Referee Report

What Do We Learn from a Conditional?

RECOMMENDATION: REJECT

Overall recommendation

The manuscript is exceptionally clear and addresses a timely problem. Its central observation - that certainty equivalence does not entail evidential or update equivalence - is correct and important. Nevertheless, I do not think the paper, in its present form, meets the threshold for Mind, Noûs, or Philosophy and Phenomenological Research. The formal results are sound but follow fairly directly once the substantive constraints have been imposed, while the principal philosophical burden - showing that trivalent semantics privileges Domain Preservation - is not discharged. The paper ultimately establishes a conditional result about one possible information-generating process rather than a distinctively semantic account of learning conditionals.

Summary

The paper considers a trivalent content E, with true, false, and void cells E^T, E^F, and E^V, and defines its domain as $D_E = E^T \cup E^F$. It objects to the trivalent conditionalization rule recently proposed by Égré, Rossi, and Sprenger. For a simple $A \rightarrow B$, that rule has the same effect as conditioning on $\neg A \vee B$; consequently, learning the conditional lowers P(A) whenever both $A\neg B$ and $\neg A$ initially have positive probability.

The alternative begins with an acquisition event S satisfying:

$$P(S \mid D_E) = P(S \mid E^V).$$

Bayes's theorem then entails $P(D_E \mid S) = P(D_E)$. Combining this "Domain Preservation" constraint with a target posterior trivalent probability $\pi_Q(E) = s$, and minimizing any strictly convex f-divergence, yields an update that preserves probability ratios within each truth-value cell. For simple conditionals, this is Adams conditioning. The manuscript then argues for a layered treatment on which conditional content determines revision within the domain while pragmatic, causal, or source-related information may separately affect the probability of entering the domain.

General assessment

There is much to admire here. The exposition is unusually good. The reactor example is effective, the distinction between admissible posteriors and the evidence-supported posterior is well drawn, and the comparison of three treatments of Edgington's coin is illuminating. The manuscript also identifies a genuine weakness in one informal motivation offered for the materialization result. Contents can impose the same posterior certainty condition without providing the same evidence.

My central reservation is that the paper oscillates between two conclusions:

- Conditional learning is underdetermined until an acquisition model is supplied.
- Trivalent semantics supplies a privileged domain-preserving baseline.

The argument convincingly supports the first. It does not yet support the second. Once the manuscript retreats to the first conclusion, however, its positive proposal becomes a familiar combination of Adams conditioning, source independence, and minimum-change reasoning. That is worthwhile, but the paper does not establish sufficiently clearly what new result of top-journal significance remains.

Major comments

1. The crucial bridge from voidness to preservation is missing

Equation (4) entails Domain Preservation immediately. But equation (4) is itself a substantive independence assumption about S and D_E; it does not follow from the conditional's being void outside D_E.

The paper acknowledges this in §6.1, but continues to claim that trivalence "privileges" the neutral baseline and that the distinction between falsity and voidness should constrain learning. I do not yet see the argument for these stronger claims. Semantics identifies an event relative to which neutrality can be formulated. It does not make the relevant signal neutral, nor does it establish a presumption of neutrality.

This matters dialectically. Égré, Rossi, and Sprenger do not formally derive their update rule merely from the requirement of posterior certainty. They define trivalent conditionalization by $p^E(H) = p(E \rightarrow H)$; update equivalence is then a consequence of that rule and their logic. Their informal appeal to certainty equivalence is insufficient, as the manuscript correctly observes, but exposing that insufficiency does not refute the independently proposed update rule. The paper needs either:

- a substantive bridge principle connecting semantic voidness with evidential neutrality; or
- a more modest conclusion according to which semantics underdetermines the update and Domain Preservation describes one important acquisition model.

The second position strikes me as more defensible, but it would require reframing the manuscript's central claim.

2. Source neutrality is prior-relative, and the examples need a generative model

The condition $P(S \mid D_E) = P(S \mid E^V)$ is not simply an intrinsic property of a source. The first term averages the signal's likelihood across E^T and E^F using the agent's prior conditional distribution. In the categorical case, where $P(S \mid E^F) = 0$, neutrality requires:

$$P(S \mid E^T) / P(S \mid E^V) = 1 / P(E^T \mid D_E).$$

Thus a single fixed information channel will generally not be domain-neutral relative to agents with different priors. This is compatible with treating neutrality as agent-relative independence, but it sits uneasily with the manuscript's description of it as an independently assessable fact about the design of an experiment or the causal route of the report. Those process descriptions may justify independence, but the manuscript needs to state the richer probabilistic assumptions under which they do so.

The reactor example makes the problem particularly vivid. The engineer presumably knows some stable mechanical or causal fact about what a restart would produce. In an enriched state space, that fact varies among the ¬R-worlds. The report will therefore normally convey information within E^V, contrary to the update in (6), which preserves all probabilities there pointwise. If the state space contains only the four actual-value combinations of R and O, then it lacks the dispositional or causal variable through which the engineer obtains her information.

The paper recognizes the general possibility of within-void information in its "stable-type" treatment of Edgington's coin. But that concession threatens the motivating examples for the baseline rule. In Table 2, the "domain-neutral content" model assigns relative likelihoods 2, 0, 1, 1. Why should a reliable source be just as likely to issue the heads conditional in the double-tailed/no-flip world as in the double-headed/no-flip world, while being twice as likely to issue it when the coin is actually flipped and lands heads? The numbers satisfy the desired independence, but no natural information-generating mechanism is supplied.

The manuscript should provide an explicit generative model - ideally a causal graph or a set of structural assumptions - in which the report genuinely constitutes information supporting the conditional, is domain-neutral for principled reasons, and carries no finer information within the three semantic cells. Without such a model, the proposed baseline looks reverse-engineered from the desired posterior.

3. Domain Preservation does not determine the advertised update

Source neutrality yields only $Q(D_E) = P(D_E)$. Equation (6) additionally preserves all prior ratios within E^T, E^F, and E^V. The manuscript is explicit that this is a separate minimum-change assumption, but its rhetoric sometimes attributes too much of the result to the trivalent semantics or to source neutrality.

The representation theorem does not provide an independent argument for those constraints. Given the target values of the three cell masses, cellwise proportional rescaling is an immediate application of Jensen's inequality. The theorem is correct, but nearly all the substantive work has already been done by specifying the constraints. The paper should therefore be more restrained in describing the rule as "uniquely characterized by familiar minimum-change principles." It is uniquely selected only after Domain Preservation and the new conditional probability have fixed every cell's total mass.

Relatedly, interpreting the ratios $Q(w)/P(w)$ as relative signal likelihoods does not independently vindicate the update. Any absolutely continuous posterior can be reverse-engineered in this way. What is needed is a prior account of the source that predicts those likelihoods.

4. The novelty relative to the existing literature is insufficiently established

Antecedent preservation under Adams conditioning is familiar, as is the point that it is appropriate only in cases where the conditional information is irrelevant to the antecedent. More importantly, Eva, Hartmann, and Rafiee Rad already motivate antecedent preservation through the independence of the reason for a conditional report from the antecedent, and explicitly distinguish interpreting evidence from updating on an interpretation. Their discussion of the "disabler" variable in the sundowners case is especially close to the present manuscript's reactor argument. Douven likewise discusses the circumstances in which Adams conditioning is appropriate.

The manuscript cites these works but does not adequately distinguish its contribution from them. The principal additions appear to be:

- using a trivalent semantics to identify the relevant domain for arbitrary compounds;
- contrasting this explicitly with the materialization theorem in Égré, Rossi, and Sprenger; and
- giving an all-f-divergence representation result.

These are genuine additions, but the first depends on accepting the particular trivalent semantics, while the third is mathematically straightforward. For publication at one of the journals in question, the paper needs a much sharper account of what earlier source-sensitive and Adams-conditioning treatments could not explain, and what substantive consequence the compositional extension has beyond allowing the same formula to be applied to more syntactically complex contents.

5. The status of materialization as an acquisition model needs qualification

The manuscript says that materialization "builds" a domain-negative answer into the learning event. Mathematically, conditioning on $E^T \cup E^V$ can indeed be represented by a signal whose likelihood is one on that set and zero on E^F. But this does not show that proponents of materialization are committed to a psychological or causal source that literally has those world-relative reporting propensities. It may instead be a rule for updating on semantic content, analogous to ordinary conditioning.

The likelihood reconstruction usefully displays the rule's probabilistic consequences. It does not by itself establish that the rule presupposes a mistaken empirical model of speakers or evidence sources. The manuscript should distinguish:

- a formal likelihood representation of an update;
- a claim about the actual causal production of testimony; and
- a normative principle governing "pure" content-level learning.

At present these are sometimes run together.

Minor and technical points

- The theorem calls g(x) = x log x differentiable on the nonnegative reals, although its ordinary derivative is not finite at 0. The statement should use the standard continuous extension and formulate the regularity condition appropriately.
- The terminology should be made consistent: the text alternates between f-divergence and g-divergence.
- The paper should distinguish its "Domain Preservation" more explicitly from earlier preservation conditions in the conditional literature.
- The undefined case $Q(D_E) = 0$ deserves mention when certainty is identified with $Q(E^F) = 0$. The minimum-divergence solution avoids it under the manuscript's assumptions, but the equivalence is not unrestricted.
- The claims that materialization treats void worlds as "positive survivors" and that it "erases" the semantic distinction are rhetorically stronger than what has been shown. Dynamic equivalence does not itself imply semantic identification.

Recommendation

This is a polished and intelligent manuscript containing a real criticism of the inference from certainty equivalence to update equivalence. But the positive proposal currently derives its conclusion from a neutrality assumption that is largely equivalent to the desired invariance, and it does not show that trivalent semantics gives that assumption privileged status. Moreover, the central source-independence motivation is substantially anticipated in the literature, while the formal representation result is comparatively modest.

Repair would require a significant reconception: a clearer thesis about underdetermination versus a privileged baseline, an explicit generative account of domain-neutral conditional information, and a much more detailed demonstration of novelty. For that reason, rather than recommending revision and resubmission at this journal, I recommend rejection.

Works discussed

Douven, Igor. 2012. "Learning Conditional Information." Mind & Language 27(3): 239-263.
Égré, Paul, Lorenzo Rossi, and Jan Sprenger. 2026. "Probability for Trivalent Conditionals." Mind, forthcoming.
Eva, Benjamin, Stephan Hartmann, and Soroush Rafiee Rad. 2020. "Learning from Conditionals." Mind 129(514): 461-508.