...that P is relevant for Q: Indicative conditionals and learning from testimony

Karolina Krzyżanowska (k.h.krzyzanowska@uva.nl)
Institute for Logic, Language and Computation, University of Amsterdam
Science Park 107, 1098 XG Amsterdam, The Netherlands

Peter J. Collins (p.j.collins.65@gmail.com)
Department of Psychology, Goldsmiths, University of London
New Cross, London, SE14 6NW

Ulrike Hahn (u.hahn@bbk.ac.uk)
Department of Psychological Sciences, Birkbeck, University of London
Malet Street, London, WC1E 7HX U.K.

Abstract

Our beliefs change with learning, and much of what we learn comes from the testimony of other people. How much our beliefs change may depend on how many people are the sources of a given piece of information, and how reliable their expertise makes them. It is not clear, however, what exactly the effects of reliability or number of speakers will be when the testimony has the form of an indicative conditional. Here, we test the hypothesis that learning a conditional amounts to increasing the degree to which the antecedent of that conditional is relevant for its consequent. Furthermore, we investigate whether this is affected by number of speakers and by their expertise.

Keywords: indicative conditionals; probabilistic relevance; testimony; source reliability

Introduction

We form beliefs in response to what we learn. We learn from our own experiences, from observation of the world, but, also, from other's testimony: that is, from their saying, uttering, or asserting a proposition. Often, the testimony we receive has a form of an indicative conditional, typically: "If φ, (then) ψ," where φ and ψ—the conditional's antecedent and consequent, respectively—can, in principle, stand for any sentences. For instance, a doctor may tell a patient that if they do not quit smoking they put themselves at high risk of developing lung cancer. Students in a biology class may hear that if an animal has gills, it can breathe under water, and an anonymous internet user may try to convince other internet users that if we cannot see the horizon curving when we look out of the window, the Earth must be flat.

Sometimes we accept those testimonies with unconditional trust; on other occasions, we take them with a grain of salt. How does the reliability of the speaker affect the way we change our beliefs in response to what we are being told? And what happens when the same information is repeated by more than one speaker? When testimony takes the form of a categorical statement, the answer appears relatively straightforward. For instance, when a professor of medicine tells us that Bill, a malaria patient at an infectious diseases ward, will make a good recovery, we may become fully confident that Bill will recover. By contrast, when the same testimony comes from a waiter in the hospital's cafeteria, we might be less inclined to form a strong belief about Bill's prognosis—our subjective degrees of belief in the proposition "Bill will make a good recovery" remains lower than when the testimony comes from an expert in the relevant field. The professor of medicine is clearly a more reliable source of information about patients' prognoses than a waiter, and we should expect this reliability to factor in belief change (Briñol & Petty, 2009; Collins, Hahn, von Gerber, & Olsson, 2018; Hahn, Harris, & Corner, 2009; Petty & Briñol, 2008). However, when we hear the same information not only from a waiter, but also from another patient at the same ward, from a visitor to the hospital, and from a relative on the phone, we might become somewhat more confident that Bill will make a good recovery than if that information came from any of those sources alone (see, e.g. Bovens & Hartmann, 2003; Schum & Martin, 1982).

But how do we change our beliefs when the testimony has a form of an indicative conditional? What happens when, instead of a categorical statement about the prognosis of Bill, a patient on an infectious diseases ward, we learn (1)?

(1) If Bill has malaria, then he will make a good recovery.

If source reliability affects the strength of the beliefs we form based on testimony, what is actually affected when the testimony is conditional in form?

Although reasoning with conditionals has attracted a lot of attention in philosophy and in psychology of reasoning in the recent years (e.g., Evans & Over, 2004; Oaksford & Chater, 2007; Douven, 2016), the question of learning from conditionals remains under-researched. Conditional reasoning tasks simply presuppose that participants treat the conditional premises as something they have learned, notable exception being Stevenson and Over (2001) who studied the effect of the expertise of a speaker asserting the premises on conditional reasoning. This is not to say that no one tried to answer the question of how to update on a conditional, that is, how to accommodate a new piece of information conveyed by a conditional. In fact, in the belief revision literature, various update rules, that is, procedures for calculating one's posterior degrees of belief, have been proposed (see Douven, 2016, ch. 6, for an overview). However, this work has tended to concern itself with the question of how to maintain a coherent set of beliefs upon learning that the conditional probability of a conditional's consequent, ψ, given its antecedent, ϕ, is very high. Although a high conditional probability is likely part of what someone asserting an indicative conditional communicates (Baratgin, Politzer, & Over, 2013; Over, Hadjichristidis, Evans, Handley, & Sloman, 2007), and hence it is a plausible assumption that learning a conditional amounts to fixing one's $\Pr(q \mid p)$ to a high value, it is not necessarily the only information we receive when we learn a conditional (cf. Collins, 2017). What it is that we learn when someone asserts a conditional in the sense discussed here is thus a question orthogonal to that pursued in the belief revision literature.

Our question of what a person learns upon hearing a conditional asserted is closely tied to the question of what a conditional means, that is, what kind of information constitutes its "core," conventional meaning. Indeed, the aforementioned assumption underlying Bayesian belief revision—that assertion of a conditional communicates that the corresponding conditional probability is high—constitutes the gist of the Suppositional Theory (e.g., Adams, 1975; Bennett, 2003; Edgington, 1995). While the Suppositional Theory has received a lot of empirical support (see, e.g., Over et al., 2007; Oberauer, Weidenfeld, & Fischer, 2007; Politzer, Over, & Baratgin, 2010), recent developments in philosophy and psychology suggest that high conditional probability might not be all there is to the "core" meaning of indicative conditionals. Recent work has sought evidence for an approach to conditionals, dubbed *inferentialism*, that claims antecedent and consequent must be connected for a conditional to be acceptable (e.g., Skovgaard-Olsen, Singmann, & Klauer, 2016; Douven, Elqayam, Singmann, & van Wijnbergen-Huitink, 2018). We next discuss how the connection between the clauses of a conditional may be operationalised, and what follows for the question of learning from conditional testimony.

Connecting antecedents and consequents

While few would deny that conditionals seem to convey that φ and ψ are somehow connected, the connection between φ and ψ tends to be relegated to pragmatics.[1] What distinguishes inferentialist approach from the alternative views on conditionals, such as the Suppositional Theory (ST) mentioned above or the Mental Models Theory (MMT) (e.g. Johnson-Laird, Khemlani, & Goodwin, 2015) is that the relationship between the antecedent and consequent of a conditional is treated as a part of the conventional, semantic meaning of a conditional. This relationship is typically understood as a form of an inductive, abductive, or deductive *inferential connection* (e.g. Krzyżanowska, Wenmackers, & Douven, 2013; Douven et al., 2018), or as a probabilistically understood *reason relation* (e.g. Skovgaard-Olsen et al., 2016; Skovgaard-Olsen, Kellen, Hahn, & Klauer, 2019). Consequently, learning that "If φ, then ψ," on this view, amounts to learning that φ is a reason for ψ or that ψ can be inferred from φ. In the present paper, we will focus on the latter version of the approach.

In the psychology of reasoning, it is customary to characterise the relevance relation between φ and ψ in terms of the Δp-rule (see, e.g., Over et al., 2007; Skovgaard-Olsen et al., 2016). Δp is defined as the difference between the conditional probability of a conditional's consequent given its antecedent, $\Pr(\psi \mid \varphi)$ and the conditional probability of that conditional's consequent given the negation of its antecedent $\Pr(\psi \mid \neg\varphi)$:

$$\Delta p = \Pr(\psi \mid \varphi) - \Pr(\psi \mid \neg\varphi) \quad (1)$$

When Δp is greater than 0, φ is said to be *positively relevant* for ψ. For instance, if the probability of Bill making a good recovery is higher under the assumption that he has malaria than under the assumption that he has any other disease treated on an infectious diseases ward, we may say that the malaria diagnosis increases the probability of good recovery and hence the antecedent of (1) is positively relevant for its consequent. But when $\Delta p = 0$, φ is said to be *irrelevant* for ψ. Take, for instance, the following sentence:

(2) If Bill likes jazz, he will make a good recovery.

Given that, to the best of our knowledge, the musical preferences of patients on infectious diseases wards have no bearing on their prognoses, the probability of Bill's good recovery will be the same on the supposition that he is fond of jazz as on the supposition that he does not like it at all. Therefore, the antecedent of (2) is probabilistically irrelevant for the consequent of this conditional.

Finally, whenever $\Delta p < 0$, the relationship between φ and ψ is that of negative probabilistic relevance, which means that φ decreases the probability of ψ, for instance:

(3) If Bill has terminal cancer, he will make a good recovery.

When our initial degree of belief that Bill, a patient on an infectious diseases ward, will make a good recovery is positive, learning that he has terminal cancer may decrease our degrees of belief in his good prospects. In such a case, the probability of the consequent of (3) given the negation of its antecedent (i.e., "Bill does not have terminal cancer") is higher than the probability of the consequent given the antecedent.

The presence or absence of the connection between a conditional's antecedent and its consequent has been shown to affect how people evaluate the probabilities of conditionals (e.g., Skovgaard-Olsen et al., 2016; Skovgaard-Olsen, Kellen, et al., 2019) as well as their truth values (Douven et al., 2018). These results suggest that the connection could be treated as an aspect of the "core" meaning of the

[1] We focus our attention here on what has been dubbed *standard* conditionals. It has became customary to classify as *non-standard* or *nonconditional* conditionals sentences by means of which a speaker actually asserts the conditional's consequent such as so-called *biscuit conditionals*, e.g., "If you're hungry, there are biscuits on the table," and *non-interference conditionals* such as "(Even) If we triple her salary, Betty will leave the department" (Douven, 2016). See also (Elder & Jaszczolt, 2016) for a useful discussion on the definition of the conditional itself.

conditional—its semantic, conventional content. If this is so and if the probabilistic relevance relation is a correct characterisation of that connection, we should expect that learning a conditional from a testimony will affect people's estimates of Δp. Moreover, assuming that source reliability depends on speaker expertise and that the number of sources repeating the same testimony affects its acceptance, we expect to see their effect on the value of Δp, too.[2]

The present experiments

Here, we report two experiments that build on the studies of Stevenson and Over (2001) who investigated the effect of source reliability on people's performance in conditional reasoning tasks. However, we are not interested in the extent to which people endorse various conclusions depending on the expertise of the speaker who asserts the premises, but, instead, we are interested in people's interpretation of the conditionals that served as the premises in Stevenson and Over's work. More specifically, we test the hypothesis that when people hear an assertion of a conditional, "If φ then ψ," what they learn is that φ is positively relevant for ψ. This amounts to predicting that Δp, calculated on the basis of people's judgements of the conditional probability of ψ given φ and of the conditional probability of ψ given $\neg\varphi$, will reliably increase on the assertion of a conditional. Additionally, we hypothesise that, when there are multiple sources or when sources are reliable, Δp will be reliably higher than when there is a single source or when sources are less reliable.

To test the main hypothesis, we compared people's Δp estimates in a context in which no conditional was asserted (Null condition) with their Δp estimates in a context in which someone asserted a conditional. To test the first of the additional hypotheses, Experiment 1 had two assertion conditions: the conditional could be asserted either by a single speaker (Single assertion) or by several different speakers (Multiple assertions). The presence and the number of assertions were manipulated between-participants. Below are examples of items belonging to each condition:

Null Imagine that you are at an infectious diseases ward.

Single Olivia is at an infectious diseases ward. She says, 'If a patient on this ward has malaria, then he'll make a good recovery.'

Multiple Olivia, Aaron, Zoe and Felix are at an infectious diseases ward. They say, 'If a patient on this ward has malaria, then he'll make a good recovery.'

In Experiment 2, we investigated how probabilistic relevance estimates change depending on the expertise of the source asserting the conditional. As discussed above, a professor of a medicine is a more reliable source of information about a patient's prognosis than a medical student, and so is, say, a butterfly keeper compared to a child visiting the butterfly house in a context of a conversation about properties of butterflies. As in Experiment 1, we compared the effect of assertions made by speakers with different expertise to a null condition where no assertion is made.

Inexpert Imagine that you are at an infectious diseases ward. A medical student tells you, 'If a patient on this ward has malaria, then he'll make a good recovery.'

Expert Imagine that you are at an infectious diseases ward. A professor of medicine tells you, 'If a patient on this ward has malaria, then he'll make a good recovery.'

This manipulation was also between-participant. In both experiments, the above items were followed by questions about the probability of the consequent (i.e., that a patient will make a good recovery) given the antecedent, (i.e., that the patient has malaria) and a question about the probability of the consequent given a negation of the antecedent (i.e., given that the patient does not have malaria).

Experiment 1: Assertion

The first experiment tests the hypothesis that the Δp calculated on the basis of people's judgements of conditional probability of ψ given φ and of conditional probability of ψ given $\neg\varphi$ increases in response to an assertion of an indicative conditional. Additionally, we probe if number of speakers affects the degree of Δp change upon learning a conditional.

Method

Participants. 175 participants (76 female; mean age 38.16) completed this experiment on the Amazon Mechanical Turk (`https://www.mturk.com/`; 5 participants had already been excluded since they were non-native English speakers. Participants were recruited via the intermediary MTurk Data (`www.mturkdata.com`). High qualifications were set for the task to improve the quality of the data and to maximize the number of native English speakers: participants had to be resident in the US, and have an overall approval rating of 99% for 1,000 previously completed tasks. Participants received a small fee, chosen to exceed the US minimum wage per minute.

Materials and procedure. Participants were assigned, in a round-robin fashion, to one of the three conditions: Null, Single Assertion, and Multiple Assertion. After giving informed consent, participants gave demographic information and were

[2] An anonymous reviewer has pointed out to us a possible limitation of this prediction. One can assert a conditional whose consequent is certain given its antecedent due to the fact that it expresses some kind of a deductive reasoning process, as in, e.g., "If it is Tuesday today, then tomorrow is Wednesday" (cf. Douven & Verbrugge, 2010; Krzyżanowska et al., 2013). In such cases, one would expect both prior and posterior (that is, calculated, respectively, before and after hearing the conditional) Δp to equal 1, and thus no change in Δp would have been observed. However, one could argue that this would not be a genuine instance of learning, unless the hearer does not know yet that Tuesdays are followed Wednesdays, in which case their prior would be different. Note, also, that we do not claim that Δp captures all there is to the meaning of a conditional on inferential accounts.

then given the instructions explaining the task and showing them an example item (included for illustration; it was not an experimental item).

Each participant was presented with 7 items. The order of presentation was counterbalanced across participants. Each item appeared on its own page, with all probability questions given below the item. These questions collected judgements of the parameters of a probabilistic model. For present purposes, only the following questions are relevant:[3]

(A) What's the probability that [Consequent] given [Antecedent]?

(B) What's the probability that [Consequent] given [Not Antecedent]?

The dependent measure was Δp, calculated by subtracting people's responses to question (B) from (A) above.

Participants provided ratings for each probability question for each item on an 11-point Likert-style scale from *not at all possible* to *certain*. Finally, participants were debriefed.

Results

Figure 1 reports the descriptive data.[4]

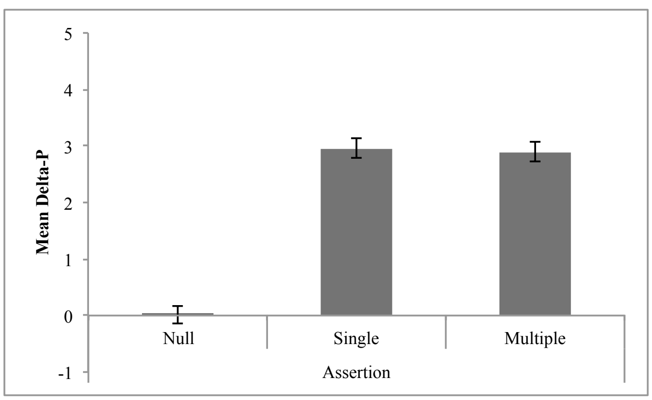

Figure 1: Mean Δp (calclulated by subtracting participants' estimates of $\Pr(\psi|\varphi)$ and $\Pr(\psi|\neg\varphi)$ on 11-point scale) by Assertion condition. Error bars are standard error.

The descriptive data suggest a clear increase from the Null Condition ($M = .02$), on the one hand, to the Single ($M = 2.95$) and Multiple ($M = 2.90$) Conditions. Inferential analysis confirmed this. Two linear mixed-effects models were fit using the maximum-likelihood method in R (R Core Team, 2016) and the lme4 package (Bates, Mächler, Bolker, & Walker, 2015). Because of convergence problems, these models excluded estimates of the covariance of random slopes and random intercepts. Otherwise, the full random-effects structure was used: random slopes for Assertion across items, and random intercepts for items and participants. The full model significantly improved fit over the null model, $\chi^2(2) = 21.54$, $p < .001$. Table 1 reports the (treatment-coded) parameter estimates (calculated using the Wald method) for the full model.

Fixed Effect	Parameter	95% CI
Intercept (Null)	$b = .021$	-1.24, 1.28
Single	$b = 2.93$	2.14, 3.71
Multiple	$b = 2.87$	2.08, 3.67

Table 1: Fixed effects of Assertion on Δp

Pairwise comparisons were conducted on the estimated marginal means using the Tukey correction for multiple comparisons with the lsmeans package (Lenth, 2016). The increase from Null to Single conditions was significant ($t(53.76) = 7.25$, $p < .001$), as was the increase from Null to Multiple conditions ($t(35.75) = 6.95$, $p < .001$). The difference between Single and Multiple conditions was not significant ($t(56.55) = -.16$, $p = .99$).

Summary These data suggest a clear effect of Assertion on the participants' perceived probabilistic relevance: the value of Δp increased significantly when the conditional was asserted. However, we did not observe any significant differences in Δp depending on the number of speakers.

The Expertise Experiment

The second experiment tests the hypothesis that the increase in Δp, calculated as in the first experiment, will depend on the reliability of the speaker asserting the conditionals, that is, their expertise on the topic the asserted conditional is about.

Method

Participants. 176 participants (75 female; mean age 35.97) completed this experiment; 5 participants had already been excluded since they were non-native English speakers. Participants were recruited via the intermediary MTurk Data (`http://www.mturkdata.com`) with the same high qualifications and remuneration system as in Experiment 1.

Materials and procedure.

Participants were assigned, in a round-robin fashion, to one of the three conditions: Null, Inexpert Source, Expert Source. After giving informed consent, participants gave demographic information and were then shown the same instructions as in the Experiment 1. Each participant was then presented with 6 items, each on its own page. The presentation order was counterbalanced across participants. These were the same conditionals that were used in Experiment 1

[3]For more details about the experimental design and the report of remaining findings, see Collins (2017, ch. 2-4). See also Collins, Krzyżanowska, Hartmann, Wheeler, and Hahn (2020) for a discussion of probabilistic models of updating on conditionals and for a discussion of related empirical findings in a broader context of philosophical and psychological literature on conditionals.

[4]Note that the figures, in this paper, present data averaged over items. However, if we compare these descriptives with the unstandardized coefficients and estimated marginal means, we can see that the figures provide a reasonable summary.

except for one item which, during pre-testing of the items, failed the source expertise manipulation.

The items were followed by the same set of questions as the previous experiment. As above, only the questions about $\Pr(Consequent = True \mid Antecedent = True)$ and $\Pr(Consequent = True \mid Antecedent = False)$ are relevant for present purposes. The same wordings were used as for the previous experiment. As in Experiment 1, the dependent measure was Δp, calculated by subtracting participant's responses to question (B) from (A) above. Participants provided ratings for each probability question for each item on a 11-point Likert-style scale from *not at all possible* to *certain*. Finally, participants received debriefing information.

Results

Figure 2 reports the descriptive data.

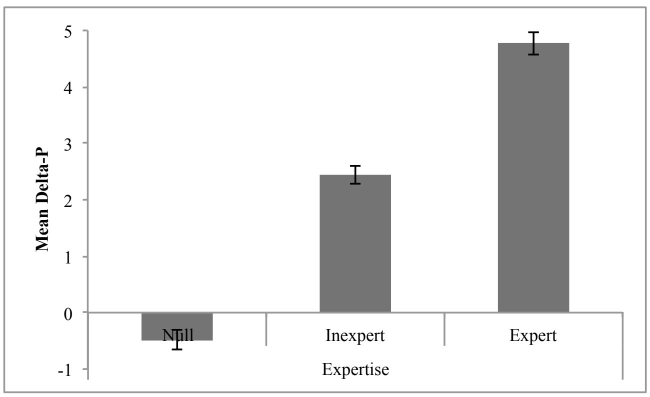

Figure 2: Mean Δp (calclulated by subtracting participants' judgments of $\Pr(\psi \mid \varphi)$ and $\Pr(\psi \mid \neg\varphi)$ on 11-point scale) by Expertise Condition. Error bars are standard error.

The descriptive data suggest a strong linear effect of Expertise, with Δp increase from the Null ($M = -.048$) to Inexpert ($M = 2.44$) and Expert ($M = 4.76$) conditions. Two linear mixed-effects models were fit using the maximum-likelihood method: a full model including the fixed effect, and a null model excluding it. Both models included the full random structure justified by the design: random slopes for Expertise across items, random intercepts for items, and random intercepts for participants. The covariance of slopes and intercepts was also estimated. The full model significantly improved fit over the null model, $\chi^2(2) = 31.32$, $p < .001$. Table 2 reports the (treatment-coded) parameter estimates (calculated using the Wald method).

Fixed Effect	Parameter	95% CI
Intercept (Null)	$b = -.48$	-1.51, .55
Inexpert	$b = 2.92$	1.96, 3.88
Multiple	$b = 5.24$	4.54, 5.95

Table 2: Fixed effects of Expertise on Δp

Once again, pairwise comparisons were conducted on the estimated marginal means using the Tukey correction for multiple comparisons with the lsmeans package (Lenth, 2016). The increase from the Null to Inexpert conditions was significant ($t(14.06) = 5.71$, $p < .001$), as was the increase from Null to Expert conditions ($t(32.34) = 14.40$, $p < .001$), and from Inexpert to Expert conditions ($t(18.65) - 5.05$, $p < .001$).

Summary

As in Experiment 1, we observed that the assertion of a conditional significantly increased the value of Δp calculated on the basis of the participants' conditional probability judgements. This increase was observed both when the conditional was asserted by a reliable speaker (Expert) and when it was asserted by an unreliable speaker (Inexpert). Furthermore, we found a significant difference between Inexpert and Expert conditions: the probabilistic relevance relation conveyed by a conditional was stronger—Δp was higher—in the Expert condition compared to the Inexpert condition.

Discussion

Our data clearly show that when people receive information in the form of an indicative conditional, they learn that the antecedent is positively relevant for the consequent. Expressed mathematically, they set the probability of that conditional's consequent given its antecedent to a higher degree than the probability of the consequent given the negation of the antecedent. In other words, as predicted by the inferential view on conditionals, the participants interpret the speaker as asserting that the conditional's antecedent is a reason for the consequent. Moreover, the more reliable the speaker, the stronger the relationship between the asserted conditional's antecedent and consequent appears to be.

The data did not meet our expectations in one respect: in the Assertion Task, multiple assertion did not reliably increase Δp. Indeed, there was a (very) small decrease from the single to multiple assertion conditions. We suspect that, underlying this result, is an unintended ambiguity in the experimental materials. In the multiple assertion condition, sources could be understood as making separate, independent observations or reporting the same dependent observation. While we do not have data to test this possibility directly, Collins (2017) reports an experiment which explores a similar data pattern with the conditional probability. That experiment suggests that a condition with unambiguously independent testimony reliably increased the conditional probability. It seems reasonable to expect a similar outcome for Δp.

We should note that our results are not incompatible with other prominent views on conditionals in psychology of reasoning, ST and MMT. Nevertheless, on the ST, a speaker asserting a conditional conveys that the conditional probability of that conditional's consequent given its antecedent is high, but this can happen when the conditional probability of the consequent given the negation of the antecedent is equally high or even higher, resulting in Δp being 0 or negative, respectively.

On the MMT, a conditional is said to convey a conjunction of possibilities "Possibly($\varphi \wedge \psi$) and Possibly($\neg\varphi \wedge \neg\psi$) and Possibly($\neg\varphi \wedge \psi$)," and, implicitly, that $\varphi \wedge \neg\psi$ is impossible (Johnson-Laird et al., 2015; Khemlani, Byrne, & Johnson-Laird, 2018). While it is not clear how to assign probabilities to modal statements, we can entertain a charitable interpretation on which a possibility of a proposition φ implies that φ has a non-zero probability value. For instance, we could take the probability of the impossibility to be 0 and assign a positive probability value to all the remaining, possible conjunctions, summing up to 1. If we calculate the conditional probabilities based on such probability distribution, we will see that $\Pr(\psi|\varphi) = 1$, while $0 < \Pr(\psi|\neg\varphi) < 1$, resulting in a positive Δp. However, such an interpretation of the probabilities corresponding to the MMT conditional leaves no space for uncertainty, and, hence would not be able to account for our results on the effects of speaker expertise. Interestingly, it seems dubious that the proponents of the MMT would approve of such an interpretation of their theory given that they emphasise that the connection between antecedents and consequents does *not* belong to the "core" meaning of the conditional, whereas, on this interpretation, a positive probabilistic relevance relation would be built in into it.

Of course, advocates of the above theories do not deny that the connection between φ and ψ can belong to what "If φ then ψ" communicates, but they argue that this is a result of pragmatics of indicative conditionals, rather than an aspect of their core meaning (see, e.g., Over et al., 2007, p. 92, or Johnson-Laird & Byrne, 2002, p. 600). None of the theories, however, elaborate on what kind of a pragmatic mechanism is meant to be responsible for this effect or in virtue of which pragmatic principles, which maxims of good conversations, a conditional conveys such an additional meaning. Importantly, recent experiments have shown that the connection between antecedents and consequents cannot be easily captured in pragmatic terms. For instance, it cannot be explained as a simple matter of discourse coherence (Krzyżanowska, Collins, & Hahn, 2017), and it does not share the signatures of uncontroversially pragmatic phenomena such as conversational implicatures or *not-at-issue* content (Skovgaard-Olsen, Collins, Krzyżanowska, Hahn, & Klauer, 2019).

Although our studies can be seen as supporting the inferential approach to conditionals, it is important to acknowledge their shortcomings. The reported experiments were envisaged as a test of a specific prediction derived from the inferential account; were we to find that an assertion of a conditional does not affect the participants' Δp at all, the account would have been falsified. However, our data confirmed the prediction. We should emphasise nonetheless that we do *not* claim that our findings constitute any form of ultimate proof that inferentialism is *the* account of conditionals, nor was this the purpose of our research. Instead, our findings offer but a glimpse into what is going on when people learn a conditional from a testimony—which is an immensely under-researched issue—which will, hopefully, serve as the starting point for future studies.

Acknowledgments

All authors were funded by awards of the Alexander von Humboldt Foundation to Ulrike Hahn (Anneliese Maier Research Award). Peter Collins was funded by ESRC studentship ES/J500021/1. Karolina Krzyżanowska was supported by "The Logic of Conceivability" Project funded by the European Research Council (ERC CoG), Grant Number 681404. Karolina Krzyżanowska is additionally affiliated with Arché Research Center, at the University of St Andrews. Peter Collins is now at the Department of Psychology, Social Work, and Counselling, at the University of Greenwich. The Authors would like to thank three anonymous reviewers for their helpful comments.

References

Adams, E. W. (1975). *The logic of conditionals*. Dordrecht: D. Reidel.

Baratgin, J., Politzer, G., & Over, D. E. (2013). Uncertainty and the de finetti tables. *Thinking & Reasoning*, *19*(3), 308-328.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models uding lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Bennett, J. (2003). *A philosophical guide to conditionals*. Oxford: Oxford University Press.

Bovens, L., & Hartmann, S. (2003). *Bayesian epistemology*. Oxford: Oxford University Press.

Briñol, P., & Petty, R. E. (2009). Source factors in persuasion: A self-validation approach. *European Review of Social Psychology*, *20*, 49–96.

Collins, P. J. (2017). *Rationality, pragmatics, and sources*. Unpublished doctoral dissertation, Birkbeck, University of London.

Collins, P. J., Hahn, U., von Gerber, Y., & Olsson, E. J. (2018). The bi-directional relationship between source characteristics and message content. *Frontiers in Psychology*, *9*(18).

Collins, P. J., Krzyżanowska, K., Hartmann, S., Wheeler, G., & Hahn, U. (2020). *Conditionals and testimony.* (Manuscript)

Douven, I. (2016). *The epistemology of indicative conditionals*. Cambridge: Cambridge University Press.

Douven, I., Elqayam, S., Singmann, H., & van Wijnbergen-Huitink, J. (2018). Conditionals and inferential connections: A hypothetical inferential theory. *Cognitive Psychology*, *101*, 50-81.

Douven, I., & Verbrugge, S. (2010). The Adams family. *Cognition*, *117*, 302-318.

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.

Hahn, U., Harris, A. J., & Corner, A. (2009). Argument content and argument source: An exploration. *Informal Logic*, *29*(4), 337–367.

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review*, *109*(4), 646-678.

Johnson-Laird, P. N., Khemlani, S. S., & Goodwin, G. P. (2015). Logic, probability, and human reasoning. *Topics in Cognitive Sciences*, *19*(4), 201-214.

Khemlani, S. S., Byrne, R. M. J., & Johnson-Laird, P. N. (2018). Facts and possibilities: A model-based theory of sentential reasoning. *Cognitive Science*, *42*(6), 1887-1924.

Krzyżanowska, K., Collins, P. J., & Hahn, U. (2017). Between a conditional's antecedent and its consequent: Discourse coherence vs. probabilistic relevance. *Cognition*, *164*, 199–205.

Krzyżanowska, K., Wenmackers, S., & Douven, I. (2013). Inferential conditionals and evidentiality. *Journal of Logic, Language and Information*, *22*(3), 315-334.

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, *69*(1), 1–33.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford 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.

Petty, R. E., & Briñol, P. (2008). Persuasion: from single to multiple to metacognitive processes. *Perspectives on Psychological Science*, *3*(2), 137–147.

Politzer, G., Over, D. E., & Baratgin, J. (2010). Betting on conditionals. *Thinking & Reasoning*, *16*(3), 172-197.

R Core Team. (2016). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.r-project.org/`

Schum, D. A., & Martin, A. W. (1982). Formal and empirical research on cascaded inference in jurisprudence. *Law & Society Review*, *17*(1), 105-152.

Skovgaard-Olsen, N., Collins, P., Krzyżanowska, K., Hahn, U., & Klauer, K. C. (2019). Cancellation, negation, and rejection. *Cognitive Psychology*, *108*, 42-71.

Skovgaard-Olsen, N., Kellen, D., Hahn, U., & Klauer, K. C. (2019). Norm conflicts and conditionals. *Psychological Review*, *126*(5), 611-633.

Skovgaard-Olsen, N., Singmann, H., & Klauer, K. C. (2016). The relevance effect and conditionals. *Cognition*, *150*, 26-36.

Stevenson, R. J., & Over, D. E. (2001). Reasoning from uncertain premises: Effects of expertise and conversational context. *Thinking & Reasoning*, *7*(4), 367-390.