

## New Experimental Evidence against the Similarity Approach to Conditionals

**Summary.** The similarity approach to conditionals (Stalnaker 1968, Lewis 1973) predicts Reciprocity to be valid: whenever  $A > B$ ,  $B > A$  and  $A > C$  are true,  $B > C$  is true (where  $A > B$  denotes *if A would B*). We ran an experiment to test the validity of this rule. Strikingly, half of our participants judged the rule invalid, i.e. judged in at least one scenario that it does not preserve truth. Our data also challenge Kratzer’s (2012) and Fine’s (2012) semantics of conditionals, but we show that McHugh’s (2022) aboutness approach can account for our data.

**Background.** Ever since Stalnaker (1968) and Lewis (1973), semanticists have been attracted to the idea that a *would*-conditional “if A, would C” (henceforth denoted  $A > C$ ) is true just in case C is true at the most similar world(s) to the actual world where A is true. This is arguably the dominant approach to the semantics of *would*-conditionals today, and was influentially extended to modality in general via Kratzer’s (1981) ordering source. The similarity approach validates the following rule of inference, known as Reciprocity (Egré and Rott 2021; Nute 1980 calls it ‘CSO’ and Starr 2022 calls it ‘Substitution of Subjunctive Equivalents’).

*Reciprocity.* Whenever  $A > B$ ,  $B > A$  and  $A > C$  are true,  $B > C$  is true.

This rule is part of the logic of similarity approach to conditionals: no matter how one intuitively interprets the similarity order, Lewis’s and Stalnaker’s semantics predict it to be valid.

**Experiment.** We tested the judgements of 80 native English speakers, recruited via Prolific, in three sentence-picture verification tasks, displayed in a random order. To illustrate, in one of the tasks we presented participants with Figure 1. The light is on just in case switch A is in the middle and switch B is up or in the middle. Part of the image is shaded. Currently A is in the middle and B is down, so the light is off. After a training phase to ensure participants understood the scenario, we presented the following sentences, in addition to six fillers, in a random order. T is a true control, F a false control, and P1–C is an instance of Reciprocity.

- T If both switches were in the middle, the light would be on.
- P1 If switch B were in the shaded area, both switches would be in the shaded area.  $B > both$
- P2 If both switches were in the shaded area, switch B would be in the shaded area.  $both > B$
- P3 If switch B were in the shaded area, the light would be on.  $B > on$
- C If both switches were in the shaded area, the light would be on.  $both > on$
- F If both switches were outside the shaded area, the light would be on.

Following Romoli et al. (2022), for each sentence we asked the participants whether it is true, false, or indeterminate. If they answered indeterminate we asked them whether they strongly feel that there is no correct answer or just do not know, and excluded the latter responses from the analysis. We added an indeterminate option because some theories (e.g. Stalnaker’s 1980 supervaluationism) allow conditionals to be neither true nor false. Participants understood our scenarios well, with a mean accuracy of 89% on the filler items. We excluded from the statistical analysis two participants whose error rates on the fillers were above 30%.

**Results.** Table 1 and Figure 2 present the participants’ responses from all three scenarios. Using logistic regression we found no significant difference in rates of true judgements between the true control and any of the premises P1–P3, but each premise was significantly different from the conclusion C ( $\beta \approx 26\%$ ,  $p < 0.0001$ ). Analysing each scenario individually revealed effects of similar size and significance. On average, 25% of the time when a participant accepted all of the premises P1–P3 they rejected the conclusion C ( $p < 0.0001$ ).

**Discussion.** Our results offer a novel and striking challenge to the similarity approach to conditionals. It is impossible to construct a model of Lewis or Stanlaker’s semantics in which premises P1–P3 are true but the conclusion C is not. However, a significant number of participants reasoned in this way; indeed, half of the participants judged in at least one scenario that the premises are true but conclusion is not (42/78). This remains (39/78) when we restrict to those who correctly answered the true and false controls, indicating that they understood the scenario and reasoned correctly with conditionals of equal complexity to the test items.

On the basis of our results, we propose abandoning the similarity approach. The question is what alternative framework to adopt in its place. Kratzer (1989, 2012:133) proposes a semantics of *would*-conditionals using situation semantics. As the reader may verify, on the most plausible choice of base set—in the switches scenario,  $\{A \text{ in the middle, } B \text{ down, light is on iff } A \text{ is middle and } B \text{ is the middle or up}\}$ —Kratzer’s semantics predicts the conclusion to be true: if both switches were in the shaded area the light would be on. Kratzer therefore cannot account for the significant number of participants who did not judge this sentence true.

Fine (2012) proposes a semantics of *would*-conditionals using truthmaker semantics:  $A > C$  is true at a world  $w$  just in case for every exact verifier  $t$  of  $A$  and possible outcome  $u$  of  $t$  at  $w$ ,  $u$  contains an exact verifier of  $C$ . Fine’s semantics satisfies a principle he calls *Universal Realisability of the Antecedent*: a conditional is true only if it is true for any way in which its antecedent might be true (Fine 2012:236). The state of both switches being up exactly verifies that both switches are in the shaded area, in which case the light would be off. Thus Fine predicts the conclusion  $C$ , *If both switches were in the shaded area the light would be on*, to be false. This does not account for the majority of participants who judged it true.

More recently, McHugh (2022, 2023) has proposed a semantics of conditionals based on a notion of aboutness. The thought is that when we interpret a conditional, we allow the part of the world its antecedent is about to vary. The guiding idea of aboutness is that a sentence is about the part of the world that exactly determines its truth value (McHugh 2023:108). This allows for variation in what a sentence is about; for example, McHugh (2023:123–24) offers two views on aboutness. On the truthmaker view, a sentence is about its exact verifiers and falsifiers. The state of switch A being in the middle does not exactly verify or falsify *both switches are in the shaded area*, so this view predicts that we fix switch A (but allow B to vary) when we interpret  $C$ , *If both switches were in the shaded area, the light would be on*: this view predicts  $C$  to be true. On the subject matter view, a sentence is about the exact verifiers and falsifiers of its atomic sentences. Since *switch A is in the shaded area* is an atomic sentence of *both switches are in the shaded area*, this view predicts that we vary switch A when we interpret  $C$ , which is therefore predicted to be false. Assuming that our concept of aboutness is vague, allowing for both of these views, correctly predicts  $C$ ’s intermediate status.

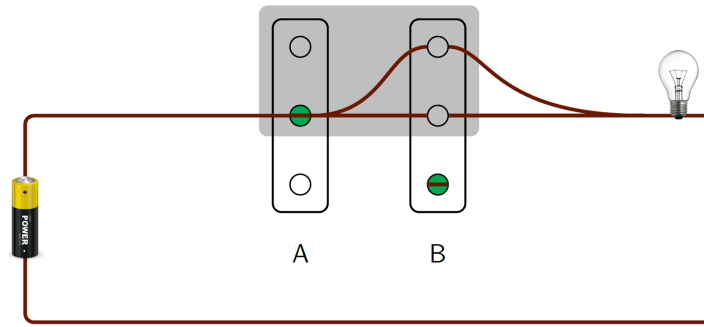


Figure 1: The switches scenario, one of the three scenarios tested in our experiment.

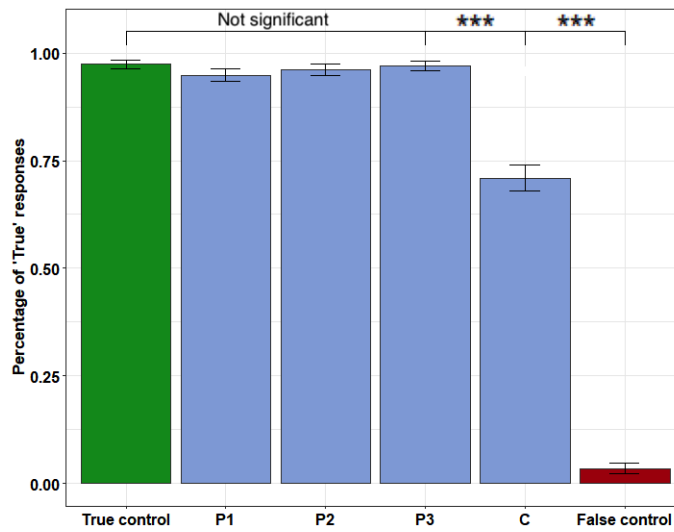


Figure 2: Percentage of 'True' responses. Error bars denote Standard Errors.

Sentence	True	Indeterminate	False	Not sure
True control	232	0	8	0
Premise 1	225	0	11	4
Premise 2	227	2	8	3
Premise 3	225	2	12	1
Conclusion	163	26	43	8
False control	9	0	229	2

Table 1: Responses from all three scenarios.

**References.** Egré, P. & Rott H. (2021). The logic of conditionals, *SEP*. • Fine, K. (2012) Counterfactuals without possible worlds. • Kratzer, A. (1981) The notional category of modality. (1989) An investigation of the lumps of thought. (2012) *Modals and Conditionals*. • Lewis, D. (1973) *Counterfactuals*. • McHugh, D. (2022) Aboutness and modality. *Proceedings of the 23rd Amsterdam Colloquium*. (2023) *Causation and Modality*. • Nute, D. (1980) *Topics in conditional logic*. • Romoli, J. Santorio, P. & Wittenberg, E. (2022) Alternatives in counterfactuals: what is right and what is not. • Stalnaker, R. (1968) A theory of conditionals. (1980) A defense of conditional excluded middle. • Starr, W. (2022) Counterfactuals, *SEP*.