. Canavotto, I., Horty, J.F.: Piecemeal knowledge acquisition for computational normative reasoning. In: AIES '22. pp. 171–180. ACM (2022)
7. Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B.: Bayesian data analysis. Chapman and Hall/CRC (1995)
8. Gray, M., Savelka, J., Oliver, W., Ashley, K.: Toward automatically identifying legally relevant factors. In: Legal Knowledge and Information Systems, pp. 53–62. IOS Press (2022)
9. Gray, M., Savelka, J., Oliver, W., Ashley, K.: Empirical legal analysis simplified: reducing complexity through automatic identification and evaluation of legally relevant factors. Philosophical Transactions of the Royal Society A **382** (2024)
10. Horty, J.F.: The result model of precedent. Legal Theory **10**(1), 19–31 (2004)
11. Horty, J.F.: Rules and reasons in the theory of precedent. Legal theory **17**(1), 1–33 (2011)
12. Kelly, J.S.: Social choice theory: An introduction. Springer Science & Business Media (2013)
13. Letham, B., Rudin, C., McCormick, T.H., Madigan, D.: Interpretable classifiers using rules and bayesian analysis: Building a better stroke prediction model. The Annals of Applied Statistics **9**(3) (2015)
14. Peters, J.G., Bex, F.J., Prakken, H.: Model-and data-agnostic justifications with a fortiori case-based argumentation. In: Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law. pp. 207–216 (2023)
15. Peters, J.G., Bex, F.J., Prakken, H., et al.: Justifications derived from inconsistent case bases using authoritativeness. In: Proceedings of ArgXAI 2022. vol. 3209, pp. 1–13. CEUR WS (2022)
16. Rempell, S.: Factors. 70 Buff. L. Rev. 1755 (2022)
17. Robert, C.P., Casella, G., Casella, G.: Monte Carlo statistical methods, vol. 2. Springer (1999)
18. Schoenmackers, S., Davis, J., Etzioni, O., Weld, D.S.: Learning first-order horn clauses from web text. In: Proceedings of the 2010 conference on empirical methods in natural language processing. pp. 1088–1098 (2010)
19. van Woerkom, W., Grossi, D., Prakken, H., Verheij, B.: A fortiori case-based reasoning: from theory to data. J. of Artificial Intelligence Research **81**, 401–441 (2024)
20. von Wright, G.H.: Deontic logics. American Philosophical Quarterly **4**(2), 136–143 (1967)