Cognition 150 (2016) 26–36

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

The relevance effect and conditionals ☆,☆☆

Niels Skovgaard-Olsen [a,b,*], Henrik Singmann [c], Karl Christoph Klauer [b]

[a] Department of Philosophy, University of Konstanz, Konstanz, Germany
[b] Institut für Psychologie, Albert-Ludwigs-Universität Freiburg, Freiburg, Germany
[c] Department of Psychology, University of Zürich, Zürich, Switzerland

ARTICLE INFO

Article history:
Received 25 August 2015
Revised 26 December 2015
Accepted 28 December 2015

Keywords:
Indicative conditionals
The New Paradigm
Relevance
Reasons
Concessive conditionals
The Equation

ABSTRACT

More than a decade of research has found strong evidence for P(if A, then C) = P(C|A) (“the Equation”). We argue, however, that this hypothesis provides an overly simplified picture due to its inability to account for relevance. We manipulated relevance in the evaluation of the probability and acceptability of indicative conditionals and found that relevance moderates the effect of P(C|A). This corroborates the Default and Penalty Hypothesis put forward in this paper. Finally, the probability and acceptability of concessive conditionals (“Even if A, then still C”) were investigated and it was found that the Equation provides a better account of concessive conditionals than of indicatives across relevance manipulations.

1. Introduction

In philosophy, there is a widely shared consensus that Stalnaker's Hypothesis is wrong and that Adams' Thesis is correct, due to formal problems affecting the former but not the latter – known as the triviality results.

> Stalnaker's hypothesis: P(if A, then C) = P(C|A) for all probability distributions where
> P(A) > 0 and ‘If A, then C’ expresses a proposition.
> Adams' thesis: Acc(if A, then C) = P(C|A) for all simple conditionals (i.e., conditionals whose antecedent and consequent clauses are not themselves conditionals), where ‘Acc(if A, then C)’ denotes the degree of *acceptability* of ‘If A, then C’.[1]
> Triviality Results: Lewis' triviality results show that there is no proposition whose probability is equal to P(C|A) for all probability distributions without the latter being subject to trivializing features such as that P(C|A) collapses to P(C) or that positive probabilities can only be assigned to two pairwise incompatible propositions (Bennett, 2003: chap. 5; Woods, 1997: chap. 4, p. 114–8).

In psychology, there has been a tendency to endorse a thesis very similar to Stalnaker's hypothesis, known as the Equation, which avoids the problems affecting the former by either denying that conditionals express propositions altogether or by endorsing three-valued de Finetti truth tables (Table 1).

At present, the theories united under the heading ‘the New Paradigm of Reasoning’, which endorse the Equation, have branched out in different directions. To name just a few, in Baratgin, Politzer, and Over (2013) and Politzer, Over, and Baratgin (2010), the Equation is studied in relation to three-valued de Finetti truth tables in general and its relation to conditional bets is emphasized. In Pfeifer and Kleiter (2011) and Pfeifer (2013), the Equation is endorsed on the basis of a coherence-based probability logic that works with intervals of imprecise probabilities. However, what matters for our purposes is not so much the exact theory in which the Equation is embedded but rather the general commitment to the Equation. As it stands, over a decade of empirical research has found strong evidence in favor of the Equation and a recent study has begun to challenge Adams' Thesis, as nicely outlined in Douven (2015b: chap. 3, 4).

In contrast, a basic intuition that has emerged repeatedly throughout the history of philosophy is that in conditionals like ‘If it rains, then the match will be cancelled’ the antecedent and

☆ This work was supported by grants to Wolfgang Spohn and Karl Christoph Klauer from the Deutsche Forschungsgemeinschaft (DFG) as part of the priority program “New Frameworks of Rationality” (SPP 1516).
☆☆ Supplementary materials including all data and analysis scripts are available at: https://osf.io/j4swp/.

* Corresponding author at: Department of Philosophy, University of Konstanz, Konstanz, Germany.
E-mail addresses: niels.skovgaard.olsen@psychologie.uni-freiburg.de, n.s.olsen@gmail.com (N. Skovgaard-Olsen).

[1] One of the reviewers pointed out that Adams later abandoned this position in Adams (1998). To sidestep such exegetical issues we use the phrase ‘Adams' Thesis’ to denote the position attributed to him in the literature based on his earlier work.

http://dx.doi.org/10.1016/j.cognition.2015.12.017
0010-0277/

Table 1
De Finetti truth table.

A	C	If A, then C
⊤	⊤	⊤
⊤	⊥	⊥
⊥	⊤	*void*
⊥	⊥	*void*

The Equation: P(if A, then C) = P(C|A), but 'If A, then C' does not express a classical proposition (Bennett, 2003; Evans & Over, 2004; Oaksford & Chater, 2007).

the consequent should somehow be connected or relevant for one another as one aspect of the conditionals' meaning (for references see Douven, 2015b; Krzyżanowska, 2015; Skovgaard-Olsen, in press). This intuition is especially salient when we observe examples in which the relevance expectation is violated, as in conditionals such as 'If blood is red, then Oxford is in England', for which the truth-value of the antecedent leaves the truth-value of the consequent unaffected. However, surprisingly this intuitive idea is not preserved in any of the theories of conditionals currently endorsed in the psychology of reasoning, as we shall see.

1.1. The paradoxes of the material implication

Before the Equation became popular in the psychology of reasoning (Evans & Over, 2004; Oaksford & Chater, 2007; Pfeifer, 2013), the dominant theory was mental model theory, which is based on the material implication analysis of natural language conditionals (Johnson-Laird & Byrne, 2002).[2] Since the material implication is always true except in cases when its antecedent is true and its consequent is false, the theory validates the following argument-schemes that are known to give rise to nonsensical results once natural language content is substituted:

$$\frac{\neg A}{\therefore \text{ if } A, C} \qquad \frac{C}{\therefore \text{ if } A, C} \tag{1}$$

With no restrictions on the relationship between the antecedent and the consequent, *any* conditional could be inferred from a false antecedent or a true consequent, no matter how odd. Hence, from the true premise 'It is not the case that Europe has been ruled by France since Napoleon' the conditional 'If Europe has been ruled by France since Napoleon, then the sun emits light' could be inferred. And from the true premise 'The sun emits light', the conditional 'If Europe has been ruled by France since Napoleon, then the sun emits light', or indeed 'If Europe was liberated from occupation by Napoleon's France, then the sun emits light' could be inferred. Unsurprisingly, participants in psychological experiments tend to find such inferences odd as well (Pfeifer & Kleiter, 2011). Of course, this fact has not escaped the proponents of mental model theory. In accounting for the oddness of such inferences, they exploit the logical equivalence of the material implication with '¬A v C' and argue that the reason why we are reluctant to endorse the valid argument schemes in [1] is due to the problem with endorsing the following equally valid argument schemes:

$$\frac{\neg A}{\therefore \neg A \vee C} \qquad \frac{C}{\therefore \neg A \vee C} \tag{2}$$

Since more possibilities are excluded by the premises than by the conclusions in [2], information is lost in the conclusion, and according to Johnson-Laird and Byrne (2002; Byrne & Johnson-Laird, 2009) this is really the source of our intuitive problems with [1]. However, in the absence of a prior theoretical commitment to the logical equivalence of natural language conditionals to disjunctions, a much more straightforward diagnosis of the oddness of [1] runs as follows. The problem is not so much that *fewer* possibilities are excluded by the conclusion than by the premises, but rather that different conditions are imposed by the premises and the respective conclusions. The premises are silent on the relationship between A and C and impose conditions on a set of possible worlds by being factual propositions; the conclusions impose constraints on epistemic states (i.e., that A is epistemically relevant for C).

In contrast, the probabilistic approaches that are currently replacing the mental model theory under the heading 'the New Paradigm of Reasoning' endorse the Equation and reject [1]. Purportedly this is because the premises do not probabilistically constrain the conclusion when the latter is interpreted as a conditional probability as long as 0 < P(premise) < 1 (Bennett, 2003, p. 139; Evans & Over, 2004; Oaksford & Chater, 2007; Pfeifer & Kleiter, 2011). However, as argued in Skovgaard-Olsen (in press), it can be claimed that these theories reject [1] for the wrong reasons. The most obvious diagnosis of the oddness of [1] remains that no restrictions on the relevance of A for C are introduced by the premises, whereas indicative conditionals fit for the speech act of assertions seem to require A to be relevant for C. Yet these probabilistic approaches within the New Paradigm of Reasoning are unable to account for this. According to the latter, indicative conditionals should be seen as a linguistic device by which the participants activate a mental algorithm known as the Ramsey test, which consists in temporarily adding the antecedent to their knowledge-base and evaluating the consequent under its supposition (Evans & Over, 2004; Oaksford & Chater, 2007; Pfeifer, 2013). As such, indicative conditionals can have a high probability of being true as long as P(C) is high, even if the antecedent is irrelevant for the consequent. Accordingly, none of the main contenders in contemporary psychological accounts of conditional reasoning are willing to make relevance part of the core meaning of natural language conditionals.[3]

1.2. P(If A, then C) and relevance

The next surprise is that until quite recently,[4] when the role of relevance in the interpretation of conditionals was empirically investigated it was either found that no support could be provided (Oberauer, Weidenfeld, & Fischer, 2007; Singmann, Klauer, & Over, 2014), or that it was only weakly supported by the data (Over et al., 2007). So perhaps relevance should be set aside for our theories of conditionals after all. In these studies, relevance was operationalized in terms of the Δp rule, which is well-known from the psychological literature on causation, where $\Delta p > 0$ has been taken to be a necessary, but not sufficient, condition for inferring causality (Cheng, 1997).

The Δp Rule: $\Delta p = P(C|A) - P(C|\neg A)$

As P(C|A) is already occupied as a predictor of P(if A, then C) by the Equation, Over et al. (2007) and Singmann et al. (2014) try to obtain an orthogonal predictor for the relevance approach by using P(C|¬A). The evidence clearly favored P(C|A) as a predictor.

[2] As one of the reviewers pointed out, mental model theory has recently been revised so as to avoid being committed to the material implication analysis of the natural language conditional in Johnson-Laird, Khemlani, and Goodwin (2015). However, we here restrict our focus to the previous version of the theory.

[3] However, it should be noted that Over and Evans (2003) did entertain the possibility that relevance could characterize a subgroup of conditionals (i.e. causal conditionals). Yet this idea was later rejected in Over, Hadjichristidis, Evans, Handley, and Sloman (2007).

[4] An exception is Douven, Elqayam, Singmann, Over, and Wijnbergen-Huitink (in press). In this study it was found in a novel experimental task that the participants used clues about the inferential relations between A and C in evaluating the conditionals used in that task.

However, as Spohn (2013: 1092) observes, the Δp of the stimulus material used in Over et al. (2007) ranged from .23 to .32 and it would thus seem that a fairer test of the relevance approach would cover the whole spectrum of *positive relevance*, *irrelevance*, and *negative relevance*[5]:

Positive Relevance: $\Delta p > 0$
Irrelevance: $\Delta p = 0$
Negative relevance: $\Delta p < 0$

To be sure, Oberauer et al. (2007) did include $\Delta p = 0$ conditions. But in contrast to Over et al. (2007), they did not use realistic stimulus material that would enable the participants to form their own relevance expectations based on their background knowledge. Instead they supplied the participants with frequency information about a deck of cards relating properties in artificial relations. Accordingly, it is unclear whether a failure to take relevance into account in their study is due to: (a) the independence of the participants' assessment of P(if A, then C) with respect to relevance assessments or (b) the participants' failure to incorporate novel frequency information about artificial stimuli into their degrees of belief in conditionals. Hence, one goal of the present study was to use realistic stimuli that activate the participants' background knowledge while measuring P(if A, then C) across systematic manipulations along the relevance dimension.

A further issue is that it is not entirely obvious what the relationship between P(if A, then C) and relevance should be on the relevance approach to conditionals. There are more options available than simply considering Δp as a predictor of P(if A, then C). Indeed, in Douven (2015a) and Olsen (2014: chap. 3; see also Skovgaard-Olsen, 2015) it is suggested that an alternative could also provide a solution to the unsolved problem of where the participants' conditional probabilities come from, if we do not want to assume that they are calculated from unconditional probabilities using the Kolmogorov ratio definition (i.e., $P(C|A) = P(C\&A)/P(A)$). Reference to the Ramsey test can only be part of the solution, because it does not in itself tell us which psychological mechanisms are involved in determining P(C) once A has been added as a supposition to the participants' knowledge base, as recognized by Over et al. (2007):

> Explaining how the Ramsey test is actually implemented—by means of deduction, induction, heuristics, causal models, and other processes—is a major challenge, in our view, in the psychology of reasoning (p. 63).

Douven's (2015a) suggestion is that once A has been added to the knowledge base, assessments of the strength of arguments from A to C (given background knowledge) are used in determining P(C) in performing the Ramsey test. Olsen's (2014: chap. 3) suggestion is that heuristic assessments of the extent to which A is a predictor of C are used.

A third possibility, which we propose in this paper, is *the Default and Penalty Hypothesis*. The Default and Penalty Hypothesis holds that in evaluating either P(if A, then C) or Acc(if A, then C) the participants evaluate whether A is a sufficient reason for C. Applying the explication of the reason relation given in Spohn (2012: chap. 6), this requires at least two things: (i) evaluating whether positive relevance is fulfilled and (ii) evaluating P(C|A). The default assumption is that positive relevance is given, so the participants jump directly to evaluating P(C|A), which explains the existing evidence for the Equation. However, once the default assumption of positive relevance is violated, the violation of the participants' expectations will disrupt the equality between P(C|A) and both P (if A, then C) and Acc(if A, then C).

How exactly this disruption takes place is a matter for further research. Conceptually, the idea is that the negative surprise of the lack of positive relevance makes the participants apply a simple penalty to P(if A, then C) or Acc(if A, then C) (amounting to a main effect of the relevance condition). However, the discovery that A is not a reason for C may also lead the participants to rely less on P(C|A) (amounting to an interaction between the effect of P(C|A) and the relevance condition), since P(C|A) is used to assess the sufficiency.

The Default and Penalty Hypothesis thus describes a heuristic that the participants rely on when asking themselves, in one way or another, how plausible the indicative conditional is. It can be motivated by the observation that we use conditionals to display and discuss the inferential relations we are prepared to use in arguments (Brandom, 2010: 44–8, 104; Fogelin, 1967). Processing conditionals accordingly makes us expect an inferential relation to be displayed and so we expect there to be a relationship of epistemic relevance between A and C. However, this default assumption can, of course, be overridden. Perhaps one way of accounting for so-called non-interference conditionals, where there is no apparent connection between A and C, is to say that they are exactly like cases in which the context indicates that this default assumption is to be set aside (Skovgaard-Olsen, in press). That is to say, in these special cases relationships between sentences can be displayed that are so absurd in the first place that a rhetorical point is made either of the absurdity of the antecedent (e.g., 'If you can lift that, then I am a monkey's uncle') or of the fact that the consequent is endorsed come what may (e.g., 'If it snows in July, the government will fall'). However, such non-interference conditionals are the exception and the default assumption is one of the positive relevance of the antecedent for the consequent.

One virtue of the last two possibilities for relating P(If A, then C) to the relevance approach is that each offers an explanation for the substantial body of evidence found in favor of the Equation. According to Douven (2015a) and Olsen's (2014: chap. 3) suggestions, inferential relations and predictor relationships, respectively, play a role in determining P(C|A), which in turn is used in determining P(if A, then C). According to the Default and Penalty Hypothesis, upon processing the antecedent with realistic materials[6] and the conditional form, the participants will by default assume the positive relevance of the antecedent for the consequent. Hence, as long as we are primarily investigating positive relevance stimulus material then participants should jump directly to the second step and evaluate P(if A, then C) solely on the basis of P(C|A). But

[5] In these definitions we follow Spohn (2012: chap. 6, 2013). Importantly, this notion of relevance is different from the notion of relevance as introduced by Sperber and Wilson (1986) in that it does not attribute a role to processing costs as negatively correlated with perceived relevance. In Skovgaard-Olsen, Singmann, and Klauer (draft), there is further discussion of how they relate to other popular ideas in the psychological literature, such as the dual processing framework and Sperber and Wilson's relevance theory.

[6] As this indicates, we formulate the Default and Penalty Hypothesis in the first instance as a psychological hypothesis concerning the processing of realistic material implementing a more general relevance approach to conditionals. How the theory extends to the processing of abstract material is an open issue. But one point is clearly that as the stimulus material blocks the participants' ability to make relevance expectations on the basis of their background knowledge, they can only rely on the information provided by the experimenters. Accordingly, since false antecedent cases fail to supply the participants with useful information for assessing whether A is a sufficient reason for C, evaluation of the conditional under these circumstances may provoke a presupposition failure. In the literature on presupposition failures in general, it is thought that they either make the afflicted sentence false or truth-valueless (Heim & Kratzer, 1998: chap. 4, 6). Perhaps this accounts for the effect known as 'the defective truth table' in the literature, which has been taken as evidence for de Finetti truth table (Evans & Over, 2004). An alternative explanation is that participants inclined to a conjunctive reading treat A as a presupposition of the indicative conditional. (We owe parts of this argument to discussions with Seth Yalcin.)

as soon as this tacit assumption of positive relevance is violated, then we can experimentally distinguish the Default and Penalty Hypothesis from the Equation. One of the goals of the present study is to test this hypothesis.

1.3. Relevance and the core meaning of conditionals

As current probabilistic theories of reasoning do not make relevance, or inferential relations, part of the core meaning of conditionals, they have to treat it as a pragmatic component that is introduced by contextual factors. One option is to attribute the expectancy of the antecedent's relevance to the consequent to an implicature that arises due to Gricean norms of non-misleading discourse.[7] However, as Douven (2015a) points out, it is not entirely obvious how exactly the pragmatic mechanism is supposed to work.

Moreover, if this pragmatic explanation were true, we would only expect to find an effect of relevance manipulations on the *acceptability* of utterances in conversational contexts, as Gricean maxims *in the first instance* apply to conversational contexts.[8] In contrast, studies investigating P(If A, then C) have been used by proponents of the Ramsey test and the Equation to arbitrate between conflicting accounts of the core meaning of conditionals (Over & Evans, 2003). Hence, we should not expect to find a relevance effect on P(If A, then C), if the latter quantity is an indicator of semantic content and expectations of relevance are to be excluded from the semantic analysis. On the contrary, if relevance expectations are part of the core meaning of conditionals, then we should expect to find a relevance effect in both the P(if A, then C) and Acc(if A, then C) conditions.

It is more difficult to say how evidence of a relevance effect on P (If A, then C) would affect the mental model theory. This is because the theory has been formulated in such a way that the core meaning of conditionals is only investigated directly using abstract stimulus materials. When realistic stimulus materials are applied, the theory allows for both pragmatic and semantic modulation (Johnson-Laird & Byrne, 2002). In mental model theory, it is assumed that there are different levels of processing conditionals. In the most superficial mode, indicative conditionals are thought to be processed as conjunctions by constructing a mental model of the first cell of the truth table for the material implication, where both the antecedent and the consequent are true (while adding a mental footnote in the form of an ellipsis representing that there are further implicit models that would be consistent with the truth of the conditional, which distinguishes its mental-model representation from that of a conjunction):

A C
...

Pragmatic modulation is thought to occur when contextual factors modify the mental models constructed of the truth table cells in which the conditional is true. Semantic modulation occurs when the content of the antecedent and the consequent modifies the mental models constructed of the truth table cells in which the conditional is true. In both cases, this can take the form of adding information to the models, preventing the construction of models, or by aiding the participant in replacing the mental footnote with an explicit representation of all the cases in which the conditional is true.

However, as it stands the mental model theory has not been formulated in such a manner as to generate the general expectation that the antecedent should be epistemically relevant for the consequent once natural language content is used. As such, it too would be faced with the explanatory challenge of how to account for a relevance effect on P(if A, then C) without relying on *ad hoc* principles.

1.4. Concessive conditionals

As outlined above, previous studies that do not include a systematic comparison between positive relevance, negative relevance, and irrelevance based on realistic materials have not found a relevance effect on P(if A, then C). Yet Douven and Verbrugge (2012) did find that the categorical acceptance of indicative conditionals requires the antecedent to provide evidential support for the consequent as a necessary condition, in addition to high conditional probabilities. Moreover, they also found that the evidential support relation could be used to differentiate between the acceptability of indicative conditionals and concessive conditionals such as 'Even if it rains, then Michael will still go outside for a smoke'. Accordingly, a further goal of the present study is to investigate whether these findings can be extended from the case of categorical acceptability to quantitative degrees of acceptability.

In Skovgaard-Olsen (in press), it is suggested that concessive conditionals could be used to deny that A is a sufficient reason against C in contexts where there is a presupposition of A being a reason against C. This can occur either because the speaker denies that A is a *sufficient* reason against C or because the speaker denies that A is a *reason against* C – perhaps because A is taken to be *irrelevant* for or indeed to constitute a *reason for* C by the speaker. So whereas the Default and Penalty Hypothesis makes us predict that the participants will find indicative conditionals defective in the negative relevance and irrelevance conditions, in general there should be nothing defective about the use of concessive conditionals here. The exception, of course, is when P(C|A) = low and A is indeed a sufficient reason against C.

We can distinguish two versions of this hypothesis about concessive conditionals. In one version the concessive conditional simply expresses a denial of A as a good objection against C, if A were true (i.e., P(C|A) $\neq$ low). In another version, the concessive conditional expresses an unconditional commitment to C and the assumption that the degree of justification for C would be stable with respect to the truth of A (i.e., P(C) = high and P(C|A) $\neq$ low). On this latter proposal, the unconditional commitment to C distinguishes concessive conditionals from indicative conditionals. In the case of the indicative conditional, 'If A, then C', we adopt a *conditional* commitment to C under the supposition that A is true, *because* A is viewed as a sufficient reason for C. In the case of the concessive 'Even if A, then (still) C', we retain an *unconditional* commitment to C even if A is true, because it is denied that A is a good objection against C. For both of these proposals, it might seem a bit redundant to deny that A is a good objection against C when it is obvious that A is either irrelevant or positively relevant for C. But strictly speaking one would not be committing an epistemic error in doing so, because it is true after all that C would not be undermined by the truth of A, if A raises the probability of C ($\Delta p > 0$) or leaves it unchanged ($\Delta p = 0$).

Igor Douven (personal communication, September, 2015) notes in relation to the second version of the hypothesis about concessive conditionals that it may actually fit the non-interference conditionals we encountered above (e.g., 'If it snows in July, the government will fall') better than concessive conditionals. In the case of non-interference conditionals, substitution of 'whether or

[7] To be sure, Grice also has a maxim to the effect that one should make one's contributions to the conversation relevant (with an unspecified notion of relevance). However, it should be noted that the latter maxim applies to the level of whole speech acts, whereas when we talk about relevance in relation to conditionals, we are dealing with an internal relation between the antecedent and the consequent in one sentence.

[8] The reason for this qualification is that Douven (2010) has made a case for the claim that Gricean maxims of conversation not only apply to conversational contexts but also to individual reasoning.

not', 'regardless of whether', and sometimes 'even if' for 'if' makes little difference to their assertability (Douven, 2015b: 11). Accordingly, it should be possible to test whether the participants are interpreting a given concessive as a non-interference conditional by testing their sensitivity to such substitutions. In our experiment below, we will, however, only test the first version of the hypothesis about concessive conditionals, which is more clearly distinguished from non-interference conditionals.

2. The current experiment

A general lesson to take from the discussion above is that if we want to make progress on the issue of whether expectations of epistemic relevance should be included in the core meaning of indicative conditionals, then we should use realistic stimulus material that allows the participants to form expectations about relevance. We should then systematically violate those expectations through manipulations of relevance that also implement conditions of negative relevance and irrelevance. We did this in our experiment. As discussed above, we predict that in the case of indicative conditionals we find an effect of the relevance condition: for positive relevance (PO) the Equation holds, whereas for irrelevance (IR) and negative relevance (NE) it does not hold. In contrast, for concessive conditionals we predict no such effect. Here the Equation is expected to hold throughout, for all three relevance conditions.

3. Method

3.1. Participants

The experiment was conducted over the Internet to obtain a large and demographically diverse sample. A total of 577 people took part in the experiment. The participants were sampled through the Internet platform www.Crowdflower.com from the USA, the UK, and Australia and were paid a small amount of money for their participation.

The following exclusion criteria were used: not having English as native language, failing to answer two simple SAT comprehension questions correctly in a warm-up phase, completing the task in less than 160 s or in more than 3600 s, and answering 'not seriously at all' to the question of how seriously they would take their participation at the beginning of the study. The final sample consisted of 348 participants: 94 were assigned to the P(if A, then C) condition, 89 to the Acc(if A, then C) condition, 78 to the P(Even if A, then still C) condition, and 87 to the Acc(Even if A, then still C) group (see below). The mean age was 37.2 years, ranging from 17 to 72 years; 39.4% of the participants were male; 57.8% indicated that the highest level of education that they had completed was an undergraduate degree or higher.

3.2. Design

The experiment implemented a mixed design. There were two factors that were varied within participants: relevance (with three levels: PO, NE, IR), and priors (with four levels: HH, HL, LH, LL, meaning, for example, that P(A) = high and P(C) = low for HL). The prior manipulation had the goal of increasing the spread of the conditional probability of the consequent given the antecedent. This ensured a robust estimation of the relationship with the dependent variables. Two further factors were varied between participants, leading to the four experimental groups: conditionals, with two levels: indicative ('if A, then C') and concessive ('Even if A, then still C'); and mode of evaluation, with two levels: probability and acceptability.

3.3. Materials and procedure

Participants were randomly assigned to the four experimental groups. The 12 within-participants conditions were randomly assigned to 12 different scenarios for each participant. More specifically, we performed a large pre-study (Skovgaard-Olsen et al., draft) in which we measured prior probabilities and perceived relevance for a set of 18 scenarios from which we obtained the 12 different scenarios employed here. From each of the 12 selected scenarios we could construe all 12 within-participants conditions. Consequently, the mapping of conditions to scenarios was counterbalanced across participants, thereby preventing confounds of condition and content.

To reduce the dropout rate once the proper experiment had begun, participants first saw three pages stating our academic affiliations, asking for their email addresses (which were not paired with their responses, however), presenting two SAT comprehension questions in a warm-up phase, and posing a seriousness check about how careful the participants would be in their responses (Reips, 2002). Following this, the experiment began with the presentation of the 12 within-participants conditions. Their order was randomized anew for each participant.

For each of the 12 within-participants conditions, the participants were presented with three pages. The (randomly chosen) scenario text was placed at the top of each page. One participant might thus have seen the following scenario text:

> Sophia's scenario: Sophia wishes to find a nice present for her 13-year-old son, Tim, for Christmas. She is running on a tight budget, but she knows that Tim loves participating in live role-playing in the forest and she is really skilled at sewing the orc costumes he needs. Unfortunately, she will not be able to afford the leather parts that such costumes usually have, but she will still be able to make them look nice.

The underlying idea was to use brief scenario texts concerning basic causal, functional, or behavioral information that uniformly activates stereotypical assumptions about the relevance and prior probabilities of the antecedent and the consequent of 12 conditionals that implement our experimental conditions for each scenario. To introduce the 12 experimental conditions for the scenario text above we, *inter alia*, exploited the fact that the participants would assume that receiving things belonging to orc costumes would raise the probability of Tim being excited about his present (PO), receiving a Barbie doll would lower the probability of Tim being excited about his present (NE), and that whether Sophia regularly wears shoes would leave the probability of Tim being excited about his present unchanged (IR). A pretest with 495 participants reported in Skovgaard-Olsen et al. (draft) showed that the average Δp was .32 for the positive relevance conditions, $-.26$ for the negative relevance conditions, and .014 for our irrelevance conditions.

On the first page of each within-participants condition, the scenario text was followed by two questions presented in random order. One of those questions measured the conditional probability of the consequent given the antecedent, which is here illustrated for the NE–LH condition (=negative relevance, P(A) = low, P(C) = high) for the scenario text above:

> Suppose Sophia buys a Barbie doll for Tim.
> Under this assumption, how probable is it that the following sentence is true on a scale from 0 to 100%:
> Tim will be excited about his present.

The other question measured the probability of the conjunction of the antecedent and the consequent. We included this question to measure the probability of the premise of an inference task presented on the third page of the study. On the second page, the scenario text was either followed by a question asking the participants to evaluate P(if A, then C), Acc(If A, then C), P(Even if A, then still C), or Acc(Even if A, then still C), depending on which experimental group they were in:

> Could you please rate the probability that the following sentence is true on a scale from 0 to 100%:
>
> IF Sophia buys a Barbie doll for Tim, THEN Tim will be excited about his present.

If the participants were in one of the acceptability groups (i.e., Acc(If A, then C) or Acc(Even if A, then still C)), and it was their first scenario then they would first receive the following instruction:

> When we ask – here and throughout the study – how 'acceptable' a statement is, we are not interested in whether the statement is grammatically correct, unsurprising, or whether it would offend anybody. Rather we ask you to make a judgment about the adequacy of the information conveyed by the statement. More specifically, we ask you to judge whether the statement would be a reasonable thing to say in the context provided by the scenarios.

On the third page, the participants were presented with a short argument with the conditional as the conclusion. The results of that task are not reported here.

Thus, for each of the 12 within-participants conditions, each mapped to a different scenario, participants went through 3 pages. For each question, the participants were instructed to give their responses using sliders ranging from 0% to 100%. The full list of scenarios, the raw data, the data preparation script, and the analysis script can all be found in the supplemental materials at https://osf.io/j4swp/.

4. Results

Figs. 1 and 3 provide an overview of the data per mode of evaluation and relevance condition with the estimated conditional probability P(C|A) on the *x*-axis and the dependent variables (either P(if A, then C), Acc(if A, then C), P(Even if A, then still C), or Acc(Even if A, then still C)) on the *y*-axis (similar plots further divided as a function of prior manipulation are provided in the supplemental materials; they essentially show the same pattern, albeit with more noise). Regarding the statistical analysis it is important to note that the data has replicates on both the level of the participant (each participant provided one response for each of the twelve within-participants conditions; i.e., four responses per relevance condition) as well as on the level of the scenario (each scenario could appear in each relevance condition; we obtained between 19 and 41 responses for each scenario-by-relevance-condition combination across all four groups). This dependency structure, with conditions repeated within participants and scenarios, can be accommodated by a linear mixed model (LMM) analysis with crossed random effects for participants and scenarios (Baayen, Davidson, & Bates, 2008). Details of the model specification can be found in the Appendix A.

4.1. Indicative conditionals

Fig. 1 seems to support our first hypothesis; for indicative conditionals the relevance condition seemed to affect the results but the mode of evaluation (P(if A, then C) vs. Acc(if A, then C)) seemed to have little influence. In the PO condition the agreement between the conditional probability and the dependent variable seemed to be very strong. If it were not for some data point in the upper left corners the agreement would have been perfect and the regression line would have lain exactly on the main diagonal. However, in the

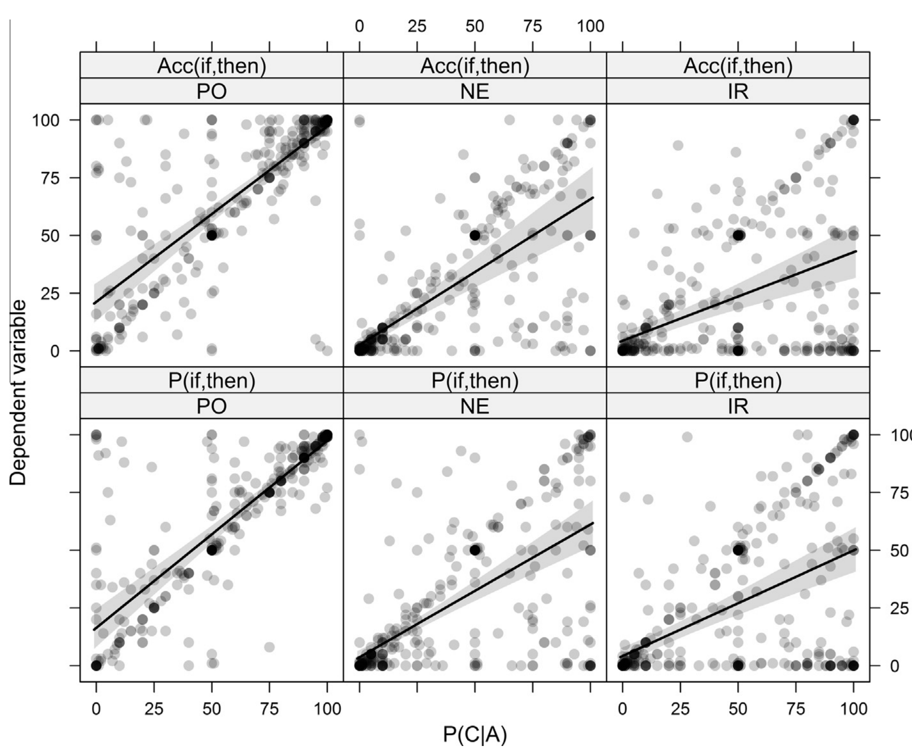

Fig. 1. Raw data values (plotted with 80% transparency) and LMM estimated linear effect of P(C|A) as a predictor on Acc(if A, then C) (upper row) and P(if A, then C) (lower row) across relevance manipulations (PO = left column, NE = center column, IR = right column). The confidence bands show the 95% confidence region of the effect of P(C|A).

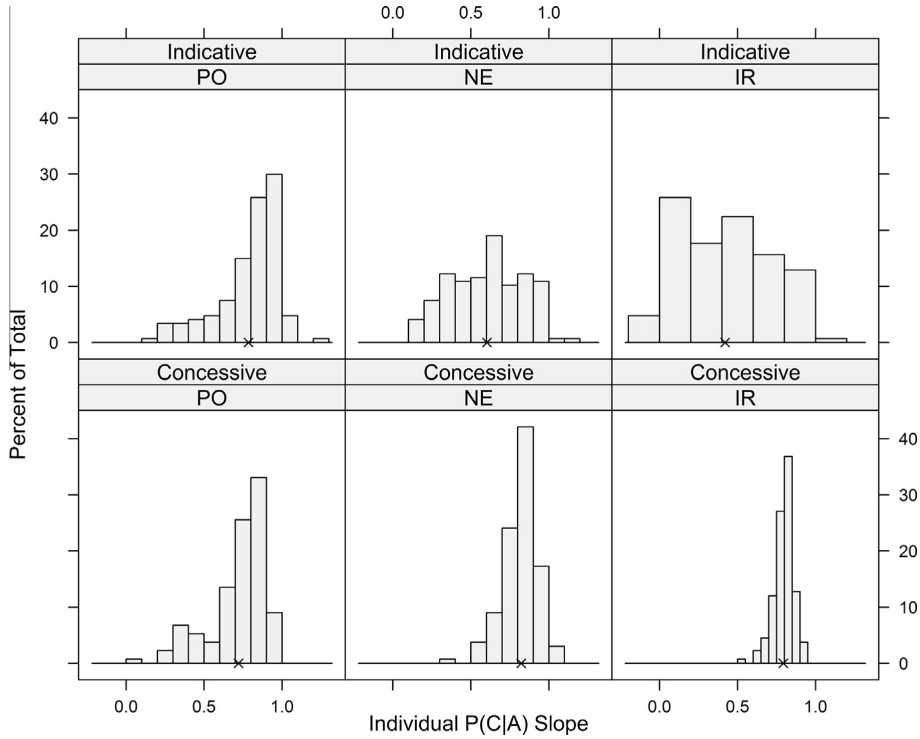

Fig. 2. Individual slope estimates for the effect of conditional probability P(C|A) on the dependent variable across conditions. These estimates are derived from the random effects terms of the LMM. In each plot each participant provided one slope estimate. The × denotes the fixed effects estimate.

other two conditions this relationship seemed much weaker, mainly because of a larger cluster of data points in the lower right corners. In addition, there seemed to be a difference in intercept, such that the overall level of responses to the dependent variable seemed to be considerably lower in the NE and IR condition compared to the PO condition.

This pattern was confirmed in an LMM analysis with fully crossed fixed effects for the conditional probability P(C|A), relevance condition (PO, NE, and IR), and modes of evaluation (P(if A, then C) and Acc(if A, then C)). Interestingly, this LMM showed no effects of the mode of evaluation, all $F < 1.5$, $p > .28$, indicating that the probability of the conditional was judged in exactly the same way as the acceptability of the conditional. We found a main effect of conditional probability, $F(1,33.45) = 505.16$, $p < .0001$, which was further qualified by an interaction between conditional probability and the relevance condition, $F(2,18.10) = 20.21$, $p < .0001$. Follow-up analysis on the interaction showed that the slope in the PO condition ($b = 0.78$, 95%-CI = [0.71, 0.86]) was significantly larger than the slope in the NE condition ($b = 0.60$, 95%-CI = [0.49, 0.72]), $t(20.05) = 3.27$, $p_H = .008$,[9] as well as significantly larger than the slope in the IR condition ($b = 0.42$, 95%-CI = [0.32, 0.52]), $t(22.08) = 6.00$, $p_H < .0001$. Additionally, the slopes from the NE and IR conditions also differed significantly, $t(12.26) = 2.36$, $p_H = .04$.

In other words, in the PO condition an increase in perceived conditional probability by 1% led to an increase of around 0.8% in the perceived probability or acceptability of the conditional – an almost perfect relationship. In the other conditions the same increase in perceived conditional probability led to a markedly lower increase in the perceived probability or acceptability of the conditional of 0.6% and 0.4%, respectively.

We also found a main effect of the relevance condition, $F(2,16.97) = 89.25$, $p < .0001$, indicating that the level of perceived probability or acceptability of the conditional differed across the three conditions. However, given the significant interaction with conditional probability, the pattern was slightly less simple. Across the conditional probability scale, PO conditionals received higher ratings than both NE and IR conditionals, $t > 4.49$, $p_H < .002$. However, while there was clearly no difference between NE and IR at the far left of the scale (i.e., at 0%), $t(10.36) = -0.71$, $p_H = .49$, this result was less clear at the midpoint of the scale (i.e., at 50%), $t(12.28) = 2.27$, $p_H = .10$ as well as at the far right end of the scale (i.e., at 100%), $t(12.30) = 2.37$, $p_H = .10$. The estimated marginal means [EMM] were PO = 18.8%, NE = 3.1%, and IR = 4.3% at 0%, PO = 58.1%, NE = 33.3%, and IR = 25.3% at 50%, and PO = 97.3%, NE = 63.4%, and IR = 46.3% at 100%.

Careful inspection of Fig. 1 suggests that there was a further difference between the three relevance conditions. The effect of conditional probability on the dependent variable seemed to be quite uniform across participants in the positive relevance condition. In contrast, in the other two conditions there seemed to be more inter-individual variability in the slope, some participants seemed to maintain a slope of one (i.e., their responses lay on the main diagonal) whereas other decreased the slope. This decrease seemed to be specifically strong in the irrelevance condition, where some slopes seemed to be at zero. To assess this hypothesis Fig. 2 (upper row) plots the distribution of individual slope estimates derived from the LMM. As can be seen, the distribution of conditional probability slopes in the PO condition clearly peaks, whereas the one in the IR condition is a lot flatter with (at least) one rather weak peak at 0. This is supported by the empirical standard deviations of the individual slopes estimates, 0.20 in the PO condition, 0.24 in the NE condition, and 0.29 in the IR condition, as well as by the empirical kurtosis,[10] 0.98 in the PO condition, −0.95 in the NE condition, and −1.11 in the IR condition. Ensuring the distribution of random

[9] We controlled for the family-wise error rate of follow-up tests, for each set of follow-up tests separately, using the Bonferroni-Holm correction (indicated by the index "H").

[10] "For symmetric unimodal distributions, positive kurtosis indicates heavy tails and peakedness relative to the normal distribution, whereas negative kurtosis indicates light tails and flatness" (DeCarlo, 1997, p. 292).

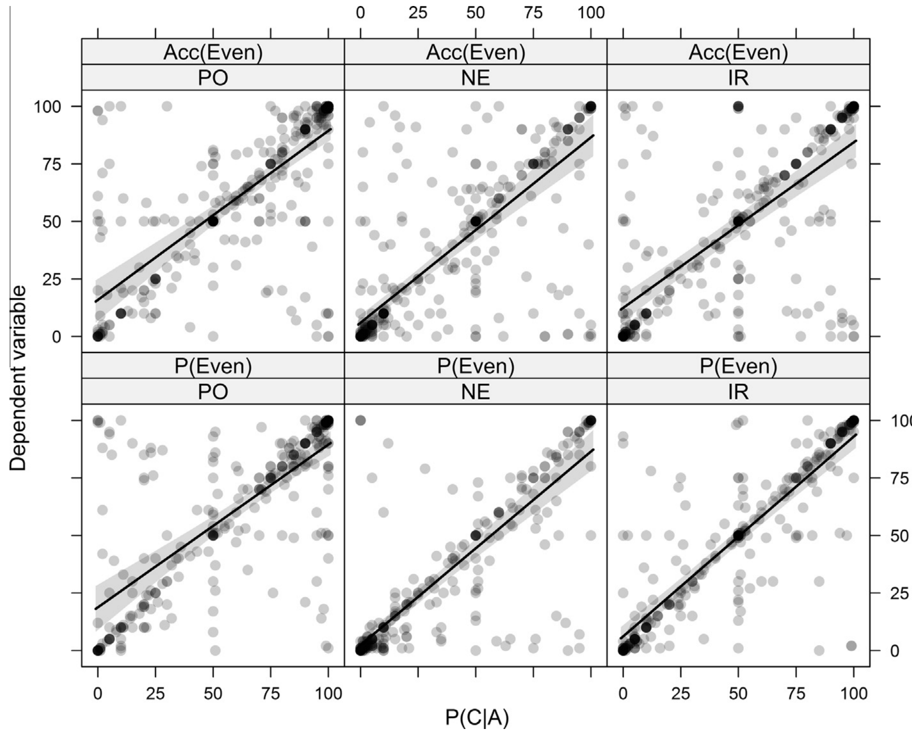

Fig. 3. Raw data values (plotted with 80% transparency) and LMM estimated linear effect of P(C|A) as a predictor on Acc(Even if A, then still C) (upper row) and P(Even if A, then still C) (lower row) across relevance manipulations (PO = left column, NE = center column, IR = right column). The confidence bands show the 95% confidence region of the effect of P(C|A).

effects is not an artifact of the hierarchical modeling approach; since it shrinks extreme estimates toward the mean estimate, we also estimated separate regressions for each individual and relevance condition. This analysis essentially showed the same pattern of results (see supplementary materials).

4.2. Concessive conditionals

Inspection of Fig. 3 suggests a more homogenous pattern for the concessive conditionals. With small exceptions in the PO conditions, the agreement between the conditional probability and the dependent variable was almost perfect. There seemed to be few other effects. This was supported by an LMM with the same structure as the indicative conditionals, which revealed a main effect of conditional probability, $F(1, 16.63) = 827.43$, $p < .0001$, but no interaction between conditional probability and the relevance condition, $F(2, 13.21) = 2.00$, $p = .17$. The overall effect of conditional probability was $b = 0.78$, 95%-CI [0.73, 0.83], again suggesting that an increase in conditional probability of 1% results in an increase in probability or acceptability of the concessive conditional of around 0.8%, but this time for all three relevance conditions. As before, there was no effect of mode of evaluation, all $F < 1.8$, $p > .22$. We also found a small main effect of relevance condition, $F(2, 11.79) = 4.06$, $p = .046$. Follow-up analysis revealed that (at the midpoint of the conditional probability scale) PO conditionals (EMM = 53.4%) received higher ratings than NE conditionals (EMM = 45.3%), $t(14.81) = 2.99$, $p_H = .03$. IR conditionals (EMM = 49.0%), however, differed from neither of the other conditions, $|t| < 1.8$, $p_H > .21$.[11] The distribution of individual conditional probability effects (Fig. 2, lower row) shows clearly peaked distribution with the largest variability for PO (SD = 0.18), followed by NE (SD = 0.11), and IR (SD = 0.06) and positive kurtosis in each case (PO = 0.82; NE = 1.22; IR = 1.80).

5. Discussion

As we saw in the introduction, earlier studies that did not systematically contrast PO, NE, and IR stimulus material failed to find a relevance effect on indicative conditionals. In contrast, we introduced this manipulation and our results indicate that relevance affects the rating of the probability and the acceptability of indicative conditionals. These findings corroborate the predictions of the Default and Penalty Hypothesis that participants make a default assumption of positive relevance when processing the antecedent and the 'If..., then...' form with realistic material. As long as this assumption is fulfilled, reasoners proceed to the second step of the evaluation of whether A is a sufficient reason for C by evaluating whether P(C|A) = high. Yet, when the participants' expectations of positive relevance are violated in the irrelevance or negative relevance conditions, they react to the perceived defect of the indicatives by providing lower ratings and by showing less sensitivity to P(C|A). The analysis of the random effects estimates of the LMM reveals, however, that there is quite some individual variability present in the interaction between relevance and P(C|A). As shown in Fig. 2 (upper row), there appears to be a minority for whom P(C|A) continues to have a steep slope even in the irrelevance condition for indicative conditionals while P(C|A) has either a weak relationship, or no relationship at all, for the remaining participants in the same condition.

Both Adams' Thesis and the Equation are challenged by these findings. To account for the effects, proponents of the Equation would have to attribute the relevance effect to pragmatic modulation.[12]

[11] When analyzing the effect of relevance condition at each end of the scale separately, we found that at the left end (i.e., 0%), PO conditionals received higher ratings than NE ($p_H = .04$) and IR differed from neither ($p_H > .38$), whereas there was no difference between the three conditions at the right end of the scale (i.e., 100%), all $p_H > .99$. Note also that the interaction between conditional probability and relevance was not significant. Hence, not a lot of weight should be attached to the differences of effects at different scale positions.

[12] It could be argued that this strategy is not available to proponents of Adams' Thesis. After all, if Gricean maxims are required to account for a relevance effect on our judgments of the acceptability of indicatives, then presumably they should also explicitly enter into the theory for it to be descriptive of acceptability conditions.

For the Acc(If A, then C) group this might be accomplished by invoking Gricean maxims of conversation, since the instructions explicitly introduced a conversational context. Yet the same strategy cannot be applied to account for a relevance effect in the P(If A, then C) group – unless it is assumed that pragmatic factors are implicitly infused in the experimental task.

However, adopting this latter interpretation would put proponents of the Equation in a somewhat odd dialectical position. On the one hand, studies investigating P(If A, then C) have been used as direct evidence against the mental model theory to show that it got the core meaning of natural language conditionals wrong, since the dominant response is P(If A, then C) = P(C|A) (Evans & Over, 2004; Over & Evans, 2003). On the other hand, this fictive opponent would now insist that the same type of task, which was once used to arbitrate in disputes over the core meaning of conditionals, can no longer be interpreted as an investigation of semantic content in the absence of pragmatic factors now that a relevance effect has been found. Of course, the immediate problem would then be to explain what can prevent proponents of the mental model theory from using the same dialectical strategy to account for the evidence for P(If A, then C) = P(C|A) by claiming that it arose merely due to an infusion of pragmatic factors into the core meaning of the natural language conditional, which continues to be given by the material implication. Furthermore, we may note that if this strategy is to avoid having the appearance of an *ad hoc* attempt to dodge an unpleasant objection, then the burden of justification is on those engaging in this line of defense to produce positive experimental evidence that this is indeed what is happening in the concrete task under the relevance manipulations.

To be sure, finding a theory-neutral way of operationalizing the distinction between semantic and pragmatic content continues to be a vexing problem. Part of the reason is that the distinction persists in being deeply controversial in both the philosophical and linguistic literature on purely theoretical grounds (Bach, 1997; Birner, 2013: chap. 3). However, until this theoretical dispute is resolved, we propose to adopt the following strategy: to interpret our results as minimally raising an explanatory challenge to proponents of the Equation. If investigations of P(If A, then C) can be used to challenge and replace one theory of the core meaning of conditionals, then it seems legitimate to use the same task for documenting a relevance effect on the core meaning of conditionals.

As we have noted, the dialectical situation is somewhat different when it comes to the mental model theory, insofar as it holds that the core meaning of conditionals is to be investigated through the use of abstract stimulus materials. (In contrast, the Equation has been defended on the basis of tasks that use abstract and realistic materials interchangeably, Over & Evans, 2003.) However, the mental model theory is likewise faced with the explanatory challenge of showing how semantic modulation gives rise to the general expectation of positive, epistemic relevance of the antecedent for the consequent based on systematic principles, since its preferred account of the core meaning of conditionals ignores relevance considerations altogether.

Turning to our results concerning the concessive conditional, it is somewhat ironic that the present results indicate that P(C|A) is actually a better predictor of P(Even if A, then still C) than of P(If A, then C) across relevance manipulations. This is ironic since the Equation and the Ramsey Test were offered as explanatory hypotheses specifically concerning indicative conditionals that were silent on concessive conditionals. In contrast, it was predicted by our account that there would be a defect affecting indicative conditionals but not concessive conditionals in the irrelevance and negative relevance conditions. The reason we gave was that concessive conditionals deny that A is a good objection against C. For the negative relevance condition this requires that P(C|A) does not have a low probability. So here this qualitative analysis coincides with an account that uses P(C|A) as a predictor of the probability or acceptability of 'Even if A, then still C'. A little surprising, however, is that the acceptability ratings of the concessive were still high in the positive relevance condition, since denying that A is a good objection *against* C seems to be a bit redundant when A is actually a reason *for* C. Indeed, Douven and Verbrugge (2012: 485) think that a categorical acceptance of the concessive is positively odd for the positive relevance condition. But strictly speaking, the denial continues to be accurate under this condition, since A is not a good objection against C, if A in fact raises the probability of C.

In their investigation into the categorical acceptance of indicative and concessive conditionals, Douven and Verbrugge (2012) found that there was a tendency to accept the indicative conditional in positive relevance conversational contexts, whereas there was a tendency to accept the concessive conditional in negative relevance or irrelevance conversational contexts. In contrast, we found little difference between the degree of Acc(Even if A, then still C) across PO, NE, and IR. However, in comparing these results it must be kept in mind that Douven and Verbrugge (2012) were investigating comparative judgments of the *categorical* acceptance of indicative conditionals versus concessive conditionals across relevance manipulations, whereas we are making between-subject comparisons of *the absolute degrees of* acceptance of indicative conditionals and concessive conditionals across relevance manipulations. So while in their study only a small group of participants accepting the concessive in the PO condition, this may simply have been because their participants had the choice to select the indicative instead. In contrast, our study involved a between-groups comparison. So we did not ask for comparative judgments between concessives and indicatives. This might explain why there is a difference in how acceptable the concessive was found to be in the PO condition across the two studies.

Douven and Verbrugge (2010) report differences between the acceptability and probability of indicative conditionals when contrasting conditionals with inductive, abductive, and deductive inferential relations between the antecedents and consequents. In light of these findings, it is somewhat surprising that no differences between the probability and acceptability of both the indicative and concessive conditionals were found in our experiment. However, as Douven (personal communication, November, 2015) points out, one explanation might be that the differences were most marked for conditionals expressing inductive relations (where the connection is based on purely frequentist information), whereas the positive relevance conditionals we investigated seemed to have a more abductive character. Their findings suggest that the type of inferential relation may exercise an influence on the assessment of the reason relationship. But as Douven and Verbrugge (2012) readily admit, it is difficult to formulate a deeper understanding of their findings in the absence of a more detailed account of the processing of deductive, inductive, and abductive inferential relations.

One final thing that may seem surprising about our results is that the Equation only seemed to hold in the positive relevance case. Yet, as Igor Douven (personal communication, September, 2015) points out, the triviality results seem to show that P(If A, then C) = P(C|A) entails the probabilistic independence of the antecedent and the consequent. So it may seem surprising that in our experiment we found that the Equation only holds when the antecedent is probabilistically dependent on the consequent. However, as Douven and Verbrugge (2013) point out, the triviality results actually rely on the following assumption, which is stronger than the Equation:

$$\textsc{Generalized Equation}: \ \mathrm{P}(\text{if } \varphi, \text{ then } \psi|\chi) = \mathrm{P}(\psi|\varphi, \chi),$$
for any ψ, φ, χ such that $\mathrm{P}\ (\varphi, \chi) > 0$

And in their experiments, Douven and Verbrugge (2013) found evidence that this stronger assumption fails to hold for normal conditionals. Whether this is the right explanation is a subject for further research.

6. Conclusion

More than a decade of research has offered the Equation (P(If A, then C) = P(C|A)) strong empirical support. Moreover, not only do the prevalent theories in the psychology of reasoning not make the expectation of relevance of the antecedent for the consequent part of the core meaning of conditionals, but previous studies also appear to suggest that the presence of such an expectation is not supported by the data.

In the present study, results were presented that challenge this consensus by showing a relevance effect on P(If A, then C). This raises an explanatory challenge for psychological theories of conditionals like the recent probabilistic theories and the mental model theory, which deny that relevance plays a role in the core meaning of indicative conditionals. Moreover, it was found that P(C|A) actually provides a better predictor of the probability and acceptability of concessive conditionals than for indicative conditionals across relevance manipulations. This new finding is also surprising given that the probabilistic theories use P(C|A) as their main predictor for indicative conditionals, but have so far been silent on concessive conditionals.

Acknowledgements

We are very grateful to Wolfgang Spohn, Igor Douven, David Over, Nicole Cruz, Shira Elqayam, Seth Yalcin, and the audiences of talks at 'Gießener Abendgespräche Kognition und Gehirn' (University of Gießen), 'What if' (University of Konstanz), and 'Working Group in the History and Philosophy of Logic, Mathematics, and Science' (UC Berkeley) for valuable discussions. Moreover, careful comments from the reviewers of *Cognition* substantially improved the manuscript.

Appendix A. Details of the LMM analysis

All analyses were performed using lme4 (Bates, Maechler, Bolker, & Walker, 2015) for the statistical programming language R (R Core Team, 2015). To ease interpretation of the numerical covariate P(C|A), we centered it and the dependent variable on the midpoint of the scale (at 50%). For numerical stability in the estimation we also divided both by 100 so that all variables were on the scale from -1 to 1 (as factors were coded with 1 and -1). We followed the suggestions of Barr, Levy, Scheepers, and Tily (2013) and employed the maximal random effects structure as detailed below. Tests of fixed effects were Wald-tests using the Kenward-Roger approximation for degrees of freedom (Halekoh & Højsgaard, 2014). Follow-up analyses were based on the methods implemented in lsmeans (Lenth, 2015) and also employed the Kenward-Roger approximation for deriving standard errors and degrees of freedom.

Each of the LMMs (one for the indicative and one for the concessive conditionals) had crossed random effects for participants and scenarios. For participants, we estimated random intercepts as well as by-participant random slopes for P(C|A), the relevance condition, and their interaction. For scenarios we also estimated random intercepts as well as by-scenario random slopes for P(C|A), the relevance condition, the mode of evaluation, as well as all corresponding interactions. We also estimated all correlations among random effects for both the by-participant random effects as well as the by-scenarios random effects. Note that we did not estimate random slopes for the prior manipulation for either random effects term. This followed the consideration that the prior was only manipulated to achieve a certain spread of conditional probabilities in each relevance manipulation. Furthermore, including random slopes for the priors would have prevented us from estimating random slopes for the conditional probabilities in each relevance condition (as such a model would have been oversaturated), which were of substantive interest in the present study (see Fig. 2). For an analysis of the effect of the priors see the supplementary materials.

Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://osf.io/j4swp/.

References

Adams, E. (1998). *A primer of probability logic*. Stanford, CA: CLSI Publications.
Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. http://dx.doi.org/10.1016/j.jml.2007.12.005.
Bach, K. (1997). The semantics-pragmatics distinction: What it is and why it matters. *Linguistiche Berichte, 8*, 33–50. http://dx.doi.org/10.1007/978-3-663-11116-0_3.
Baratgin, J., Politzer, G., & Over, D. E. (2013). Uncertainty and the de Finetti tables. *Thinking & Reasoning, 19*(3), 308–328. http://dx.doi.org/10.1080/13546783.2013.809018.
Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language, 68*, 255–278. http://dx.doi.org/10.1016/j.jml.2012.11.001.
Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* (in press). arXiv:1406.5823.
Bennett, J. (2003). *A philosophical guide to conditionals*. Oxford: Oxford University Press.
Birner, B. J. (2013). *Introduction to pragmatics*. Malden, MA: Wiley-Blackwell.
Brandom, R. (2010). *Between saying & doing. Towards an analytic pragmatism*. Oxford: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199542871.001.0001.
Byrne, R. M. J., & Johnson-Laird, P. N. (2009). 'If' and the problems of conditional reasoning. *Trends in Cognitive Sciences, 13*(7), 282–287. http://dx.doi.org/10.1016/j.tics.2009.08.003.
Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104*(2), 367–405. http://dx.doi.org/10.1037//0033-295X.104.2.367.
DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods, 2*(3), 292–307. http://dx.doi.org/10.1037/1082-989X.2.3.292.
Douven, I. (2010). The pragmatics of belief. *Journal of Pragmatics, 42*(1), 35–47. http://dx.doi.org/10.1016/j.pragma.2009.05.025.
Douven, I. (2015a). How to account for the oddness of missing-link conditionals. *Synthese*, 1–14. http://dx.doi.org/10.1007/s11229-015-0756-7.
Douven, I. (2015b). *The epistemology of indicative conditionals. Formal and empirical approaches*. Cambridge: Cambridge University Press.
Douven, I., Elqayam, S., Singmann, H., Over, D., & Wijnbergen-Huitink, J. V. (2015). Conditionals and inferential connections (in press).
Douven, I., & Verbrugge, S. (2010). The adams family. *Cognition, 117*(3), 302–318. http://dx.doi.org/10.1016/j.cognition.2010.08.015.
Douven, I., & Verbrugge, S. (2012). Indicatives, concessives, and evidential support. *Thinking & Reasoning, 18*(4), 480–499. http://dx.doi.org/10.1080/13546783.2012.716009.
Douven, I., & Verbrugge, S. (2013). The probabilities of conditionals revisited. *Cognitive Science, 37*(4). http://dx.doi.org/10.1111/cogs.12025.
Evans, J. St. B. T., & Over, D. (2004). *If*. Oxford: Oxford University Press.
Fogelin, R. J. (1967). Inferential constructions. *American Philosophical Quarterly, 4*(1), 15–27.
Halekoh, U., & Højsgaard, S. (2014). A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models – The R package pbkrtest. *Journal of Statistical Software, 59*(9), 1–32.
Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar*. Malden & Oxford: Blackwell Publishers.
Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review, 109*, 646–678. http://dx.doi.org/10.1037//0033-295X.109.4.646.
Johnson-Laird, P. N., Khemlani, S. S., & Goodwin, G. P. (2015). Logic, probability, and human reasoning. *Trends in Cognitive Science, 19*, 201–214. http://dx.doi.org/10.1016/j.tics.2015.02.006.

Krzyżanowska, K. (2015). *Between "If" and "Then": Towards an empirically informed philosophy of conditionals*. PhD dissertation, Groningen University. <http://karolinakrzyzanowska.com/pdfs/krzyzanowska-phd-final.pdf>.
Lenth, R. V. (2015). *lsmeans: Least-squares means*. R package version 2.18. <http://CRAN.R-project.org/package=lsmeans>.
Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford: 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. http://dx.doi.org/10.1080/13546780601035794.
Olsen, N. S. (2014). *Making ranking theory useful for psychology of reasoning*. PhD dissertation, University of Konstanz. <http://kops.uni-konstanz.de/handle/123456789/29353>.
Over, D. E., & Evans, J. St. B. T. (2003). The probability of conditionals: The psychological evidence. *Mind & Language, 18*(4), 340–358. http://dx.doi.org/10.1111/1468-0017.00231.
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*(1), 62–97. http://dx.doi.org/10.1016/j.cogpsych.2006.05.002.
Pfeifer, N. (2013). The new psychology of reasoning: A mental probability logical perspective. *Thinking & Reasoning, 19*(3–4), 329–345. http://dx.doi.org/10.1080/13546783.2013.838189.
Pfeifer, N., & Kleiter, G. D. (2011). Uncertain deductive reasoning. In K. Manktelow, D. E. Over, & S. Elqayam (Eds.), *The science of reason: A Festschrift for Jonathan St. B.T. Evans* (pp. 145–166). Hove: Psychology Press.
Politzer, G., Over, D., & Baratgin, J. (2010). Betting on conditionals. *Thinking & Reasoning, 16*(3), 172–197. http://dx.doi.org/10.1080/13546783.2010.504581.
R Core Team (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org/>.
Reips, U. D. (2002). Standards for internet-based experimenting. *Experimental Psychology, 49*(4), 243–256. http://dx.doi.org/10.1027//1618-3169.49.4.243.
Singmann, H., Klauer, K. C., & Over, D. (2014). New normative standards of conditional reasoning and the dual-source model. *Frontiers in Psychology, 5*, 316. http://dx.doi.org/10.3389/fpsyg.2014.00316.
Skovgaard-Olsen, N. (2015). Ranking theory and conditional reasoning. *Cognitive Science*. http://dx.doi.org/10.1111/cogs.12267.
Skovgaard-Olsen, N., 2015. Motivating the relevance approach to conditionals. *Mind & Language* (in press).
Skovgaard-Olsen, N., Singmann, H., & Klauer, K. C. (draft). *Relevance and reason relations*.
Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Oxford: Blackwell.
Spohn, W. (2012). *The laws of beliefs*. Oxford: Oxford University Press.
Spohn, W. (2013). A ranking-theoretic approach to conditionals. *Cognitive Science, 37*, 1074–1106. http://dx.doi.org/10.1111/cogs.12057.
Woods, M. (1997). *Conditionals*. Oxford: Oxford University Press (Edited by Wiggins, D., Commentary by: Edgington, D.).