

Between a conditional's antecedent and its consequent: discourse coherence vs.
probabilistic relevance

Karolina Krzyżanowska
Ludwig-Maximilians-Universität München

Peter J. Collins and Ulrike Hahn
Birkbeck, University of London

Forthcoming in Cognition.

Author Note

Correspondence concerning this article should be addressed to Karolina Krzyżanowska, Center for Advanced Studies, Ludwig-Maximilians-Universität München, Geschwister-Scholl-Platz 1, 80539 Munich, Germany (k.krzyzanowska@lmu.de).

We are grateful to David Over and two anonymous reviewers for valuable feedback.

This research was supported by the Alexander von Humboldt Foundation's Anneliese Maier Research Award to Ulrike Hahn. Additionally, Peter Collins was supported by a studentship from the ESRC's Bloomsbury Doctoral Training Centre, award number ES/J500021/1. Karolina Krzyżanowska was supported by Alexander von Humboldt Foundation and the Senior Researcher in Residence Programme of the Center for Advanced Study, LMU Munich.

Abstract

Reasoning with conditionals is central to everyday life, yet there is long-standing disagreement about their meaning. One example is the puzzle of so-called missing-link conditionals such as “if raccoons have no wings, they cannot breathe under water.” Their oddity may be taken to show that conditionals require a connection between antecedent (“raccoons have no wings”) and consequent (“they cannot breathe under water”), yet most accounts of conditionals attribute the oddity to natural-language pragmatics. We present an experimental study disentangling the pragmatic requirement of discourse coherence from a stronger notion of connection: probabilistic relevance. Results indicate that mere discourse coherence is not enough to make conditionals assertable.

Keywords: indicative conditionals; probabilistic relevance; discourse coherence; assertability

Between a conditional's antecedent and its consequent: discourse coherence vs.
probabilistic relevance

Introduction

Indicative conditionals, that is, sentences of the form “If p , then q ,” are everywhere. We utter them in everyday conversations, in political or legal debates, or in scientific discourse, for instance:

- (1) a. If I don't leave in five minutes, I will be late for the meeting.
- b. If we keep emitting greenhouse gases at our current pace, the oceans will rise and many cities will be flooded.

It is unsurprising then that there has been huge interest in conditionals from psychology (Evans & Over, 2004; Johnson-Laird & Byrne, 2002), linguistics (Declerck & Reed, 2001; Elder & Jaszczolt, 2016; Kratzer, 1986), and philosophy (Adams, 1975; Bennett, 2003; Douven, 2016; Edgington, 1995). Despite long-standing interest, many basic questions about how to interpret conditionals remain unresolved. One puzzle is the oddity of so-called missing-link conditionals, for instance:

- (2) a. If Russia never joined the European Union, Bogota is the capital of Colombia.
- b. If raccoons have no wings, they cannot breathe under water.

A natural response is to claim that an indicative conditional conveys some kind of connection between antecedent, p , and consequent, q . In its strong form, this view takes such a connection to be part of the semantics of conditionals. However, the most common approach has been to attribute their “weirdness” to pragmatics— non-literal, speaker-intended meaning that makes “Can you pass the salt?” a request for salt, not a question about an ability. As Grice (1989) observed, pragmatic considerations may make a true sentence unassertable in context. For instance, the sentence:

- (3) Some horses are mammals.

is a strange thing to say, because it suggests, or *implicates*, that not all horses are mammals. Someone asserting (3), however, would not be saying anything false (from a purely semantic perspective).

A whole host of otherwise divergent theories which do not posit an intrinsic connection between antecedent and consequent have appealed to pragmatics to explain the oddity of missing-link conditionals, alongside other deviations between theory and participant's response. In particular, most theories of conditionals render valid an inferential transition from two true statements (such as "raccoons have no wings" and "raccoons cannot breathe under water") to not only their conjunction ("raccoons have no wings and cannot breathe under water") but also the conditional ("if raccoons have no wings, they cannot breathe under water"). That is, whenever two statements are true, the conditional combining them is true also. Such inferences, however, are obviously *not* valid, if conditionals require a connection between the clauses. Known technically as "centering"¹ this inference is entailed by most accounts of indicative conditionals in the philosophical and psychological literature, such as the material interpretation (Grice, 1989; Jackson, 1987), Stalnaker's interpretation (Stalnaker, 1968, 1975), Mental Models Theory (Johnson-Laird & Byrne, 2002), and the suppositional theory (Adams, 1975; Edgington, 1995; Evans & Over, 2004).

The proponents of these accounts do not deny that sentences such as (2a) and (2b) are odd, or that people might hesitate to endorse "centering" for them (see, e.g., Cruz et al., 2016; Skovgaard-Olsen et al., 2016a), but claim that the oddity of missing-link conditionals is due to pragmatics (see, e.g., Johnson-Laird and Byrne 2002, p. 651 or Over, Hadjichristidis, Evans, Handley, and Sloman 2007, p. 92). Few authors, however,

¹The inference from "*p* and *q*" to "if *p* then *q*" is referred to as "one-premise centering" or "conjunctive sufficiency," and it can be distinguished from (psychologically not equivalent) "two-premise centering," which takes *p* as one premise and *q* as another premise. Whether or not lay people endorse either form of centering is a matter of controversy. For example, Cruz, Baratgin, Oaksford, and Over (2015), Cruz, Over, Oaksford, and Baratgin (2016) and Politzer and Baratgin (2015) found evidence in favour of one-premise centering. By contrast, a recent paper by Skovgaard-Olsen, Singmann, and Klauer (2016a) finds evidence against.

hint at any specific pragmatic mechanism that would explain this phenomenon. An exception seems to be a recent paper by Cruz et al. (2016, p. 1108) who claim that:

... what matters for the intuition that there is something odd in conditionals like “If Hillary Clinton runs for president in 2016, the earth weighs more than 2 kilograms,” is the absence of a common topic of discourse between p and q , rather than the absence of a specific connection.

This suggests that the connection between antecedents and consequents may be nothing more than what we normally expect of any two statements constituting a coherent piece of discourse. A direct way to test the pragmatic hypothesis is by creating scenarios that pit discourse coherence against a stronger, probabilistic connection, and examine the impact of these two factors on peoples’ judgements. The present paper reports a study to this effect.

Discourse coherence and probabilistic relevance

It is not without merit to argue that antecedents and consequents of indicative conditionals need to be connected, because speakers expect *any* consecutive elements of discourse to be connected in some way. Discourse, after all, is not a random collection of sentences, but has implicit organization. Compare the following sentences from (Hobbs, 1979):

- (4) a. James took a train from Paris to Istanbul. He has family there.
- b. James took a train from Paris to Istanbul. He likes spinach.

Both examples consist of two pieces of information, which are not strange themselves and which we can easily imagine to be true about a particular James. However, while (4a) is a perfectly natural thing to say, (4b) raises eyebrows. The reason seems rather straightforward: spinach does not have much to do with a train trip to Istanbul; the two sentences seem disconnected.

Consecutive, self-standing elements of discourse, p and q , can be related in various ways. For instance, q may elaborate on, or explain, p . In most cases, unless signalled

otherwise, p and q are expected to at least be on the same topic. In (4b), this expectation is violated, unless we can conjure up a link (e.g., the speaker wants to suggest that eating a lot of spinach is correlated with a fear of flying, or that Istanbul is culinary heaven for spinach lovers; cf. Asher and Lascarides 2003; Kehler 2002; Stojnić 2016).

Where p and q are the antecedent and consequent of a conditional, however, we arguably expect a stronger connection than just any discourse coherence relation. One candidate is probabilistic relevance, conventionally operationalized as a Δp rule:²

$$\Delta p = \Pr(q|p) - \Pr(q|\neg p)$$

Whenever $\Delta p > 0$, we say that p is positively relevant for q . $\Delta p = 0$ indicates irrelevance, while $\Delta p < 0$ indicates negative relevance (see, e.g., Oberauer, Weidenfeld, & Fischer, 2007; Over et al., 2007; Skovgaard-Olsen, Singmann, & Klauer, 2016b). In other words, probabilistic relevance means that p either raises or lowers the probability of q .

Though we contrast probabilistic relevance with discourse coherence in this paper it is important to understand that probabilistic relevance itself *gives rise to discourse coherence*. While it is not uncommon for two coherent elements of discourse to be probabilistically independent, whenever p actually raises the probability of q , p and q occurring together in a conversation should constitute a coherent piece of discourse, as the following examples indicate:

- (5) a. John forgot to take his umbrella. Bob never carries an umbrella.
- b. John forgot to take his umbrella. He will get wet.

The probability of “Bob never carries an umbrella” would not be expected to be higher under the supposition that John forgot to take his umbrella, but the two sentences together nevertheless appear coherent—they can constitute two consecutive elements of

²One could also think of the connection in terms of evidential support (Douven, 2008) or inferential relations (Krzyżanowska, 2015).

discourse, as do sentences in the example (5b). In this case, however, the conditional probability of John’s getting wet given that he forgot to take his umbrella is higher than the probability of “John will get wet” on its own. In other words, while it is difficult to envision probabilistic relevance without discourse coherence, it is readily possible to have discourse coherence without probabilistic relevance. This possibility forms the basis of our experimental test.

The present experiment

We investigated whether people expect a stronger connection between the antecedent and consequent of an indicative conditional than between other consecutive elements of discourse. More specifically, we aimed to disentangle the effect of probabilistic relevance from (mere) discourse coherence. We compared how people evaluate conditionals with how they evaluate the consequents of those conditionals in conversational contexts in which the antecedents have already been asserted (see figures 1 and 2).

Our test factorially combined probabilistic relevance (positive relevance, irrelevance) and discourse coherence (same topic, different topics). However, given that probabilistic relevance, whenever it is salient to interlocutors, establishes discourse coherence, the combination positive relevance/no discourse coherence is practically not possible. This left the following conditions for comparison:

1. positive relevance & the same topic (PR-ST)
2. irrelevance & same topic (IR-ST)
3. irrelevance & different topic (IR-DT)

These within-participant conditions were combined with a between-participant manipulation of type of discourse (conditional/conversational exchange): the two clauses were presented either in a conditional or as two consecutive turns in a conversational exchange (detailed examples follow below).

We asked participants two questions: one probed how natural the clauses were to assert (“assertability”) and one asked whether it made sense for a speaker to say them (“sensibleness”). We expected neither the exchange nor the conditional to be

assertable (sensible) where the two clauses were on different topics. Conversely, we expected equal assertability (sensibleness) when the two clauses were both on the same topic and connected via probabilistic relevance. Crucially, however, we expected only the exchange, and not the conditional, to be assertable where there was only a common topic, but not probabilistic relevance.

Methods

Participants

101 individuals completed the online survey on the MTurk platform (<https://www.mturk.com/>). 100 participants indicated English as their first (native) language. We removed one (male) participant whose first language was not English. Of the 100 participants who remained, 42 were female, 56 male, 1 agender and 1 bigender. Their mean age was 33.77 (range 18-66). All participants received a small remuneration.

Materials and procedure

Each participant was randomly assigned to one of the two groups: the conditionals or the conversational exchanges condition. Within these conditions, each participant received 18 blocks presented in a random order one at a time. Each block consisted of a conversational context followed by two questions concerning conditionals or exchanges exemplifying different combinations of probabilistic relevance and same/different discourse topic:

PR-ST: *“If Sophie likes the Alps, then she will enjoy hiking in the Pyrenees.”*

IR-ST: *“If Sophie likes the Alps, then mountaineering can be dangerous.”*

IR-DT: *“If Sophie likes the Alps, then more and more people in Western Europe care about animal welfare.”*

The first question concerned assertability and was immediately followed by the question about sensibleness. The dependent variables were marked on 7-point

Patrick plans to take his girlfriend, Sophie, for short holidays. He is discussing different ideas with Matt. They have been to the Alps plenty of times, so now Patrick considers a hiking trip in the Pyrenees.

	Definitely no						Definitely yes
In this context, would it be natural for Matt to assert "If Sophie likes the Alps, then she will enjoy hiking in the Pyrenees"?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Or, relatedly:							
In this context, would it make sense for Matt to say "If Sophie likes the Alps, then she will enjoy hiking in the Pyrenees"?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 1. An example of a PR-ST item in the conditionals condition.

Patrick plans to take his girlfriend, Sophie, for short holidays. He is discussing different ideas with Matt. They have been to the Alps plenty of times, so now Patrick considers a hiking trip in the Pyrenees.

Patrick tells Matt: "Sophie likes the Alps."

	Definitely no						Definitely yes
In this context, would it be natural for Matt to assert "She will enjoy hiking in the Pyrenees."	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Or, relatedly:							
In this context, would it make sense for Matt to say "She will enjoy hiking in the Pyrenees."	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 2. An example of a PR-ST item in the conversational exchanges condition.

Likert-style scale. Figures 1 and 2 show example items from conditionals and conversational exchanges conditions, respectively.

Results

Figure 3 and 4, which show histograms of the individual ratings in each condition, already clearly reveal marked differences in the response distributions across conditions. Supplemental statistical analyses were performed on the mean rating (averaged across items) for each participant. Because of the necessarily missing cell (positive relevance PR & different topic DT), we split the analyses by same/different topic.

For the "different topic" conditions, we ran independent-samples comparisons by type of discourse (i.e. conditionals versus conversational exchange). For the "same topic" conditions, we ran 2×2 mixed analyses (probabilistic relevance \times type of

discourse). Throughout, the residuals showed moderate departures from the normal distribution. Moreover, there were unequal numbers of participants in the between-participants conditions. In such circumstances, standard analyses can perform poorly (Wilcox, 2016). We followed Wilcox’s (2016) guidelines for such situations, running both standard and robust analyses. Robust analyses used 20% trimmed means, and were performed in R (R Core Team, 2016) using functions from Wilcox (2016) and the WRS2 package (Mair, Schoenbrodt, & Wilcox, 2016). To correct for multiple comparisons, we chose significance levels of $p = .025$ for the mixed analyses and for the different-topics analyses, and $p = .004$ for the simple-effects analyses.

Table 1 below shows the descriptive data of the “assertability” question.

		PR-ST	IR-ST	IR-DT
Conditionals	Mean	5.91	2.44	1.37
	SD	0.82	1.07	0.98
Exchanges	Mean	5.77	4.92	1.48
	SD	0.87	1.09	1.01

Table 1

Descriptive statistics for the “assertability” question.

We begin with the analysis of the “different topic” (DT) conditions.

Different Topics. As Figure 3 shows, the absence of a common discourse topic renders both the conversational exchange and the conditionals equally unassertable. In fact, there were no significant differences between conditionals and conversational exchanges either with conventional independent-samples t-tests, $t(99) = -.55$, $p = .58$, $d = .11$, or with the robust analysis with 20% trimmed means $t(45.81) = .81$, $p = .42$, $\xi = .13^3$ (robust independent-samples t-test using “yuen” function).

³ ξ is a robust explanatory measure of effect size. 0.10, 0.30, and 0.50 correspond to small, medium, and large effect sizes. For discussion, see Mair and Wilcox (2016).

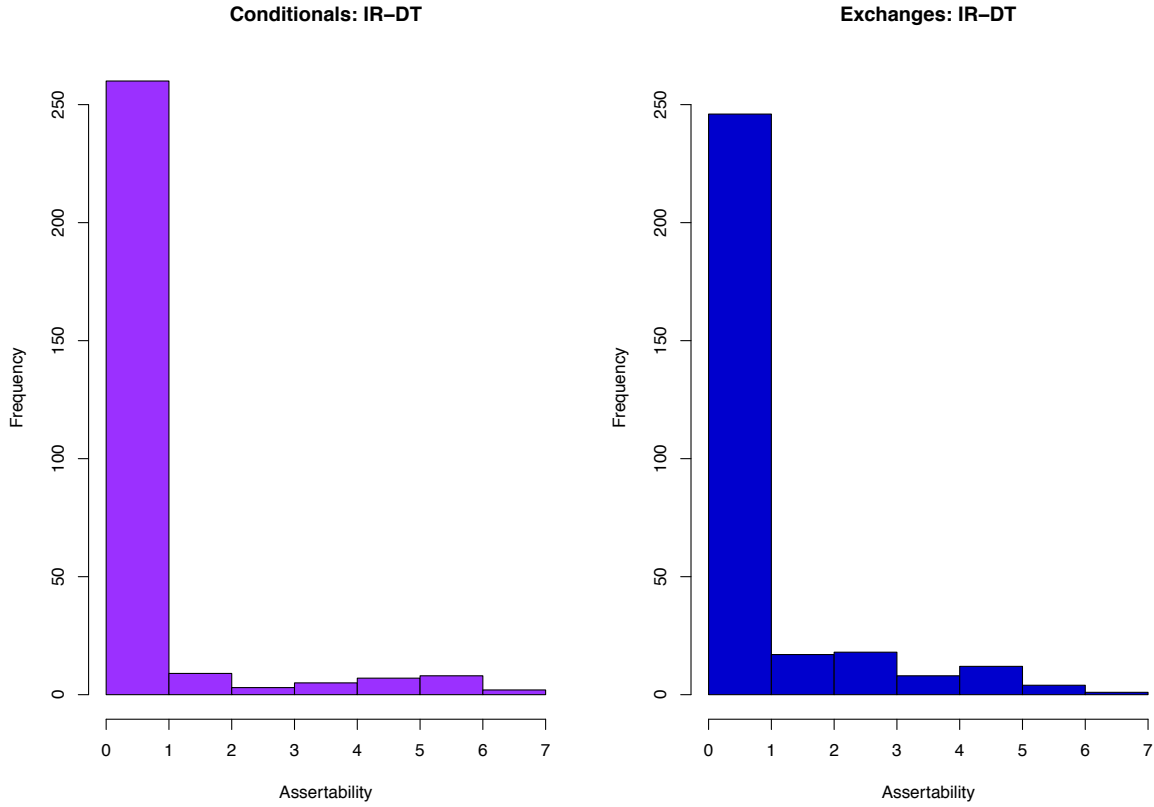


Figure 3. Results for the “assertability” question, DT condition.

Same Topic. Figure 4 shows the distribution of responses in the “same topic” (ST) conditions. Visual inspection suggests an interaction : whereas probabilistically irrelevant clauses on the same topic receive “definitely yes” as the modal response in the discourse condition, the modal response for conditionals is “definitely no”. Statistical analysis confirms this impression.

A 2×2 mixed ANOVA yielded a significant main effect of probabilistic relevance, $F(1, 99) = 337.16, p < .001$, partial $\eta^2 = .77$. The main effect of task was also significant, $F(1, 99) = 58.49, p < .001$, partial $\eta^2 = .37$, as was the interaction between probabilistic relevance and task, $F(1, 99) = 124.08, p < .001$, partial $\eta^2 = .56$.

To explore the interaction further, we ran simple-effects analyses. Firstly, we compared the effect of probabilistic relevance for each type of discourse; these are the most theoretically interesting simple effects. For conditionals, relevant items were given ratings on average 3.47 higher than irrelevant items. This difference was significant,

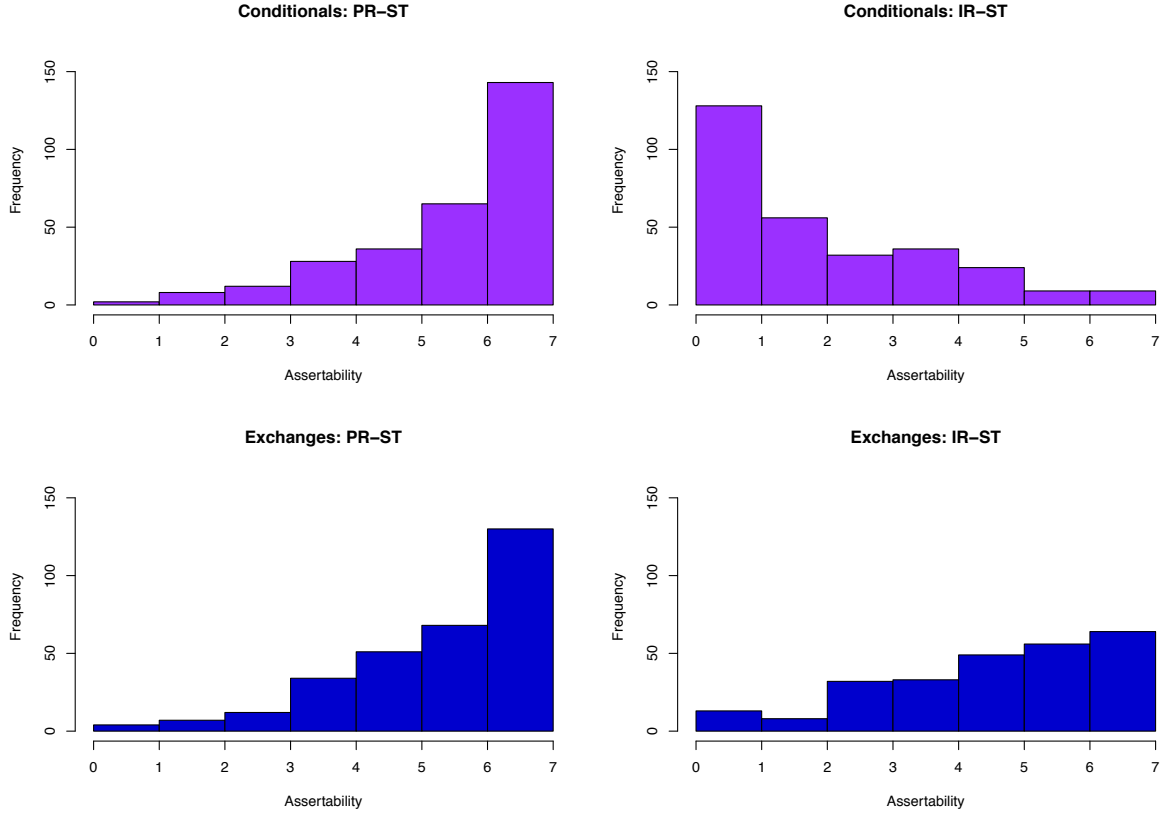


Figure 4. Results for the “assertability” question, ST condition. Note the different scale for the y axis.

$F(1, 99) = 422.62, p < .001$, partial $\eta^2 = .81$. For exchanges, relevant items were given ratings on average .85 higher than irrelevant items. This difference was also significant— $F(1, 99) = 26.88, p < .001$, partial $\eta^2 = .21$ —but note the substantially smaller effect size. Secondly, we compared the effect of type of discourse for each level of probabilistic relevance. For probabilistically relevant items, conditionals were given ratings on average .14 higher than exchanges. This difference was not significant, $F(1, 99) = .65, p = .42$, partial $\eta^2 = .01$. For probabilistically irrelevant items, exchanges were given ratings on average 2.48 higher than conditionals. This difference was significant, $F(1, 99) = 133.25, p < .001$, partial $\eta^2 = .57$.

To support these analyses, we ran complementary robust analyses using the WRS2 package. These analyses used 20% trimmed means. A robust mixed ANOVA (the “bwtrim” function) yielded a significant main effect of probabilistic relevance,

$Q = 355.86$, $p < .001$, and of task, $Q = 59.05$, $p < .001$. The interaction was also significant, $Q = 150.37$, $p < .001$.⁴

We followed up the interaction by carrying out simple effects analyses with robust t-tests. Firstly, we explored the effect of probabilistic relevance for each type of discourse, using robust paired t-tests (the “yuend” function). For conditionals, relevant items were given ratings on average⁵ 3.81 higher than irrelevant items. This difference was significant, $t(30) = 18.69$, $p < .001$, $\xi = .97$. For exchanges, relevant items were given ratings on average .81 higher than irrelevant items, $t(31) = 5.96$, $p < .001$, $\xi = .56$. Secondly, we compared the effect of type of discourse for each level of probabilistic relevance, using robust independent-samples t-tests (the ‘yuen’ function). For relevant items, conditionals were given ratings on average .22 higher than exchanges. This difference was not significant, $t(60.16) = 1.11$, $p = .27$, $\xi = .17$. For irrelevant items, exchanges were given ratings on average 2.78 higher than conditionals. This difference was significant, $t(60.41) = 12.67$, $p < .001$, $\xi = .97$.

Summary. The conventional and robust analyses agreed on the following picture. There was a significant interaction of probabilistic relevance and type of discourse (conditionals vs. exchanges). Probabilistic relevance had a significant effect for both conditionals and exchanges, but the effect was considerably more pronounced with conditionals. For relevant items, conditionals and exchanges did not differ; only in the irrelevant condition did same topic conditionals and exchanges differ significantly.

Finally, all of these findings were mirrored exactly in the “how sensible would it be to say” question. Table 2 shows the descriptive data for this question. Figures 5 and 6 show the distributions of responses for this second question. As can be easily confirmed by visual inspection, these data are virtually indistinguishable from those for the assertability question. For the sake of brevity, we thus omit report of the full statistical analysis, noting only that the corresponding analyses replicate all of the findings of the

⁴The bwtrim function returns the test statistic Q , which is approximately F-distributed. The function returns neither degrees of freedom nor effect sizes.

⁵The mean differences reported for robust analyses are the differences in 20% trimmed means.

		PR-ST	IR-ST	IR-DT
Conditionals	Mean	6.11	2.56	1.37
	SD	0.67	1.19	0.95
Exchanges	Mean	5.88	5.13	1.50
	SD	0.90	0.88	1.16

Table 2

Descriptive data for the “sensibleness” question.

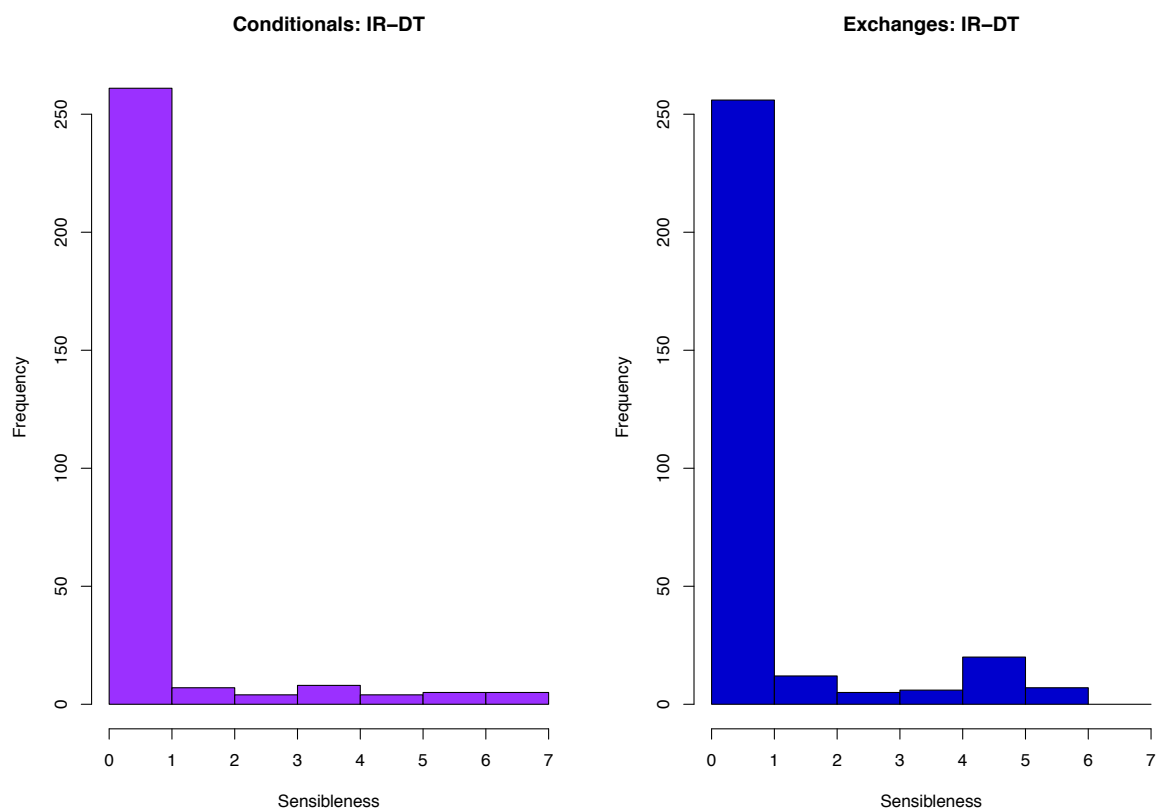


Figure 5. Results for the “sensibleness” question, DT condition.

assertability question.⁶

⁶In response to reviewer suggestion we added a comparison between same and different topic irrelevant conditionals as reported by Skovgaard-Olsen et al. (2016a). While these authors found no significant difference between these, we did. Items with the same topic were rated on average 1.06 higher than with different topics, $t(48) = 8.70$, $p < .001$, $d = 1.03$, and 1.20 higher with 20% trimmed means, robust

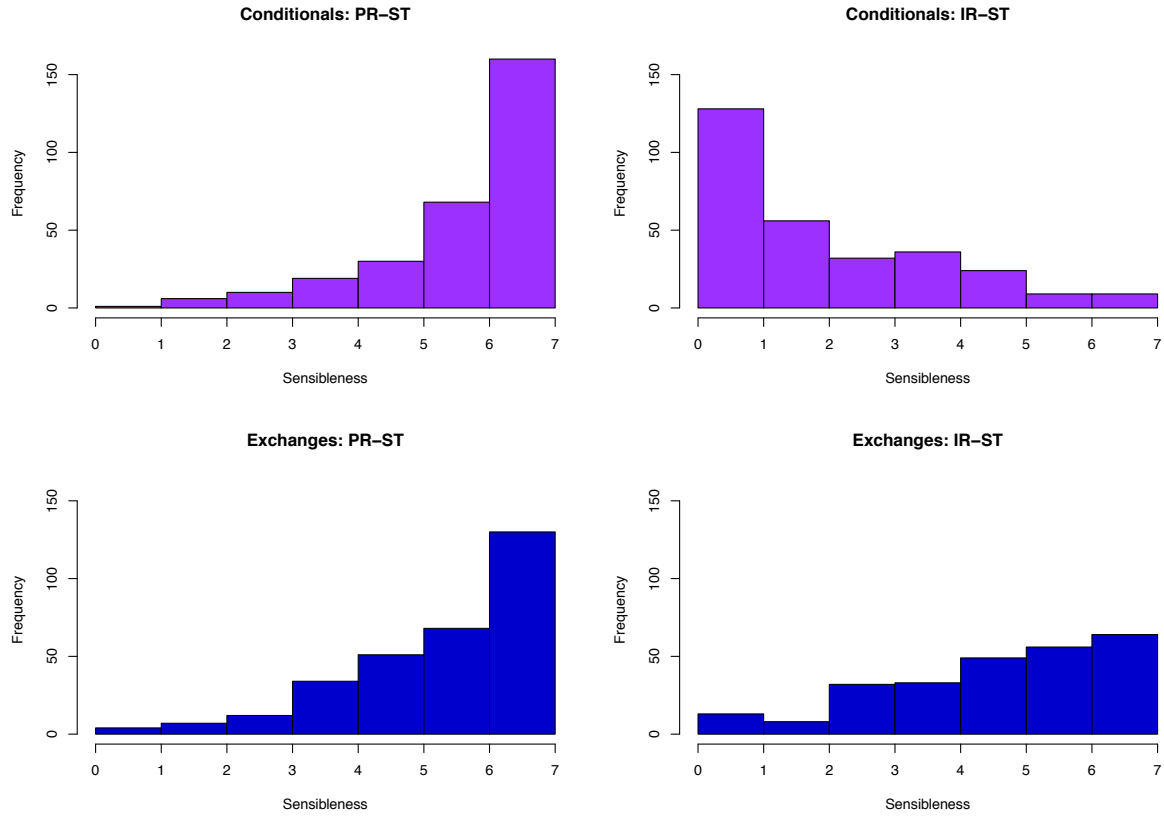


Figure 6. Results for the “sensibleness” question, ST condition. Note the different scale for the y axis.

Discussion

Our data support the hypothesis that people expect a stronger link between the clauses of an indicative conditional than between other consecutive elements of discourse.

Neither self-standing discourse elements nor conditionals are judged as natural to assert or sensible to say when even the minimal discourse coherence requirement is violated (different topics condition). Both types of discourse are judged to be natural and

$t(30) = 9.05, p < .001, \xi = .79$. The same effect was found in the exchanges: mean difference 3.44, $t(51) = 15.81, p < .001, d = 3.27$; and trimmed mean difference of 3.93, $t(31) = 21.22, p < .001, \xi = .96$. These analyses are included in the correction for multiple comparisons. Closer scrutiny of Skovgaard-Olsen et al. (2016a)’s materials reveals a different notion of “same topic” than ours (which is determined by the presence or absence of discourse coherence relations, cf., Lascarides and Asher 1993, p.462-463, on the topic of narration). So their study includes “same topic” items which would be “different topic” items on our terms. There is thus no tension between the two sets of results.

sensible when there is both minimal discourse coherence (i.e. the clauses are on the same topic) and probabilistic relevance between antecedent and consequent. However, patterns of responses diverge when there is minimal discourse coherence but no probabilistic relevance between antecedent and consequent. In this case, self-standing discourse elements were moderately assertable or sensible, while conditionals were largely judged to be unassertable. This allows us to conclude that a conditional is assertable only if its antecedent is relevant for the consequent. A coherence relation between p and q (which may be sufficient to make q a natural response to p in a conversation) is not sufficient for a conditional “if p then q ” to be assertable.

Conversely, this means that the lack of discourse coherence is not a sufficient explanation for the oddity of missing-link conditionals. This poses a challenge to any account of conditionals that does not posit the need for a connection between antecedent and consequent (Douven, 2015). This means another explanation for their oddity is required, such as the fact that they make poor arguments (Oaksford & Hahn, 2007).

At the same time, the results suggest the need to take another look at centering. It might be a valid inference in very specific cases (such as those examined in Cruz et al. 2016). However, it seems unlikely that this inference will be perceived as valid in general—any more than inferences from co-occurrence to causation are necessarily licensed. Future research should be directed at these issues.

References

- Adams, E. W. (1975). *The logic of conditionals*. Dordrecht: D. Reidel.
- Asher, N., & Lascarides, A. (2003). *Logics of conversation*. Cambridge: Cambridge University Press.
- Bennett, J. (2003). *A philosophical guide to conditionals*. Oxford: Oxford University Press.
- Cruz, N., Baratgin, J., Oaksford, M., & Over, D. (2015). Bayesian reasoning with ifs and ands and ors. *Frontiers in Psychology*, 6(192). doi: 10.3389/fpsyg.2015.00192
- Cruz, N., Over, D., Oaksford, M., & Baratgin, J. (2016). Centering and the meaning of conditionals. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th annual conference of the cognitive science society* (p. 1104-1109). Austin, TX: Cognitive Science Society.
- Declerck, R., & Reed, S. (2001). *Conditionals: A comprehensive empirical analysis*. Berlin/New York: Mouton de Gruyter.
- Douven, I. (2008). The evidential support theory of conditionals. *Synthese*, 164(1), 19–44.
- Douven, I. (2015). How to account for the oddness of missing-link conditionals. *Synthese*. doi: 10.1007/s11229-015-0756-7
- Douven, I. (2016). *The epistemology of indicative conditionals*. Cambridge: Cambridge University Press.
- Edgington, D. (1995). On conditionals. *Mind*, 104(414), 235-329.
- Elder, C.-H., & Jaszczolt, K. M. (2016). Towards a pragmatic category of conditionals. *Journal of Pragmatics*(98), 36-53.
- Evans, J. S. B. T., & Over, D. E. (2004). *If*. Oxford: Oxford University Press.
- Grice, H. P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.
- Hobbs, J. R. (1979). Coherence and coreference. *Cognitive Science*, 3, 67-90.
- Jackson, F. (1987). *Conditionals*. Oxford: Blackwell.

- Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review*, 109(4), 646-678.
- Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. Stanford, CA: CSLI Publications.
- Kratzer, A. (1986). Conditionals. *Chicago Linguistic Society*, 22(2), 1-15.
- Krzyżanowska, K. (2015). *Between “if” and “then.” towards an empirically informed philosophy of conditionals* (Unpublished doctoral dissertation). University of Groningen, The Netherlands.
- Lascarides, A., & Asher, N. (1993). Temporal interpretation, discourse relations and commonsense entailment. *Linguistics and Philosophy*, 16, 437-493.
- Mair, P., Schoenbrodt, F., & Wilcox, R. (2016). *WRS2: Wilcox robust estimation and testing*.
- Mair, P., & Wilcox, R. (2016). *WRS2: Robust Statistical Methods*. Retrieved from <https://cran.r-project.org/package=WRS2>
- Oaksford, M., & Hahn, U. (2007). Induction, deduction and argument strength in human reasoning and argumentation. In A. Feeney & E. Heit (Eds.), *Inductive reasoning: Experimental, developmental, and computational approaches* (pp. 269-301). Cambridge: Cambridge University Press.
- Oberauer, K., Weidenfeld, A., & Fischer, K. (2007). What makes us believe a conditional? the roles of covariation and causality. *Thinking & Reasoning*, 13(4), 340-369.
- Over, D. E., Hadjichristidis, C., Evans, J. S. B. T., Handley, S. J., & Sloman, S. A. (2007). The probability of causal conditionals. *Cognitive Psychology*, 54, 62-97.
- Politzer, G., & Baratgin, J. (2015). Deductive schemas with uncertain premises using qualitative probability expressions. *Thinking & Reasoning*, 22(1), 78-98. doi: 10.1080/13546783.2015.1052561
- R Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.r-project.org/>

- Skovgaard-Olsen, N., Singmann, H., & Klauer, K. C. (2016a). Relevance and reason relations. *Cognitive Science*. doi: 10.1111/cogs.12462
- Skovgaard-Olsen, N., Singmann, H., & Klauer, K. C. (2016b). The relevance effect and conditionals. *Cognition*, 150, 26-36.
- Stalnaker, R. C. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in logical theory* (p. 98-112). Oxford: Blackwell.
- Stalnaker, R. C. (1975). Indicative conditionals. *Philosophia*, 5(3), 269-286.
- Stojnić, U. (2016). One's *Modus Ponens*: Modality, coherence and logic. *Philosophy and Phenomenological Research*. doi: doi:10.1111/phpr.12307
- Wilcox, R. R. (2016). *Understanding and applying basic statistical methods using R*. Hoboken, NJ: John Wiley & Sons.