Second-Round Referee Report

Manuscript: What Do We Learn from a Conditional?

Recommendation: Revise and Resubmit

This is a substantially stronger paper. The revision directly addresses nearly every concern in my earlier report, and the principal grounds for my previous recommendation of rejection no longer apply. I would now recommend an R&R, with a serious prospect of eventual acceptance. I would not yet recommend conditional acceptance because one remaining issue concerns the paper's central dialectical contribution rather than merely its presentation.

What the revision successfully accomplishes

The most important improvement is the change in the central thesis. The paper no longer claims that trivalent semantics privileges Domain Preservation. It instead argues for "structured underdetermination": semantics supplies a true/false/void partition, while an acquisition model determines how the evidence bears on that partition. This is both more defensible and more precisely stated.

The revision also handles the earlier objections impressively:

- It now correctly represents trivalent conditionalization as an independently proposed formal rule, rather than suggesting that Égré, Rossi, and Sprenger derive it merely from certainty equivalence.
- It explicitly recognizes that source neutrality is prior-relative. Equation (9) states the issue cleanly, and the paper no longer treats neutrality as an intrinsic property of a source.
- Section 4 supplies the needed generative model. The variables R, M, O, and S explain both why the diagnostic is independent of restart and how it supports the conditional. The model also shows that domain preservation does not entail within-void invariance.
- The discussion of Eva, Hartmann, and Rafiee Rad is now appropriately candid. The manuscript no longer presents source independence as new, and it distinguishes its proposed contribution from the disabler analysis.
- The representation theorem has been given the right status: it is a conservative completion conditional on specified cell totals, not an independent justification of those totals. The technical issue concerning differentiability of x log x at zero has also been repaired.
- The materialization likelihood model is now explicitly presented as a formal representation of the rule's effects, not as a psychological commitment of its proponents.

These are not cosmetic revisions. They amount to a genuine reconception of the paper, and the resulting argument is clearer, more accurate, and more philosophically credible.

The principal remaining issue

The paper must sharpen the distinction between two targets:

- a norm governing the idealized update on the trivalent content E alone; and
- the total update warranted by a concrete event through which an agent comes to accept E.

The causal model in §4 establishes that a diagnostic S can make $R \rightarrow O$ certain while preserving $P(R)$ and changing beliefs within the void cell. But S conveys information about the mechanism M, which is not exhausted by the trivalent content of $R \rightarrow O$. In particular, it distinguishes no-restart worlds that the conditional itself classifies uniformly as void.

A defender of trivalent conditionalization can therefore accept the causal model while maintaining:

> Updating on the total diagnostic evidence S is not the same operation as updating on the conditional content R → O alone. Trivalent conditionalization governs only the latter.

The revised paper now acknowledges that such a content-level norm would be coherent. But this creates a question about the force of its criticism. Is the conclusion that trivalent conditionalization is incorrect, or merely that it is an additional norm not forced by the semantics? If the latter, Égré, Rossi, and Sprenger may simply agree: their conditionalization rule is precisely the additional normative bridge they propose.

The manuscript needs to say more clearly what is being established against them. Possible positions include:

- Trivalent conditionalization is insufficiently motivated, even as an idealized content-level norm.
- It is legitimate as a content-level norm but should not be described as a general theory of acquiring conditional information.
- The notion of learning "exactly the trivalent content and nothing more" is itself ill-defined unless one has already selected a bridge rule.
- Several rival content-level bridges are equally natural by standards the target theory accepts.

Any of these would be interesting. At present, the paper moves among them. Its causal model powerfully establishes the second claim, but by itself it does not establish the first.

Is the bridge problem distinctively trivalent?

A related issue is that source-sensitive underdetermination is not unique to conditionals. A report of an ordinary bivalent proposition can also convey information about the source, the circumstances of assertion, and propositions not entailed by its literal content. If this generic fact were sufficient to show that content does not determine an update, ordinary Bayesian conditionalization would face the same objection.

The manuscript contains the beginnings of the correct answer in §2: for a bivalent proposition, truth and non-falsity coincide, so the familiar idealization of learning the proposition identifies a single set of worlds. In the trivalent case, one can condition on E^T, on $E^T \cup E^V$, or on an independently modeled signal. The semantics itself does not identify which of truth or non-falsity plays the role occupied by the truth set in ordinary conditioning.

This should be developed further. In particular, the paper should explain why the trivalent bridge problem is not merely the familiar observation that testimonial evidence can contain collateral information. The special problem appears to be that even after collateral source information is idealized away, the semantic representation supplies two initially plausible candidates for the learned event. That is a stronger and more distinctive claim than the generic source-sensitivity thesis, and it should be made central.

The positive payoff from compositionality needs further development

Given the revised, more modest thesis, the strongest claim to novelty now lies in §5: a compositional trivalent semantics identifies domains for complex contents, and these domains determine the relevant questions about evidential neutrality.

This is promising, but the present discussion is too brief to carry the paper's top-journal case. The conjunction example $(A \rightarrow B) \wedge (C \rightarrow D)$ shows that the Cooper domain is $A \vee C$, so domain neutrality preserves $P(A \vee C)$, not necessarily $P(A)$ and $P(C)$ individually. That is a good example. But it presently occupies only a few paragraphs and does not yet demonstrate the broader significance of the proposal.

I recommend adding at least one case involving nesting, where the compositionally determined domain is less transparent. It would also help to show one of the following:

- that independently Adams-conditioning on the component conditionals produces a different invariance pattern;
- that competing trivalent semantics assign different domains and thereby generate different empirically or normatively assessable neutrality questions;
- that two contents with the same certainty condition but different trivalent domains impose demonstrably different requirements on an acquisition model; or
- that the domain-Bayes-factor framework yields some non-obvious constraint on sequential or compound conditional learning.

At present, the domain Bayes factor is mathematically elementary - it is the ordinary Bayes factor for the event D_E. Its novelty comes entirely from the semantic identification of D_E, especially for complex formulas. The paper should therefore develop that application rather than relying mainly on the simple-conditional case, where the domain is just the familiar antecedent.

Minor comments

- The author footnote inviting discussion "via this ChatGPT conversation" should be removed from a journal submission.
- The final sentence - "a semantic theory does not become a theory of learning merely by adding a rule called conditionalization" - is unnecessarily polemical. Égré, Rossi, and Sprenger do offer motivations for their rule, even if those motivations are not decisive.
- "Canonical domain" should always be understood as canonical relative to the selected trivalent semantics. The paper generally observes this qualification, but a few formulations could be tightened.
- Theorem 1 excludes the limiting domain Bayes factors 0 and ∞. A brief statement of the corresponding boundary cases would complete the presentation.
- Because Theorem 1 is explicitly only a realization result, the author might consider presenting it as a proposition or observation. Nothing substantive turns on this, however.
- The paper is now admirably concise. I would not restore the longer catalogue of applications from the earlier version; additional space should be devoted specifically to the dialectical target and the complex-content payoff.

Overall verdict

The revision is successful. It has transformed a paper whose positive conclusion was insufficiently supported into a careful argument that trivalent semantics structures conditional learning without uniquely determining it. The causal model, the explicit treatment of prior-relativity, and the more honest positioning relative to Eva et al. are all substantial improvements.

The remaining question is whether the paper merely observes that a proposed update rule requires an additional normative assumption, or whether it identifies a problem distinctive to trivalent learning and provides a significant compositional framework for understanding it. I think the materials for the stronger contribution are present, especially in the distinction between truth and non-falsity and in the domains of complex contents. They need to be made more central and developed further.

Accordingly, I would upgrade my recommendation from Reject to Revise and Resubmit.