# What Do We Learn from a Conditional?

Eidon Vale

**Abstract**

What should happen to an agent's beliefs when she learns a conditional? A recent trivalent theory answers: update on its material counterpart. I argue that this is a coherent idealization, but not one selected by conditional content alone. The point is distinctively trivalent. For an ordinary bivalent proposition, truth and non-falsity pick out the same event. For a trivalent content they come apart, so conditioning on its truth, conditioning on its non-falsity, and making it certain while preserving its domain are different content-level updates. All three collapse to ordinary conditionalization in the bivalent case. Concrete acquisition introduces a further question: a diagnostic or report may carry information not exhausted by the content accepted. I use a domain Bayes factor to separate these issues and give a causal model in which a diagnostic supports a conditional while preserving its domain and changing beliefs within its void cases. The positive payoff is compositional. A trivalent semantics identifies domains for conjunctions and nested conditionals that cannot be read from a surface antecedent. It thereby determines which evidential-neutrality questions an acquisition model must answer, even though it does not answer them itself.

## 1 Introduction

Suppose that a safety engineer tells us:

> If the reactor is restarted tonight, its core will overheat.

We trust the engineer and come to accept what she says. What, exactly, have we learned?

One natural answer is that we have learned what would happen on a restart. If the decision to restart will later be made by a fair coin toss, the engineer's diagnostic expertise need not tell us anything about the toss. Yet one prominent rule for learning conditionals says otherwise. The material counterpart of the engineer's statement is the proposition that either there will be no restart or there will be overheating. Conditioning on that proposition

eliminates restart-without-overheating possibilities and renormalizes globally. If the restart decision is initially independent of whether the reactor has an overheating mechanism, with both events fair, the probability of a restart falls from one-half to one-third.

This is not obviously irrational. An unsolicited warning can be evidence that its antecedent is a live possibility. A report may reveal the source's causal model, stable-property information, or reasons for mentioning one contingency rather than another. In such cases, learning the conditional may properly change the antecedent's probability. The mistake would be to suppose that the sentence alone decides whether it does.

Trivalent semantics brings the issue into focus. On a simple trivalent account, *if $A$, then $B$* is true when $A \wedge B$, false when $A \wedge \neg B$, and void when $\neg A$. The third value marks cases in which the issue posed by the conditional does not arise.[1] Its probability is the proportion of true cases among the cases in which it is not void. Thus confidence in $A \to B$ is $P(B \mid A)$, in accordance with Adams's Thesis.[2]

Recent work by Égré et al. (2025, 2026) combines this semantics with an elegant compositional probability theory. It also defines a rule of trivalent conditionalization. Under that rule, learning $A \to B$ has the same effect on unconditional probabilities as conditioning on $\neg A \vee B$. The result is a theorem of their proposed dynamics, not an algebraic error. Moreover, it follows from a general rule rather than merely from an informal inference about certainty.

Two questions must be distinguished. The first concerns an idealized *content-level update*: if all collateral information is set aside, which posterior represents learning just the trivalent content $E$? The second concerns *total acquisition*: which posterior is warranted by the report, diagnostic result, observation, or other event through which the agent comes to accept $E$? Ordinary testimony can convey information beyond what was said, so these questions differ even for bivalent propositions. A source model is needed to answer the second.

The first question, however, presents a problem that is special to trivalence. For a bivalent proposition, the worlds where it is true are exactly the worlds where it is not false. The familiar idealization of learning just that proposition therefore identifies a single event. For a trivalent content, truth and non-falsity diverge. One can filter to the true worlds, filter out only the false worlds, or eliminate falsity while holding fixed the probability of the nonvoid domain. These operations use no collateral source information, and all reduce to ordinary conditionalization when there are no void worlds. The phrase "learn exactly $E$" does not by itself choose among them.

This is the *trivalent bridge problem*. Égré, Rossi, and Sprenger offer a clear solution: treat

[1] The idea has a long history in work on conditional assertion and conditional bets; see Cooper (1968) and Belnap (1970). For recent probabilistic treatments, see Lassiter (2020) and Égré et al. (2026).

[2] See Adams (1975) and Edgington (1995).

learning as the elimination of falsity while preserving prior ratios among all nonfalse worlds. My claim is not that this rule is incoherent. It is that the rule is an additional certainty-based norm, not the unique content-level update supplied by the semantics. It is also not, without further structure, a general account of total acquisition. Those are distinct limitations, and the causal model developed below bears directly only on the second.

The positive thesis is one of *structured underdetermination*. A selected trivalent semantics supplies a true/false/void partition. That partition reveals a dimension of evidence that bivalent materialization hides: whether a signal favors the content's domain, its void cases, or neither. I represent this dimension by a *domain Bayes factor*. If a signal is independent of domain entry, Bayes's theorem requires the domain probability to be preserved; but that independence is a feature of an acquisition model, not of the sentence alone. Moreover, preserving a domain does not determine the rest of the posterior.

Compositionality gives the proposal its main positive payoff. For a simple $A \to B$, the domain is just $A$. For conjunctions and nested conditionals, the domain need not be any surface antecedent. I show, in particular, that two complex contents can have the same falsity set—and hence the same certainty condition—while imposing different domain-neutrality requirements on an acquisition model. Competing trivalent semantics can likewise disagree about the domain of a nested conditional. The framework therefore turns semantic differences into precise, assessable questions about evidence.

The resulting position is close in spirit to source-sensitive approaches in the existing literature. Douven (2012) argues that explanatory relations affect conditional learning, while Eva et al. (2020) distinguish interpreting evidence from updating on an interpretation and explicitly use an independent "disabler" to explain antecedent preservation. I do not claim to rediscover source sensitivity. The new claim is that even a unified, compositional trivalent semantics and probability theory leaves learning underdetermined in a systematic way. Trivalence identifies the missing coordinate, and thereby lets us diagnose what different update rules add to the semantics. Its further claim is that the bridge problem survives even after collateral source information is idealized away, because truth and non-falsity remain distinct semantic candidates for the learned event.

Section 2 reconstructs the materialization result and compares three content-level bridges. Section 3 introduces domain evidence. Section 4 gives the causal model and marks its dialectical limits. Section 5 develops the payoff for conjunction and nesting. Section 6 treats minimum-change updating as a conservative completion rather than an autonomous theory of learning. A short representation result appears in the Appendix.

## 2 Certainty does not determine learning

Let $W$ be a finite set of possible worlds, and let $P$ be a probability distribution over them. A trivalent content $E$ divides $W$ into three cells:

$$E^{\mathsf{T}}, \qquad E^{\mathsf{F}}, \qquad E^{\mathsf{V}},$$

the worlds at which $E$ is respectively true, false, and void. Write

$$D_E = E^{\mathsf{T}} \cup E^{\mathsf{F}}$$

for the *domain* of $E$: the worlds at which the issue settled by $E$ arises. Its trivalent probability is

$$\pi_P(E) = \frac{P(E^{\mathsf{T}})}{P(E^{\mathsf{T}}) + P(E^{\mathsf{F}})}, \tag{1}$$

provided $P(D_E) > 0$. This is just ordinary probability measured within the nonvoid domain. If $E$ is bivalent, there are no void worlds and (1) reduces to $P(E)$.

For a simple conditional with bivalent antecedent and consequent,

$$(A \to B)^{\mathsf{T}} = A \wedge B, \qquad (A \to B)^{\mathsf{F}} = A \wedge \neg B, \qquad (A \to B)^{\mathsf{V}} = \neg A.$$

Equation (1) then gives $\pi_P(A \to B) = P(B \mid A)$. The numerator measures the $A$-cases in which the conditional succeeds; the denominator measures all the cases in which it is tested. The $\neg A$-cases belong to neither.

The materialization result concerns what happens when an agent learns $E$. Égré et al. (2026, §6) define trivalent conditionalization by

$$\pi_P^E(H) = \pi_P(E \to H). \tag{2}$$

Their Proposition 5 establishes that formulas equivalent in the certainty-preserving logic induce the same update. The accompanying rationale is that learning $E$ makes $E$ certain; hence, if certainty in $E$ and certainty in $E'$ always coincide, learning either should yield the same posterior.

For a simple conditional, that principle yields materialization. Let

$$M_E = E^{\mathsf{T}} \cup E^{\mathsf{V}} = W \setminus E^{\mathsf{F}}$$

be $E$'s *material shadow*. In the simple case, $M_E$ is $\neg A \vee B$. Since $A \to B$ and $\neg A \vee B$ have

the same falsifying worlds, they are equivalent in the certainty logic. Proposition 6 therefore implies that updating on the former has the same effect on the underlying distribution as ordinary conditioning on the latter:

$$P^E_{\mathsf{Mat}}(X) = P(X \mid M_E). \tag{3}$$

It removes the false worlds and preserves all prior probability ratios among the true and void worlds that remain. Their Updating Theorem then shows that this result also minimizes every $f$-divergence from the prior subject to making $E$ certain.[3]

The content-level issue can be stated without mentioning a source. Given only $P$ and the trivalent partition, and assuming the relevant cells have positive prior probability, at least three success-inducing operations are available. *Truth filtering* conditions on $E^{\mathsf{T}}$. *Non-falsity filtering* conditions on $E^{\mathsf{T}} \cup E^{\mathsf{V}}$ and is just materialization. A *domain-stable* update sets the false-cell mass to zero, transfers that mass to the true cell, and preserves the prior mass of the domain: its posterior cell totals are $P(D_E), 0, P(E^{\mathsf{V}})$. With the natural conservative completion, prior ratios are preserved within the true and void cells. Each operation makes $E$ certain; none appeals to information about a speaker or diagnostic.

The reactor case displays the difference. Let $R$ say that the reactor is restarted, $M$ that its present mechanism would produce overheating on a restart, and $O$ that overheating actually occurs, with $O = R \wedge M$. Suppose initially that $R$ and $M$ are independent and fair. Table 1 compares the three content-level bridges for $R \to O$.

| World | Prior | Truth | Non-falsity | Domain-stable |
|---|---|---|---|---|
| $R \wedge M$ | 1/4 | 1 | 1/3 | 1/2 |
| $R \wedge \neg M$ | 1/4 | 0 | 0 | 0 |
| $\neg R \wedge M$ | 1/4 | 0 | 1/3 | 1/4 |
| $\neg R \wedge \neg M$ | 1/4 | 0 | 1/3 | 1/4 |

Table 1: Three content-level bridges for the same trivalent content.

All three updates make $P(O \mid R) = 1$. Truth filtering treats learning the conditional as learning that its true case obtains, so $P(R)$ rises to one. Non-falsity filtering eliminates $R \wedge \neg O$ and renormalizes globally, so $P(R)$ falls from 1/2 to 1/3. The domain-stable update reallocates the probability of $R \wedge \neg M$ to $R \wedge M$ while preserving the total mass of the no-restart worlds, so $P(R)$ remains 1/2. The three operations disagree about how void worlds should fare even though they make the same conditional certain.

[3]These are Definition 8, Propositions 5 and 6, and Theorem 4 of Égré et al. (2026). The Reduction Theorem extends the result to arbitrary compounds by identifying a simple conditional with the same trivalent semantic profile.

This three-way disagreement has no bivalent analogue. If $E$ is bivalent, $E^{\mathsf{V}}$ is empty, truth and non-falsity coincide, and the domain is all of $W$. Truth filtering, non-falsity filtering, and the conservative domain-stable update then all reduce to ordinary conditioning on $E$. Source sensitivity remains possible—a report that $E$ may reveal more than $E$—but once that collateral information is idealized away, the semantic truth set provides a unique filter. In the trivalent case, the idealization itself leaves a choice.

The point is general. To be certain of a trivalent content $E$ is only to assign no probability to $E^{\mathsf{F}}$. That constraint does not settle the division of probability between $E^{\mathsf{T}}$ and $E^{\mathsf{V}}$. Yet the distinction between those cells is exactly what separates an indicative conditional from its material shadow. Trivalent conditionalization makes true and void worlds dynamically alike in one respect: after falsity is eliminated, their prior ratio is preserved. That is a substantive bridge from content to change, not a consequence of certainty alone.

This observation does not refute trivalent conditionalization as a formal proposal. It has a principled reading: learning $E$ is idealized as imposing exactly the certainty constraint $Q(E^{\mathsf{F}}) = 0$, and conservativity preserves all prior odds among the remaining worlds. On that reading, the Updating Theorem supplies a strong minimum-change rationale. What remains unsupported is the claim that the certainty constraint exhausts the content-level learning task. Truth filtering and domain stability encode different answers using the same semantic partition, and they too coincide with ordinary conditioning on the bivalent fragment. The disagreement is therefore not about the algebra of trivalent conditionalization. It is about which feature of a trivalent content plays the role that the truth set plays in ordinary Bayesian learning.

An actual signal can, of course, produce any of these effects or none. That is a second layer of underdetermination. The next section represents it without conflating it with the content-level choice.

# 3 The domain-evidence coordinate

Let

$$t = P(E^{\mathsf{T}}), \qquad n = P(E^{\mathsf{F}}), \qquad v = P(E^{\mathsf{V}}), \qquad r = t + n = P(D_E),$$

and assume for now that all three cells have positive probability. Any posterior $Q$ with $Q(D_E) > 0$ can be described in part by two coordinates:

$$s = \pi_Q(E), \qquad r' = Q(D_E).$$

The first says how strongly the posterior favors truth over falsity within the domain. The second says how strongly it favors entering the domain rather than remaining in the void cases. Together they fix the posterior cell masses:

$$Q(E^{\mathsf{T}}) = sr', \qquad Q(E^{\mathsf{F}}) = (1-s)r', \qquad Q(E^{\mathsf{V}}) = 1 - r'. \tag{4}$$

They do not fix how probability is distributed inside the cells.

Now let $S$ be the event that an agent receives a report, diagnostic output, or other learning signal. Define the signal's *domain Bayes factor* by

$$\kappa_S = \frac{P(S \mid D_E)}{P(S \mid E^{\mathsf{V}})}. \tag{5}$$

The ratio is defined when the denominator is positive; zero likelihoods yield the corresponding limiting values. Bayes's theorem gives

$$P(D_E \mid S) = \frac{\kappa_S r}{\kappa_S r + v}. \tag{6}$$

Thus $\kappa_S > 1$ means that the signal supports entering the domain, $\kappa_S < 1$ means that it supports voidness, and $\kappa_S = 1$ means that it is neutral between them.

Domain Preservation is the neutral case:

$$\kappa_S = 1 \quad \Longrightarrow \quad P(D_E \mid S) = P(D_E). \tag{7}$$

For a simple $A \to B$, this is antecedent preservation. Such preservation is familiar from Adams conditioning and earlier preservation accounts.[4] Equation (7) does not yet give an updating rule. It fixes one marginal of the posterior.

Materialization occupies a different point on the same scale. A formal likelihood representation of categorical materialization assigns equal positive likelihood to $E^{\mathsf{T}}$ and $E^{\mathsf{V}}$ and zero likelihood to $E^{\mathsf{F}}$. Consequently,

$$P(S_{\mathsf{Mat}} \mid D_E) = \frac{t}{r}, \qquad P(S_{\mathsf{Mat}} \mid E^{\mathsf{V}}) = 1,$$

and hence

$$\kappa_{S_{\mathsf{Mat}}} = \frac{t}{r} = \pi_P(E). \tag{8}$$

Unless $E$ was already certain, materialization has a domain Bayes factor below one. This is

[4] See Adams (1975), Bradley (2000), and Douven (2012). The present terminology emphasizes that, for complex trivalent contents, the preserved event need not be a surface antecedent.

a description of the rule's probabilistic effect. It is not the claim that proponents of materialization are committed to a psychological source with these literal reporting propensities. They may instead offer the rule as a normative bridge governing content-level learning.

The comparison reveals a continuum rather than a forced choice. The following elementary result records the extent of the freedom.

**Proposition 1** (Cell-coarse realization)**.** *Suppose $t, n, v > 0$. For every target $s \in [0, 1]$ and every finite $\kappa > 0$, there is a signal whose likelihood is constant within each semantic cell, whose domain Bayes factor is $\kappa$, and after which the trivalent probability of $E$ is $s$.*

*Proof.* Set $r_\kappa = \kappa r/(\kappa r + v)$ and assign the three cells relative likelihoods

$$\lambda_{\mathsf{T}} = \frac{s r_\kappa}{t}, \qquad \lambda_{\mathsf{F}} = \frac{(1-s) r_\kappa}{n}, \qquad \lambda_{\mathsf{V}} = \frac{1 - r_\kappa}{v}.$$

Their prior-weighted sum is one, so Bayes's theorem gives posterior cell masses $s r_\kappa$, $(1-s) r_\kappa$, and $1 - r_\kappa$. Multiplying all three numbers by a sufficiently small common constant turns them into probabilities of a signal without changing the posterior or any likelihood ratio. Equations (5) and (6) then give the required domain Bayes factor. □

The proposition is a possibility result, not a vindication of any such channel. Likelihoods can be reverse-engineered for any absolutely continuous posterior. Its point is negative: neither trivalent content nor the demand for success selects one member of the family. Moreover, signals can distinguish worlds within a cell, in which case the same $s$ and $\kappa$ are compatible with many further posteriors. The domain effect and within-cell change are both left open.

The boundary cases explain the restriction to finite positive $\kappa$. When $\kappa = \infty$, a signal has zero likelihood in the void cell and conditions that cell away; any target $s$ can still be realized by its relative likelihoods in the true and false cells. Truth filtering is the case $s = 1$ at this boundary. As $\kappa$ tends to zero, posterior domain mass tends to zero. At the endpoint the posterior is wholly void and $\pi_Q(E)$ is undefined unless the all-void case is assigned a value by stipulation. The open boundary is therefore substantive, not merely notational.

There is an important qualification. Source neutrality is prior-relative. $P(S \mid D_E)$ averages the signal's likelihood over the true and false cells using the agent's conditional prior. In the categorical, cell-coarse case, $P(S \mid E^{\mathsf{F}}) = 0$, and neutrality requires

$$\frac{P(S \mid E^{\mathsf{T}})}{P(S \mid E^{\mathsf{V}})} = \frac{1}{P(E^{\mathsf{T}} \mid D_E)}. \tag{9}$$

A single channel with fixed cellwise likelihoods will not satisfy this equation for arbitrary priors. It is therefore misleading to call neutrality an intrinsic property of a source. It is an

independence relation between a signal and a domain in a probabilistic model.

This does not make neutrality incapable of independent support. Experimental randomization or causal structure can entail the relevant independence for a family of priors that respect that structure. But the structure must be stated. The next section supplies it for the reactor case.

# 4 A generative model of conditional evidence

The four state descriptions in Table 1 are enough to display three posteriors, but not enough to explain why an engineer's evidence about the mechanism should be neutral about a restart. For that we need to represent the information-producing process.

Let $S$ say that the engineer's diagnostic returns a positive result. Retain the variables $R$, $M$, and $O$ from Section 2. Suppose that the restart is fixed by a later randomizer, independently of $M$; the diagnostic depends on $M$ but not on the randomizer; and overheating occurs exactly when the reactor is restarted with the overheating mechanism. The causal structure is

$$R \longrightarrow O \longleftarrow M \longrightarrow S,$$

with joint distribution

$$P(R, M, O, S) = P(R)P(M)P(S \mid M)\mathbf{1}[O = R \wedge M]. \tag{10}$$

Assume the diagnostic is informative: $P(S \mid M) > P(S \mid \neg M)$, with $0 < P(M) < 1$.

This model gives the desired result for a principled reason. Because $R$ is randomized independently of $M$ and the diagnostic depends only on $M$,

$$P(R \mid S) = P(R). \tag{11}$$

At the same time,

$$P(O \mid R, S) = P(M \mid S) > P(M) = P(O \mid R). \tag{12}$$

The diagnostic therefore raises the probability of the conditional within its restart domain while leaving the probability of entering that domain fixed. If the diagnostic is perfectly reliable, $P(M \mid S) = 1$, and the conditional $R \to O$ becomes certain.

The same model exposes a limitation of the domain-stable completion in Table 1. At no-restart worlds the conditional is void, but the diagnostic still informs us about the mechanism:

$$P(M \mid \neg R, S) = P(M \mid S) > P(M).$$

Thus the signal preserves the total probability of the void cell while changing probability ratios *within* it. A total update that leaves every void world untouched would discard information the source actually provides. Domain neutrality licenses Domain Preservation, not cellwise invariance.

The scope of this conclusion matters. The causal model does not show that trivalent conditionalization is an irrational update on the abstract content $R \to O$. A defender may factor the epistemic episode into two stages: first extract the conditional content from $S$, then apply a content-level rule to that content. The model shows instead that trivalent conditionalization is not the total update warranted by $S$, because $S$ also bears on $M$. Nor does the instruction to retain "only the conditional content" yet determine how that information should be abstracted. Treating the retained content as truth, non-falsity, or domain-stable success yields the three posteriors in Table 1. The causal example establishes the insufficiency of a content rule for total acquisition; the trivalent bridge problem explains why idealization does not by itself select a unique content rule.

This is structurally close to the analysis of the Sundowners case by Eva et al. (2020). Their "disabler" variable records an independent reason why rain would prevent sundowners; updating on that reason together with the material conditional preserves the probability of rain. The reactor mechanism $M$ plays the same role. That earlier analysis anticipates the appeal to source independence, and the present proposal should not be advertised as a new justification of antecedent preservation. Its role here is different: it shows what an adequate acquisition model adds to a trivalent semantic partition, and why a preserved domain still leaves the full posterior underdetermined.

Other causal structures produce other domain Bayes factors. A speaker may be more likely to issue a warning when she expects its antecedent to occur. A common cause may influence both the source's evidence and domain entry. Learning the conditional may also alter an agent's later decision about whether to make the antecedent true. In these cases $S$ and $D_E$ are not independent, and Domain Preservation is not warranted.[5]

# 5 What trivalence contributes

The argument so far denies that semantics selects a privileged acquisition model. It does not make semantics idle. Trivalence determines the partition relative to which different dimensions of learning can be separated. The division of labor is summarized in Table 2.

For simple $A \to B$, the domain is $A$, and the domain-evidence question is the familiar

[5]For source and context effects, see Collins et al. (2020); for causal approaches to conditional learning, see Vandenburgh (2021).

| Question | Formal object | Supplied by |
|---|---|---|
| Where does the content succeed, fail, or not apply? | $E^{\mathsf{T}}, E^{\mathsf{F}}, E^{\mathsf{V}}$ | semantics |
| Does the signal favor the content's domain or its void cases? | domain Bayes factor $\kappa_S$ | acquisition model |
| Does the signal carry finer information within a cell? | statewise likelihoods | acquisition model or a conservativity assumption |

Table 2: Three questions in conditional learning.

question whether the evidence changes the probability of the antecedent. The advantage of a compositional trivalent semantics appears with complex contents. The Reduction Theorem of Égré et al. (2026) associates any compound $E$ with a simple conditional $A_E \to B_E$ having the same trivalent profile. Its domain is the bivalent event $A_E$. The theory therefore tells us which event plays the role of applicability even when there is no single surface antecedent. Table 3 displays three cases. Here $\supset$ is the bivalent material conditional.

| Content | Equivalent simple conditional | Domain |
|---|---|---|
| $(A \to B) \wedge (C \to D)$ | $(A \vee C) \to ((A \supset B) \wedge (C \supset D))$ | $A \vee C$ |
| $A \to (B \to C)$ | $(A \wedge B) \to C$ | $A \wedge B$ |
| $(A \to B) \to (C \to D)$ | $((A \supset B) \wedge C) \to D$ | $(A \supset B) \wedge C$ |

Table 3: Domains generated by the Cooper reduction principles.

The first row recovers a familiar quasi-conjunction. The conjunction is void only when both component antecedents are false, so a domain-neutral acquisition model preserves $P(A \vee C)$, not necessarily $P(A)$ and $P(C)$ separately. A diagnostic can change the relative probability of $A$-only and $C$-only cases while remaining neutral about whether at least one test occurs. Independently applying an antecedent-preservation rule to the two surface conditionals would instead preserve both component marginals and so imposes a stronger invariance pattern.

Nesting makes the gain more striking. Compare

$$E = A \to (B \to C) \qquad \text{and} \qquad F = A \to (B \supset C). \tag{13}$$

Import–Export gives $D_E = A \wedge B$, whereas $D_F = A$. Yet both contents are false at exactly the $A \wedge B \wedge \neg C$ worlds. They therefore have the same certainty condition and the same material shadow, and trivalent conditionalization updates identically on them. The domain analysis separates them: domain neutrality for $E$ is independence from $A \wedge B$, while domain neutrality for $F$ is independence from $A$.

These requirements can diverge in an elementary model. Suppose $A$ and $B$ are independent fair events and the signal $S$ is the observation that $B$. Write $\kappa_S^E$ when the Bayes factor is computed relative to $E$'s domain. Then $S$ is neutral about $D_F = A$, so $\kappa_S^F = 1$ and $P(A \mid S) = P(A) = 1/2$. But it favors $D_E = A \wedge B$ over its complement:

$$\kappa_S^E = \frac{P(B \mid A \wedge B)}{P(B \mid \neg(A \wedge B))} = 3, \qquad P(A \wedge B \mid S) = \frac{1}{2} > \frac{1}{4}.$$

The example does not decide which content an utterance expresses. It shows that certainty equivalence can conceal a difference that matters to a model of how evidence was acquired. Once a semantics supplies the domains, observed or causally modeled relations between $S$ and those domains can discriminate the corresponding neutrality claims.

The coordinate is canonical only relative to the chosen trivalent semantics. That qualification makes the framework more, not less, informative. On the Cooper semantics, the left-nested $(A \to B) \to C$ reduces to $(A \supset B) \to C$ and has domain $A \supset B$. With the de Finetti conditional it instead reduces to $(A \wedge B) \to C$ and has domain $A \wedge B$.[6] The rival semantics therefore ask whether a signal is independent of different events and identify different marginals as candidates for preservation. The domain Bayes factor does not choose between the semantics; it reveals an epistemic consequence of choosing one.

This locates the contribution relative to earlier work. Adams conditioning and Bradley (2000) supply preservation principles for simple conditionals. Douven (2012) shows that explanatory considerations can move antecedent probabilities. Eva et al. (2020) provide an explicit two-stage picture: interpret the total evidence, including causal and pragmatic structure, and then update on that interpretation. The present account accepts those lessons. It adds three claims. First, a fully compositional trivalent semantics still does not determine an update rule, even after collateral source information is bracketed: truth and non-falsity remain different candidates for the learned event. Second, the semantics supplies a domain coordinate, relative to its own truth tables, for arbitrary compounds rather than merely for simple antecedents. Third, the domain Bayes factor exposes differences that certainty equivalence suppresses, as the pair in (13) demonstrates.

The disagreement with Égré, Rossi, and Sprenger can now be stated exactly. Their trivalent conditionalization is a coherent normative bridge from semantic content to posterior belief. Proposition 6 and the Updating Theorem establish important consequences of adopting it, and the present paper does not show that the bridge is irrational. It establishes two more limited claims. As a content-level norm, trivalent conditionalization selects non-falsity and global odds preservation from several extensions that agree on the bivalent fragment;

[6] See Égré et al. (2026, Proposition 1 and p. 11, n. 6).

that selection needs a substantive defense. As a theory of total acquisition, the rule is incomplete unless the acquisition event has first been shown to contain no further relevant information. A defender may accept both conclusions while retaining the rule for the narrower, certainty-based task.

Trivalence's epistemological payoff is thus structural rather than prescriptive. It shows why learning what happens under a condition and learning whether the condition obtains are distinct dimensions. It does not say that the second dimension must remain fixed.

# 6 Conservative completions and limits

Sometimes an agent or modeler specifies only the two coordinates $s$ and $r'$ and has no further information within the semantic cells. A natural conservative completion preserves prior probability ratios within each cell. Assume first that $t, n, v > 0$; if a semantic cell is empty, its line is simply omitted:

$$P_{\mathsf{C}}^{E;s,r'}(w) = \begin{cases} P(w)\dfrac{sr'}{t}, & w \in E^{\mathsf{T}}, \\ P(w)\dfrac{(1-s)r'}{n}, & w \in E^{\mathsf{F}}, \\ P(w)\dfrac{1-r'}{v}, & w \in E^{\mathsf{V}}. \end{cases} \tag{14}$$

This is the update displayed in Table 1 when $s = 1$ and $r' = r$. It should not be confused with the full update generated by the causal model in Section 4. Equation (14) represents evidence that is coarse-grained exactly to the three semantic cells, or a modeler's decision to add no changes beyond the specified cell masses.

The representation theorem in the Appendix gives this modest claim a familiar form. Once $s$ and $r'$ have fixed all three cell totals, (14) uniquely minimizes every strictly convex $f$-divergence in the class defined there from the prior. This is an application of Jensen's inequality. It does not independently justify either target. Minimum change completes specified cell totals; it does not determine which evidential task fixes them.[7]

Several familiar rules are special cases. If $r' = r$, the completion preserves the domain. For a simple conditional it is Adams conditioning: it changes $P(B \mid A)$ to $s$, preserves $P(A)$, and preserves prior ratios within the three cells. If $E$ is bivalent and $s = 1$, the void line is omitted and the completion reduces to ordinary Bayesian conditionalization. Categorical

[7]The same moral is familiar from the Judy Benjamin literature; see van Fraassen (1981); Douven and Romeijn (2011).

materialization is obtained by setting

$$s = 1, \qquad r' = \frac{t}{t+v}.$$

These rules differ in their substantive targets, not in the mathematical principle used to fill out the posterior.

The framework has ordinary support limitations. Equation (1) is undefined when $Q(D_E) = 0$. Thus the claim that certainty is equivalent to assigning zero probability to $E^{\mathsf{F}}$ requires either the additional condition $Q(D_E) > 0$ or a stipulated value for the all-void case. Égré, Rossi, and Sprenger adopt such a stipulation; it should not be allowed to pass unnoticed. When the prior assigns zero probability to a cell that learning must revive, no update absolutely continuous with respect to the prior will suffice. Popper functions, lexicographic probabilities, imaging, or prior expansion may be needed.[8]

Nor is a true/false/void partition a complete representation of every conditional-learning problem. Causal and imaging approaches encode relations among worlds that a three-cell partition omits. Dynamic accounts can represent conditional probability through hypothetical update rather than the probability of a trivalent content (Ciardelli and Ommundsen, 2024). The present argument is internal to the trivalent project: if conditionals are represented by three-valued contents, their semantics structures but does not settle their dynamics.

# 7 Conclusion

There are two gaps between trivalent semantics and conditional learning. Even at the level of content alone, truth and non-falsity identify different filters. Truth filtering, non-falsity filtering, and domain-stable success all extend ordinary Bayesian conditioning from the bivalent fragment, but they diverge when void worlds are present. At the level of total acquisition, a report or diagnostic can also carry information not exhausted by the content accepted. The first gap is distinctively trivalent; the second is the familiar but inescapable effect of evidential source and context.

The domain Bayes factor makes that freedom explicit. Materialization assigns a particular domain effect; Domain Preservation assigns another. Neither follows from trivalence alone. An explicit reactor model shows how randomized domain entry and a diagnostic of an independent mechanism can justify preservation. It also shows that preservation fixes only a marginal: the diagnostic may carry substantial information within the void cell.

This leaves a genuine role for semantics. A compositional trivalent theory determines,

[8] See Günther (2018) and Günther and Trpin (2023).

relative to its selected truth tables, the applicability domain of complex contents and hence the event against which source effects must be assessed. Conjunction and nesting generate domains that need not be any surface antecedent. Contents with the same falsity set can have different domains, and competing semantics can assign different domains to the same nested sentence. Semantics tells us what the domain question is; an acquisition model tells us how the evidence answers it.

Trivalent conditionalization remains a coherent answer to a well-defined idealized question: what posterior results if learning $E$ is represented by eliminating exactly its false worlds while otherwise changing the prior as little as possible? Its materialization theorem is formally significant. The answer should not be mistaken for a consequence of trivalent truth conditions alone, or for the total update warranted by every episode in which a conditional is acquired. Stating which task the rule addresses is part of a theory of what was learned.

# A Representation theorem

Let $f : [0, \infty) \to \mathbb{R}$ be continuous and strictly convex, with $f(1) = 0$, and define

$$D_f(Q, P) = \sum_{w \in W} P(w) f\left(\frac{Q(w)}{P(w)}\right). \tag{15}$$

Kullback–Leibler divergence is the special case given by the continuous extension of $f(x) = x \log x$ with $f(0) = 0$. No differentiability at zero is assumed.

**Theorem 1** (Representation)**.** *Let $W$ be finite, let $P$ have full support, and suppose the three semantic cells have positive prior probability. For any $s \in [0, 1]$ and $r' \in (0, 1)$, $P_{\mathsf{C}}^{E;s,r'}$ is the unique distribution that minimizes every strictly convex divergence $D_f(Q, P)$ subject to*

$$\pi_Q(E) = s, \tag{16}$$

$$Q(D_E) = r'. \tag{17}$$

*Proof of Theorem 1.* The two constraints fix the total posterior mass of all three cells. Since $Q(D_E) = r'$ and $\pi_Q(E) = s$,

$$Q(E^{\mathsf{T}}) = sr', \qquad Q(E^{\mathsf{F}}) = (1 - s)r', \qquad Q(E^{\mathsf{V}}) = 1 - r'.$$

It remains to determine how each fixed mass is distributed within its cell. Consider any cell $C$ with $P(C) > 0$ and required posterior mass $Q(C) = m$. If $m = 0$, every admissible $Q$

vanishes throughout $C$. Suppose, then, that $m > 0$. Its contribution to the divergence is

$$P(C) \sum_{w \in C} P(w \mid C) f\left(\frac{m}{P(C)} \frac{Q(w \mid C)}{P(w \mid C)}\right).$$

The $P(\cdot \mid C)$-weighted mean of $Q(w \mid C)/P(w \mid C)$ is one. By strict convexity and Jensen's inequality, the displayed expression is uniquely minimized when

$$\frac{Q(w \mid C)}{P(w \mid C)} = 1$$

for every $w \in C$. Equivalently, $Q(w) = P(w)m/P(C)$ throughout $C$. Applying this result to the true, false, and void cells yields exactly (14). At a boundary value of $s$, the zero-mass observation applies to one of the first two cells and the Jensen argument applies to the remaining cells. □

# References


Adams, Ernest W. 1975. *The Logic of Conditionals*. Dordrecht: Reidel.

Belnap, Nuel D. 1970. Conditional assertion and restricted quantification. *Noûs* 4(1): 1–12.

Bradley, Richard. 2000. A preservation condition for conditionals. *Analysis* 60(3): 219–222.

Ciardelli, Ivano, and Adrian Ommundsen. 2024. Probabilities of conditionals: Updating Adams. *Noûs* 58(1): 26–53.

Collins, Peter J., Karolina Krzyżanowska, Stephan Hartmann, Gregory Wheeler, and Ulrike Hahn. 2020. Conditionals and testimony. *Cognitive Psychology* 122: 101329.

Cooper, William S. 1968. The propositional logic of ordinary discourse. *Inquiry* 11(1–4): 295–320.

Douven, Igor. 2012. Learning conditional information. *Mind & Language* 27(3): 239–263.

Douven, Igor, and Jan-Willem Romeijn. 2011. A new resolution of the Judy Benjamin problem. *Mind* 120(479): 637–670.

Edgington, Dorothy. 1995. On conditionals. *Mind* 104(414): 235–329.

Égré, Paul, Lorenzo Rossi, and Jan Sprenger. 2025. Certain and uncertain inference with indicative conditionals. *Australasian Journal of Philosophy* 103(3): 569–596.

Égré, Paul, Lorenzo Rossi, and Jan Sprenger. 2026. Probability for trivalent conditionals. *Mind*, in press.

Eva, Benjamin, Stephan Hartmann, and Soroush Rafiee Rad. 2020. Learning from conditionals. *Mind* 129(514): 461–508.

Günther, Mario. 2018. Learning conditional information by Jeffrey imaging on Stalnaker conditionals. *Journal of Philosophical Logic* 47(5): 851–876.

Günther, Mario, and Borut Trpin. 2023. Bayesians still don't learn from conditionals. *Acta Analytica* 38(3): 439–451.

Lassiter, Daniel. 2020. What we can learn from how trivalent conditionals avoid triviality. *Inquiry* 63(9–10): 1087–1114.

van Fraassen, Bas C. 1981. A problem for relative information minimizers in probability kinematics. *British Journal for the Philosophy of Science* 32(4): 375–379.

Vandenburgh, Jonathan. 2021. Conditional learning through causal models. *Synthese* 199: 2415–2437.