Economics Bulletin

Volume 46, Issue 3

Axiomatic characterizations of belief updating under ambiguity

Shinya Sugiura
Department of Economics, Hitotsubashi University

Abstract

We characterize three rules for updating a convex capacity at the level of beliefs, with no reference to acts or preferences: the Fagin-Halpern, maximum likelihood, and Dempster-Shafer rules. Two axioms, Caution and Unanimity, force the posterior to equal the lower envelope of the Bayesian updates of a set of admissible priors; the choice of that set selects the rule.

Citation: Shinya Sugiura, (2026) "Axiomatic characterizations of belief updating under ambiguity", *Economics Bulletin*, Volume 46, Issue 3, pages 1039-1044
Contact: Shinya Sugiura - shinya.sugiura06@gmail.com.
Submitted: May 31, 2026. **Published:** September 30, 2026.

1. Introduction

How an agent should update her beliefs after observing an event depends on how those beliefs are represented. When beliefs are given by a single probability measure, the answer is Bayes' rule, which has a simple geometric interpretation: observing an event A removes the states outside A and rescales the probabilities inside A proportionally. Alós-Ferrer and Mihm (2023) show that Bayes' rule is characterized by a few simple properties, with no reference to acts or preferences.

The situation is different when beliefs are ambiguous. Such beliefs can be represented in several ways. In a leading one, the agent's beliefs are described by a capacity v—a monotone but possibly non-additive set function (Schmeidler 1989)—together with its core, the set of priors that dominate it. Updating is then no longer unique, and three rules have become standard. The Dempster–Shafer rule updates through the conjugate capacity (Dempster 1967; Shafer 1976); the Fagin–Halpern, or generalized Bayes, rule updates every prior in the core and takes the lower envelope (Fagin and Halpern 1991; Jaffray 1992; Walley 1991); and the maximum likelihood rule keeps only the priors that make the observed event most likely (Gilboa and Schmeidler 1993). Each rule has been justified by the conditional preferences it induces (Gilboa and Schmeidler 1993). This raises the question of what each rule is as a map on beliefs alone.

This note addresses that question. We characterize the three rules at the level of beliefs, with no acts, utilities, or preferences—a belief-side counterpart to Gilboa and Schmeidler (1993), in the spirit of the preference-free characterization of Bayes' rule by Alós-Ferrer and Mihm (2023). We introduce no new rule. The proportionality behind Bayes' rule does not extend to this setting, because a capacity and its conjugate generate several competing ratios; what matters instead is which priors the agent still regards as admissible after the observation.

Two axioms make this precise. Caution requires the posterior to be no larger than the Bayesian update of any admissible prior, and Unanimity requires it to be at least as large as any common lower bound of these updates. Together they force the posterior to equal the lower envelope of the admissible updates. The admissible set then determines the rule: the whole core gives the Fagin–Halpern rule (Proposition 1), the likelihood-maximizing priors give the maximum likelihood rule (Proposition 2), and on convex capacities the Dempster–Shafer rule coincides with the latter (Corollary 1).

2. Setting

Let Ω be a finite state space and let $\mathscr{C} := 2^{\Omega}$ be the collection of events. A capacity on $\mathscr{C}$ is a normalized monotone set function $v : \mathscr{C} \to [0,1]$; that is, $v(\emptyset) = 0$, $v(\Omega) = 1$, and $v(B) \leq v(B')$ whenever $B \subseteq B'$. Write $\mathscr{V}$ for the set of capacities and $\Pi \subseteq \mathscr{V}$ for the set of additive probability measures on $\mathscr{C}$. For each event B, let $\neg B := \Omega \setminus B$. The conjugate capacity of v is $\overline{v}(B) := 1 - v(\neg B)$, and the core of v is

$$\operatorname{core}(v) := \{p \in \Pi : p(B) \geq v(B) \text{ for all } B \in \mathscr{C}\}.$$

A capacity v is convex if $v(B \cup B') + v(B \cap B') \geq v(B) + v(B')$ for all $B, B' \in \mathscr{C}$. For finite Ω, every convex capacity has a nonempty core and is its lower envelope:

$$v(B) = \min_{p \in \operatorname{core}(v)} p(B), \qquad \overline{v}(B) = \max_{p \in \operatorname{core}(v)} p(B) \qquad \text{for all } B \in \mathscr{C}.$$

In particular $v \leq \overline{v}$, with equality everywhere if and only if v is additive.

We restrict attention throughout to convex capacities. This is the natural domain for the multiple-priors interpretation, and the one on which the three rules below have a common description in terms of the core. Conditioning on A is meaningful only when A is not considered impossible, so we take as the common domain

$$\mathscr{D} := \{(v,A) : v \text{ a convex capacity},\ A \in \mathscr{C},\ v(A) > 0\}.$$

An updating function is a map $f : \mathscr{D} \to \mathscr{V}$ sending a pair (v,A) to a posterior capacity $f(v,A)$; we read $f(v,A)(B)$ as the posterior belief in B. For any $(v,A) \in \mathscr{D}$, every prior $p \in \mathrm{core}(v)$ has $p(A) \geq v(A) > 0$, so the Bayesian conditional $p(B \mid A) := p(B \cap A)/p(A)$ is well defined throughout the core, and $\max_{p \in \mathrm{core}(v)} p(A) = \overline{v}(A) > 0$.

The three rules are defined below.

Definition 1 (Fagin–Halpern rule)**.** For $(v,A) \in \mathscr{D}$,

$$f_{\mathrm{FH}}(v,A)(B) := \frac{v(B \cap A)}{v(B \cap A) + \overline{v}(\neg B \cap A)} \qquad \text{for all } B \in \mathscr{C}.$$

Definition 2 (Dempster–Shafer rule)**.** For $(v,A) \in \mathscr{D}$,

$$f_{\mathrm{DS}}(v,A)(B) := \frac{v(\neg A \cup B) - v(\neg A)}{\overline{v}(A)} \qquad \text{for all } B \in \mathscr{C}.$$

Definition 3 (Maximum likelihood rule)**.** For $(v,A) \in \mathscr{D}$, let

$$C_A(v) := \operatorname*{arg\,max}_{p \in \mathrm{core}(v)} p(A)$$

be the set of priors in the core that make A most likely. The maximum likelihood rule is

$$f_{\mathrm{ML}}(v,A)(B) := \min_{p \in C_A(v)} p(B \mid A) \qquad \text{for all } B \in \mathscr{C}.$$

On $\mathscr{D}$ all three rules are well defined. The set $C_A(v)$ is nonempty, because the core is a nonempty compact set and $p \mapsto p(A)$ is continuous, and $p(A) = \overline{v}(A) > 0$ for every $p \in C_A(v)$. The Fagin–Halpern denominator is positive: take $p^* \in \mathrm{core}(v)$ with $p^*(B \cap A) = v(B \cap A) = \min_{p \in \mathrm{core}(v)} p(B \cap A)$; since $\overline{v}(\neg B \cap A) \geq p^*(\neg B \cap A)$,

$$v(B \cap A) + \overline{v}(\neg B \cap A) \ \geq\ p^*(B \cap A) + p^*(\neg B \cap A) \ =\ p^*(A) \ \geq\ v(A) \ >\ 0.$$

3. Axioms and characterizations

Throughout, $f : \mathscr{D} \to \mathscr{V}$ is an updating function, and each axiom holds at every $(v,A) \in \mathscr{D}$. Neither axiom refers to any closed-form rule; both speak only of the agent's admissible priors and their Bayesian updates.

Axiom 1 (Caution)**.** For all $B \in \mathscr{C}$ and all $p \in \mathrm{core}(v)$,

$$f(v,A)(B) \leq p(B \mid A).$$

Axiom 2 (Unanimity)**.** For all $B \in \mathscr{C}$ and all $t \in [0,1]$, if $p(B \mid A) \geq t$ for every $p \in \mathrm{core}(v)$, then $f(v,A)(B) \geq t$.

Caution makes the posterior a lower bound on the admissible updates $p(\cdot \mid A)$, $p \in \mathrm{core}(v)$, and Unanimity makes it at least as large as every common lower bound of them. Since the lower envelope is the largest such bound, the two axioms together force the posterior to equal that envelope.

Proposition 1. *An updating function $f : \mathscr{D} \to \mathscr{V}$ is the Fagin–Halpern rule, $f = f_{\mathrm{FH}}$, if and only if it satisfies Caution and Unanimity.*

Proof. By the classical identity of Fagin and Halpern (1991) and Jaffray (1992),

$$f_{\mathrm{FH}}(v,A)(B) = \min_{p \in \mathrm{core}(v)} p(B \mid A) \qquad \text{for all } (v,A) \in \mathscr{D},\ B \in \mathscr{C}.$$

Necessity. Fix $(v,A) \in \mathscr{D}$ and $B \in \mathscr{C}$. Since $f_{\mathrm{FH}}(v,A)(B) = \min_{p \in \mathrm{core}(v)} p(B \mid A)$, we have $f_{\mathrm{FH}}(v,A)(B) \leq p(B \mid A)$ for every $p \in \mathrm{core}(v)$, which is Caution. If $t \in [0,1]$ satisfies $p(B \mid A) \geq t$ for every $p \in \mathrm{core}(v)$, then $f_{\mathrm{FH}}(v,A)(B) = \min_{p \in \mathrm{core}(v)} p(B \mid A) \geq t$, which is Unanimity.

Sufficiency. Suppose f satisfies Caution and Unanimity, and fix $(v,A) \in \mathscr{D}$ and $B \in \mathscr{C}$. By Caution, $f(v,A)(B) \leq p(B \mid A)$ for every $p \in \mathrm{core}(v)$, so $f(v,A)(B) \leq \min_{p \in \mathrm{core}(v)} p(B \mid A)$. Put $t := \min_{p \in \mathrm{core}(v)} p(B \mid A)$. Then $p(B \mid A) \geq t$ for every $p \in \mathrm{core}(v)$, so Unanimity gives $f(v,A)(B) \geq t$. Hence $f(v,A)(B) = \min_{p \in \mathrm{core}(v)} p(B \mid A) = f_{\mathrm{FH}}(v,A)(B)$. Since (v,A) and B were arbitrary, $f = f_{\mathrm{FH}}$. □

The key step is the minimization over $\mathrm{core}(v)$, not a renormalization.

To characterize the maximum likelihood rule, we impose the same axioms, but only over the priors that maximize the likelihood of A.

Axiom 3 (Likelihood Caution)**.** For all $B \in \mathscr{C}$ and all $p \in C_A(v)$,

$$f(v,A)(B) \leq p(B \mid A).$$

Axiom 4 (Likelihood Unanimity)**.** For all $B \in \mathscr{C}$ and all $t \in [0,1]$, if $p(B \mid A) \geq t$ for every $p \in C_A(v)$, then $f(v,A)(B) \geq t$.

Proposition 2. *An updating function $f : \mathscr{D} \to \mathscr{V}$ is the maximum likelihood rule, $f = f_{\mathrm{ML}}$, if and only if it satisfies Likelihood Caution and Likelihood Unanimity.*

Proof. The argument is the same as for Proposition 1, with $C_A(v)$ in place of $\mathrm{core}(v)$ and Definition 3, $f_{\mathrm{ML}}(v,A)(B) = \min_{p \in C_A(v)} p(B \mid A)$, in place of the Fagin–Halpern identity. For necessity, $f_{\mathrm{ML}}(v,A)(B) \leq p(B \mid A)$ for every $p \in C_A(v)$, which is Likelihood Caution, and $f_{\mathrm{ML}}(v,A)(B) \geq t$ whenever $p(B \mid A) \geq t$ for every $p \in C_A(v)$, which is Likelihood Unanimity. For sufficiency, suppose f satisfies the two axioms; Likelihood Caution gives $f(v,A)(B) \leq \min_{p \in C_A(v)} p(B \mid A)$, and with $t := \min_{p \in C_A(v)} p(B \mid A)$, Likelihood Unanimity gives $f(v,A)(B) \geq t$, so $f(v,A)(B) = \min_{p \in C_A(v)} p(B \mid A) = f_{\mathrm{ML}}(v,A)(B)$. □

The two characterizations differ only in the set of priors over which the axioms range—the whole core versus the likelihood-maximizing subset $C_A(v)$—so choosing a rule amounts to choosing which priors remain admissible. In each characterization the two axioms are independent (Appendix A).

4. The Dempster–Shafer rule

On convex capacities the Dempster–Shafer rule and the maximum likelihood rule coincide (Gilboa and Schmeidler 1993; Jaffray 1992). Proposition 2 therefore characterizes the Dempster–Shafer rule on this domain as well.

Corollary 1. *An updating function $f : \mathscr{D} \to \mathscr{V}$ is the Dempster–Shafer rule, $f = f_{\mathrm{DS}}$, if and only if it satisfies Likelihood Caution and Likelihood Unanimity.*

Proof. On $\mathscr{D}$ we have $f_{\mathrm{DS}} = f_{\mathrm{ML}}$ (Gilboa and Schmeidler 1993; Jaffray 1992), and the claim follows from Proposition 2. □

Remark 1. The Dempster–Shafer and maximum likelihood rules are distinct on non-convex capacities; indeed their equivalence characterizes convexity (Ui and Kajii 2005). The characterization above is therefore specific to the convex domain. Proposition 2 still characterizes the maximum likelihood rule on its own; the convex domain simply does not separate the two rules, because there they coincide.

5. Discussion

Each rule is the lower envelope of the Bayesian updates of a set of admissible priors, and the rules differ only in that set: the whole core for the Fagin–Halpern rule, and the likelihood-maximizing priors for the maximum likelihood rule. This is the familiar contrast between retaining all admissible models and selecting among them by likelihood.

Like Alós-Ferrer and Mihm (2023), we work without preferences. In their setting, additivity extends the posterior from elementary events to compound ones; here that role is played by consistency with the core, through Caution and Unanimity. This is also why a proportionality requirement alone would be uninformative in our setting: it would merely restate a closed-form rule.

The existing foundations for these rules are, by contrast, preference-based. Gilboa and Schmeidler (1993) axiomatize the Dempster–Shafer and maximum likelihood rules; Pires (2002) and Eichberger, Grant, and Kelsey (2007) axiomatize the Fagin–Halpern (full Bayesian) rule; and Cheng (2022) axiomatizes the relative maximum likelihood family, which nests full Bayesian and maximum likelihood updating of a set of priors, and hence, on convex capacities, the Fagin–Halpern and Dempster–Shafer rules. Our results are the belief-side counterpart to this line of work. At the level of beliefs, Horie (2006) unifies the naive, Dempster–Shafer, and Fagin–Halpern rules in a single constructive representation; we instead give axioms that identify each rule separately, so that the choice of rule reduces to the choice of admissible priors. More broadly, the lower-envelope view of conditioning is the capacity case of coherent updating in the imprecise-probability literature (Walley 1991).

Appendix A. Independence of the axioms

The two axioms in each characterization are logically independent. Define two updating functions on $\mathscr{D}$:

$$f^0(v,A)(B) := \begin{cases} 1 & B = \Omega, \\ 0 & B \neq \Omega, \end{cases} \qquad f^1(v,A)(B) := \begin{cases} 0 & B = \emptyset, \\ 1 & B \neq \emptyset. \end{cases}$$

Both are capacities for each (v,A), hence updating functions.

Proposition 3. *Caution and Unanimity are independent: f^0 satisfies Caution but not Unanimity, and f^1 satisfies Unanimity but not Caution. The same two functions show that Likelihood Caution and Likelihood Unanimity are independent.*

Proof. f^0 satisfies Caution, since $f^0(v,A)(B) = 0 \leq p(B \mid A)$ for $B \neq \Omega$, while $f^0(v,A)(\Omega) = 1 = p(\Omega \mid A)$. It violates Unanimity: take an additive $v = p$, so $\text{core}(v) = \{p\}$; for example $\Omega = \{1,2\}$, p uniform, $A = \Omega$, $B = \{1\}$, so that the unique prior has $p(B \mid A) = \frac{1}{2}$. Unanimity would require $f^0(v,A)(B) \geq \frac{1}{2}$, but $f^0(v,A)(B) = 0$.

f^1 satisfies Unanimity, since $f^1(v,A)(B) = 1 \geq t$ for every admissible t at every $B \neq \emptyset$, while at $B = \emptyset$ the only admissible bound is $t = 0$ and $f^1(v,A)(\emptyset) = 0 \geq 0$. It violates Caution at any B with $\emptyset \neq B \neq \Omega$ such that $p(B \mid A) < 1$ for some $p \in \text{core}(v)$, for then $f^1(v,A)(B) = 1 > p(B \mid A)$; the additive example above, with $B = \{1\}$, is one such instance.

For the likelihood axioms, replace $\text{core}(v)$ by $C_A(v)$ throughout. In the additive example $C_A(v) = \text{core}(v) = \{p\}$, so the same calculations apply. □

Remark 2. The maximum likelihood rule itself witnesses "Unanimity without Caution" in Proposition 1. Since $C_A(v) \subseteq \text{core}(v)$,

$$f_{\text{ML}}(v,A)(B) = \min_{p \in C_A(v)} p(B \mid A) \;\geq\; \min_{p \in \text{core}(v)} p(B \mid A) = f_{\text{FH}}(v,A)(B),$$

so $f_{\text{ML}} \geq f_{\text{FH}}$ and Unanimity holds; but Caution fails, because f_{ML} can exceed $p(B \mid A)$ for priors $p \in \text{core}(v) \setminus C_A(v)$. For instance, with $\Omega = \{1,2,3\}$ and the belief function with masses $m(\{1\}) = 0.4$, $m(\{2,3\}) = 0.6$, taking $A = \{1,2\}$ and $B = \{2\}$ gives $f_{\text{ML}}(v,A)(\{2\}) = 0.6$, yet the prior $p = (0.4, 0, 0.6) \in \text{core}(v)$ has $p(\{2\} \mid A) = 0$. Thus the Fagin–Halpern and maximum likelihood rules are separated by Caution alone, applied to the full core: both are lower envelopes of Bayesian updates and both satisfy Unanimity, but only the Fagin–Halpern rule satisfies Caution with respect to every prior in the core, while the maximum likelihood rule does so only with respect to the likelihood-maximizing priors.

References

Alós-Ferrer, C. and M. Mihm (2023) "An axiomatic characterization of Bayesian updating" *Journal of Mathematical Economics* **104**, 102799.

Cheng, X. (2022) "Relative Maximum Likelihood updating of ambiguous beliefs" *Journal of Mathematical Economics* **99**, 102587.

Dempster, A.P. (1967) "Upper and lower probabilities induced by a multivalued mapping" *Annals of Mathematical Statistics* **38**, 325–339.

Eichberger, J., S. Grant, and D. Kelsey (2007) "Updating Choquet beliefs" *Journal of Mathematical Economics* **43**, 888–899.

Fagin, R. and J.Y. Halpern (1991) "A new approach to updating beliefs" in *Uncertainty in Artificial Intelligence 6* by P.P. Bonissone, M. Henrion, L.N. Kanal, and J.F. Lemmer, Eds., North-Holland: Amsterdam, 347–374.

Gilboa, I. and D. Schmeidler (1993) "Updating ambiguous beliefs" *Journal of Economic Theory* **59**, 33–49.

Horie, M. (2006) "A unified representation of conditioning rules for convex capacities" *Economics Bulletin* **4**(19), 1–6.

Jaffray, J.-Y. (1992) "Bayesian updating and belief functions" *IEEE Transactions on Systems, Man and Cybernetics* **22**, 1144–1152.

Pires, C.P. (2002) "A rule for updating ambiguous beliefs" *Theory and Decision* **53**, 137–152.

Schmeidler, D. (1989) "Subjective probability and expected utility without additivity" *Econometrica* **57**, 571–587.

Shafer, G. (1976) *A Mathematical Theory of Evidence*, Princeton University Press: Princeton.

Ui, T. and A. Kajii (2005) "Equivalence of the Dempster–Shafer rule and the maximum likelihood rule implies convexity" *Economics Bulletin* **4**(10), 1–6.

Walley, P. (1991) *Statistical Reasoning with Imprecise Probabilities*, Chapman and Hall: London.