

Putting Inferentialism and the Suppositional Theory of Conditionals to the Test

Inaugural-Dissertation
zur
Erlangung der Doktorwürde
der Wirtschafts- und Verhaltenswissenschaftlichen Fakultät
der Albert-Ludwigs-Universität Freiburg. i. Br.

vorgelegt von
Niels Skovgaard Olsen
geboren am 30.07.1985 in Køge, DK

SS 2017

Tag der mündlichen Prüfung:

22. September 2017

Referenten:

Prof. Dr. Karl Christoph Klauer

Prof. Dr. Andrea Kiesel

PD Dr. Kerstin Dittrich

TABLE OF CONTENTS

Acknowledgement.....	iii
Summary.....	v
Zusammenfassung.....	vii
1. Introduction.....	1
The Suppositional Theory of Conditionals.....	2
The Psychological Evidence.....	5
Inferentialism.....	11
Putting the Two Theories to the Test.....	11
References.....	13
2. The Relevance Effect and Conditionals.....	17
Introduction.....	19
The Current Experiment.....	30
Discussion.....	41
Conclusion.....	46
References.....	46
Appendix.....	51
Supplementary Materials.....	52
3. Relevance and Reason Relations.....	61
Introduction.....	63
Experiment 1.....	64
Experiment 2.....	69
General Discussion.....	76
References.....	79
Supplementary Materials.....	82
4. Relevance Differently Affects the Truth, Acceptability, and Probability Evaluations of ‘And’, ‘But’, ‘Therefore’, and ‘If Then’.....	87
Introduction.....	89
Experiment 1.....	99
Experiment 2.....	113
General Discussion.....	122

References	127
Appendix.....	132
5. General Discussion	135
Relevance and Conditionals: A Synopsis	136
Semantic and Pragmatic Factors.....	139
Argument for Semantic Defect	147
References	150
6. Curriculum Vitae	153
7. Eidstattliche Erklärung	159

ACKNOWLEDGEMENT

A number of people deserve my complete gratitude for the completion of this dissertation. First of all, I thank Wolfgang Spohn for his generousness in allowing me to complete a second PhD while working as a PostDoc in philosophy on a grant given to him within the DFG Scherpunktprogramm, *New Frameworks of Rationality*. Second, I am very grateful for the open-mindedness of Karl Christoph Klauer in accepting an unexperienced philosopher to work empirically in his lab and for all the great advice and knowledge that he has shared in the process. I am also very grateful for the fruitful collaboration with Henrik Singmann and David Kellen – both of whom have taught me invaluable data-analyses skills and contributed with great friendship. Thirdly, I want to thank my colleagues at the Department of Socialpsychology and Methodology at the University of Freiburg for welcoming me into their working group, and to the scientific assistants (esp. Hannes Krahl) for helping me out with the hard work of conducting experiments. Fourthly, I want to thank my great colleagues at *New Frameworks of Rationality* for collaborations and good times. A special thanks goes to Markus Knauff, who has always been supportive of my research throughout the years. Fifthly, I want to thank John MarFarlane and Ulrike Hahn for productive and generous research visits at Berkeley University, Munich Center for Mathematical Philosophy, and Birkbeck University. Sixthly, I want to thank Igor Douven, Karolina Krzyżanowska, David Over, Shira Elqayam, Peter Collins, and Nicole Cruz for many good discussions over the years. But the biggest thanks goes to my parents, Niels Viggo Skovgaard Olsen and Anne Grete Skovgaard Olsen, and my brother, Jens Skovgaard Svane – to whom this dissertation is dedicated. Without their support and love I would never have been so audacious as to attempt a second dissertation within a new discipline.

SUMMARY

This dissertation is devoted to empirically contrasting the Suppositional Theory of conditionals, which holds that indicative conditionals serve the purpose of engaging in hypothetical thought, and Inferentialism, which holds that indicative conditionals express reason relations. Throughout a series of experiments, probabilistic and truth-conditional variants of Inferentialism are investigated using new stimulus materials, which manipulate previously overlooked relevance conditions. These studies are some of the first published studies to directly investigate the central claims of Inferentialism empirically.

In contrast, the Suppositional Theory of conditionals has an impressive track record through more than a decade of intensive testing. The evidence for the Suppositional Theory encompasses three sources. Firstly, direct investigations of the probability of indicative conditionals, which substantiate “the Equation” ($P(\text{if } A, \text{ then } C) = P(C|A)$). Secondly, the pattern of results known as “the defective truth table” effect, which corroborates the de Finetti truth table. And thirdly, indirect evidence from the uncertain and-to-if inference task.

Through four studies each of these sources of evidence are scrutinized anew under the application of novel stimulus materials that factorially combine all permutations of prior and relevance levels of two conjoined sentences. The results indicate that the Equation only holds under positive relevance ($P(C|A) - P(C|\neg A) > 0$) for indicative conditionals. In the case of irrelevance ($P(C|A) - P(C|\neg A) = 0$), or negative relevance ($P(C|A) - P(C|\neg A) < 0$), the strong relationship between $P(\text{if } A, \text{ then } C)$ and $P(C|A)$ is disrupted. This finding suggests that participants tend to view natural language conditionals as defective under irrelevance and negative relevance (Chapter 2). Furthermore, most of the participants turn out only to be probabilistically coherent

above chance levels for the uncertain and-to-if inference in the positive relevance condition, when applying the Equation (Chapter 3). Finally, the results on the truth table task indicate that the de Finetti truth table is at most descriptive for about a third of the participants (Chapter 4).

Conversely, strong evidence for a probabilistic implementation of Inferentialism could be obtained from assessments of $P(\text{if } A, \text{ then } C)$ across relevance levels (Chapter 2) and the participants' performance on the uncertain-and-to-if inference task (Chapter 3). Yet the results from the truth table task suggest that these findings could not be extended to truth-conditional Inferentialism (Chapter 4). On the contrary, strong dissociations could be found between the presence of an effect of the reason relation reading on the probability and acceptability evaluations of indicative conditionals (and connate sentences), and the lack of an effect of the reason relation reading on the truth evaluation of the same sentences. A bird's eye view on these surprising results is taken in the final chapter and it is discussed which perspectives these results open up for future research.

ZUSAMMENFASSUNG

Diese Abhandlung widmet sich der empirischen Gegenüberstellung der suppositionellen Theorie der Konditionalsätze und dem Inferentialismus. Laut der suppositionellen Theorie liegt die Hauptfunktion indikativer Konditionalsätzen darin, hypothetischem Denken zu veranlassen, während der Inferentialismus behauptet, dass sie Begründungsrelationen ausdrücken. Anhand einer Reihe von Experimenten werden probabilistische und wahrheitsbasierte Varianten des Inferentialismus mit neuen Stimulusmaterialien untersucht, die zuvor übersehene Relevanzbedingungen manipulieren. Die in dieser Abhandlung vorgestellten Studien gehören zu den ersten Veröffentlichungen, die die zentralen Aussagen des Inferentialismus auf direkte Weise empirisch überprüfen.

Im Gegensatz dazu weist die suppositionelle Theorie durch mehr als ein Jahrzehnt intensiver Tests eine beeindruckende Erfolgsbilanz auf. Die Befunde für die suppositionelle Theorie setzten sich aus drei Quellen zusammen. Erstens, direkte Untersuchungen der Wahrscheinlichkeit indikativer Konditionalsätze, welche die Gültigkeit von „the Equation“ ($P(\text{wenn } A, \text{ dann } C) = P(C|A)$) bestätigen. Zweitens, ein Ergebnismuster, das als „defekte Wahrheitstabelle“ bezeichnet wird und die de Finetti Wahrheitstabelle belegt. Drittens, indirektes Beweismaterial durch ungewisse „and-to-if“ Schlussfolgerungen.

Durch vier Studien wird jede dieser Beweisquellen erneut überprüft. Dazu werden neue Stimulusmaterialien angewandt, die alle Permutationen von A-priori-Wahrscheinlichkeiten und Relevanzbedingungen zweier miteinander verbundener Sätze faktoriell kombinieren. Die Ergebnisse weisen darauf hin, dass „the Equation“ für indikative Konditionalsätze lediglich unter positiver Relevanz ($P(C|A) - P(C|\neg A) > 0$) gilt. Im Fall von Irrelevanz ($P(C|A) - P(C|\neg A) = 0$) oder negativer Relevanz ($P(C|A) - P(C|\neg A)$)

< 0) wird die starke Beziehung zwischen $P(\text{wenn } A, \text{ dann } C)$ und $P(C|A)$ unterbrochen. Der Grund dafür dürfte sein, dass TeilnehmerInnen dazu tendieren, Konditionalsätze in natürlicher Sprache unter diesen Bedingungen als defekt zu betrachten (2. Kapitel). Weiterhin erweisen sich nur bei der positiven Relevanz-bedingung die meisten TeilnehmerInnen als probabilistisch kohärenter als das Zufallsniveau für ungewisse „and-to-if“ Schlussfolgerungen, wenn die „Equation“ als gültig vorausgesetzt wird (3. Kapitel). Letztendlich deuten die Ergebnisse der Wahrheitstabellen-Aufgaben darauf hin, dass die de Finetti Wahrheitstabelle die Antworten von höchstens ein Drittel der Teilnehmer beschreibt (4. Kapitel).

Andererseits lassen sich überzeugende Belege für die probabilistische Variante des Inferentialismus sowohl der Bewertung von $P(\text{wenn } A, \text{ dann } C)$ unter Relevanzbedingungen (2. Kapitel) als auch der Performanz der Teilnehmer in der ungewissen „and-to-if“ Schlussfolgerungen unter Relevanzbedingungen (3. Kapitel) erbringen. Im Gegensatz dazu sprechen die Ergebnisse der Wahrheitstabellen-Aufgabe eher gegen die wahrheitsbasierte Variante des Inferentialismus (4. Kapitel). Vielmehr deutet das Ergebnismuster darauf hin, dass starke Dissoziationen zwischen dem Einfluss von Relevanzbedingungen auf Einschätzungen von Wahrscheinlichkeiten und dem fehlenden Einfluss derselben auf Wahrheitsbeurteilungen gefunden werden können. Während der wahrgenommene Begründungszusammenhang von Antezedenz und Konsequent die Zuschreibung von Wahrscheinlichkeiten indikativer Konditionalsätze (sowie verwandter Sätze) stark beeinflusst, wurde kaum Einfluss auf Wahrheitsbeurteilungen derselben Sätze gefunden. Im abschließenden Kapitel werden diese überraschenden Ergebnisse von einem höheren Blickwinkel aus betrachtet und mögliche Erweiterungen für zukünftige Forschung erörtert.

1. INTRODUCTION

Conditionals continue to receive much attention in psychology, philosophy, linguistics, and computer science (Bennett, 2003; Douven, 2015; Nickerson 2015). Indeed the literature on conditionals is so vast that it is difficult for researchers to keep track of its many branches within a single discipline and nearly impossible to keep track of it across disciplines. Part of the reason is that as much as researchers agree on the centrality of conditionals for understanding human reasoning and argumentation, as much do they disagree over how exactly to understand their semantic content (Edgington, 2014), pragmatics, computational models (Oaksford and Chater, 2010), and logical principles (Arlo-Costa, 2007).

In adding to this literature, the present dissertation selects a somewhat narrow scope by focusing on the empirical content of two theories of indicative conditionals in contemporary philosophy and psychology of reasoning. On the one hand, the *Suppositional Theory of conditionals* which was originally imported to the psychology of reasoning from formal epistemology. On the other, Inferentialism which is based on an old idea in philosophy that conditionals can be viewed as condensed arguments. While the Suppositional Theory of conditionals has established an impressive track record through more than a decade of intensive empirical testing (Over and Evans, 2003; Evans and Over, 2004; Over and Cruz, forthcoming), the empirical testing of the more recent introduction of Inferentialism into the psychology of reasoning (Spohn, 2013; Olsen, 2014; Douven 2015; Krzyżanowska, 2015; Skovgaard-Olsen 2015, 2016) still represents uncharted territory. Indeed, one of the main goals of the present dissertation is exactly to embark on the empirical investigation of Inferentialism through a series of experiments outlined in Chapters 2-4. Like Douven, Elqayam,

Singmann, Over, and Wijnbergen-Huitink (forthcoming), these studies represent some of the first attempts to investigate the psychological merits of Inferentialism directly.

Before presenting the studies themselves, this introduction will briefly cover some of the theoretical background for the Suppositional Theory of conditionals, outline some of the existing empirical results corroborating the theory, and identify three experimental paradigms that will set the frame for the rest of the dissertation.

The Suppositional Theory of Conditionals

In a famous footnote, Ramsey (1929/1990) suggested that two interlocutors could settle their argument over a conditional ‘if A, then C’ by hypothetically adding A to their stock of beliefs and arguing over C on that basis. As outlined in Arlo-Costa (2007), this little footnote outlining, what was later to become “the Ramsey test”, has inspired at least three opposing research programs in logics:

- The AGM belief-revision theory that only works with qualitative all-or-nothing beliefs and characterizes updating behavior under revision and contraction of belief states through eight axioms. The AGM theory was originally formulated with the goal of formulating a conditional logic based on principles of belief revision until Gärdenfoss proved a triviality theorem showing that the theory was not suitable for this intended purpose. Within AGM theory, the Ramsey test can be stated as follows:

$$\mathcal{B} \text{ supports } A > B \text{ iff } \mathcal{B} * A \text{ supports } B$$

for ' \mathcal{B} ' representing a belief set, ' $A > B$ ' representing a conditional, and ' $\mathcal{B} * A$ ' representing the revision of ' \mathcal{B} ' by A (Rott, 2017).

- The Stalnaker-Lewis possible worlds semantics of conditionals that supplies an influential account of the truth conditions of counterfactuals, according to which a counterfactual (i.e. ‘if A had been the case, then C would have occurred’) is true iff the consequent is true in all the closest

possible world(-s) in which the antecedent is true. In Stalnaker's notation, this idea is made precise by introducing the idea of a selection function, $f(A, w)$, which selects the closest world (or the set of closest worlds) to w in which A is true. The conditional, ' $A > C$ ', is then true iff the selected A -world(s) is a subset of the set of worlds in which C is true, $[C]$ (Égré and Cozic, 2016).

- The probabilistic semantics of indicative conditionals of Adams (1975), which in its original form denies that indicative conditionals have truth conditions, and subscribes to either $P(\text{if } A, \text{ then } C) = P(C|A)$ or $\text{acc}(\text{if } A, \text{ then } C) = \text{acc}(C|A)$, for 'if A , then C ' referring to simple conditionals (which exclude nestings of conditionals). Here $\text{acc}(\text{if } A, \text{ then } C)$ stands for the acceptability of the conditional. Often this version of Adams' thesis is preferred, because it is unclear whether $P(\text{if } A, \text{ then } C)$ can still be interpreted as a probability in light of the so-called triviality results, as we will see in Chapter 2 (Bradley, 2007; Douven, 2015).

Through the influence of the writings of Edgington (1995) and Bennett (2003), the psychological hypothesis that the probability of indicative conditionals is evaluated as a conditional probability found its way into the psychological literature (Evans and Over, 2004), where it goes by the name "the Equation". In the psychological literature, it is assumed that the Ramsey test is implemented by treating the false antecedent cases as irrelevant and assessing the ratio of $A \& C$ and $A \& \text{non-}C$ cases (Oberauer, Geiger, Fischer, and Weidenfeld, 2007). However, the exact mental processes involved in carrying out the Ramsey test still remain unknown (Over, Hadjichristidis, Evans, Handley, and Sloman, 2007).

Since Adams (1975) denied that indicative conditionals possess truth conditions, he was faced with the challenge of providing an alternative account of the validity of argument schemes involving conditionals. The solution he came up with was to introduce the notion of p -validity. An inference is p -valid, iff the uncertainty of the conclusion is not greater than the sum of the uncertainties of the inference's premises

for all probability distributions (with the uncertainty of a proposition, p , defined as $1 - P(p)$): $[1-P(A_1)] + [1-P(A_2)] + \dots + [1-P(A_i)] \geq 1 - P(B)$. The notion of p -validity generalizes the classical notion of deductive validity—according to which an argument schema is valid iff there does not exist a model in which all the premises are true and the conclusion is false (Garmut, 1991)—to the notion that a valid inference cannot take us from premises that have a high probability to a conclusion that has a low probability. Adams observed that this latter property of probability preservation from the premises to the conclusion was one that classically valid inferences had and introduced his more general notion of p -validity to deal with both the cases where the premises are uncertain and the cases in which conditionals feature as conclusions (Edgington, 2014).

In the psychology of reasoning, the Suppositional Theory of conditionals endorsing these views is often advanced in explicit opposition to analyses of indicative conditionals in terms of the material implication (\supset),¹ which has been widely accepted earlier in the psychology of reasoning. Since the material implication is logically equivalent to the disjunction ‘not-A or C’ it has introduction rules, which are appropriate for a disjunction. In classical logic a disjunction can always be inferred from either of its disjuncts (an inference rule known as “disjunction-introduction”). Consequently, if the natural language indicative conditional has the truth conditions of the disjunction ‘not-A or C’, then it should be possible to both infer ‘if A, then C’ from ‘not-A’ and to infer ‘if A, then C’ from ‘C’. In Chapter 2, we will explain why these argument schemes, which are also known as the paradoxes of the material implication, are usually rejected.

In contrast, the paradoxes of the material implication are not valid based on p -validity, which is usually seen as a decisive advantage of the account. In addition, the following argument schemes would have been valid based on the material implication but are no longer valid based on p -validity. This too is seen as an advantage of p -

¹ The material implication is logically equivalent to ‘not-A or C’ and ‘not(A and non-C)’ and is thus true in all other cells of the truth table than the A and not-C cell.

validity, since all of these argument schemes have well-known counterexamples discussed in the philosophical literature (Bennett, 2003: ch. 9):

Strengthening of the antecedent:

$$\frac{\text{if } A, C}{\therefore \text{if } A \text{ and } B, C}$$

Contraposition:

$$\frac{\text{if } A, C}{\therefore \text{if } \sim C, \sim A}$$

Transitivity:

$$\frac{\text{if } A, B; \text{if } B, C}{\therefore \text{if } A, C}$$

However, instead of following Adams (1975) in holding that indicative conditionals lack truth conditions, the proponents of the Suppositional Theory of conditionals in the psychology of reasoning usually take the de Finetti truth table as supplementing the Ramsey test (Evans and Over, 2004):

Table 1. De Finetti Truth Table

<i>A</i>	<i>C</i>	<i>If A, then C</i>
T	T	T
T	⊥	⊥
⊥	T	<i>void</i>
⊥	⊥	<i>void</i>

Note. 'T' = True, '⊥' = False

The Psychological Evidence

On the Suppositional Theory of conditionals, the word 'if' owes its distinctive character to its role in hypothetical thought by engaging the imagination to simulate possibilities (Evans and Over, 2004; Evans, 2007). For more than a decade of research, the Suppositional Theory of conditionals has become a widely accepted theory in the

psychology of reasoning and has been gaining grounds over against the mental model theory, which remains a popular theory of other types of reasoning like spatial reasoning (Johnson-Laird, 2008). The mental model theory used to adopt the material implication reading of indicative conditionals (Johnson-Laird and Byrne, 2002), but has recently been modified in response to persistent criticism to eschew a commitment to the material implication, however (Johnson-Laird, Khemlani, and Goodwin, 2015).

To prepare the grounds for the empirical studies reported in Chapters 2-4, the following two sections introduce some of the empirical evidence that has been accumulating for the Suppositional Theory of conditionals through this decade of empirical research. We will here focus on the classical truth table task and the more recent “probabilistic truth table task” (in Chapter 3 the uncertain and-to-if inference task is also discussed). For although the Wason selection task (where the participants are asked to select the cards needed to find out whether a conditional rule is true or false) and the conditional inference task (investigating the argument schemes MP, MT, AC, and DA) belong to some of the most studied tasks, these tasks do not produce separate predictions for the material implication and the Suppositional Theory of conditionals.²

The Probabilistic Truth Table Task

In Oberauer and Wilhelm (2003) the participants were presented with a distribution of 2000 playing cards with colored letters on them, which crossed frequency of pq and the ratio of pq to $p\neg q$ across four conditions, as shown in Table 2.

² Or I should say, at least this was the stance taken in Evans, Handley, Neilens, and Over (2007). In Baratgin, Over, and Politzer (2013) new predictions are made for the Suppositional Theory of conditionals in the Wason selection task that mimick the predictions made in Oaksford and Chater (2007).

Table 2. Frequency Distribution from Oberauer & Wilhelm (2003)

Cases	HH	HL	LH	LL
pq	900	900	90	100
$p \rightarrow q$	100	900	10	100
$\neg pq$	500	100	950	900
$\neg p \rightarrow q$	500	100	950	900

Note. 'HH', high frequency of pq , high $P(q|p)$. 'HL' high frequency of pq , low $P(q|p)$. 'LH' low frequency of pq , high $P(q|p)$. 'LL' low frequency of pq , low $P(q|p)$.

For this experimental design, the Equation predicts that the participants' assessment of, say, $P(\text{if there is an A on a card, then it is red})$ are sensitive only to the ratio of pq to $p \rightarrow q$. The material implication predicts that the assessment only depends on the relative frequency of $p \rightarrow q$. Yet, a conjunctive interpretation predicts that the probability assignment is only sensitive to the relative frequency of pq . Using this paradigm, Oberauer and Wilhelm (2003) found that the majority of the participants conformed to the Equation and that a minority of the participants followed the conjunctive interpretation. Moreover, in other similar experiments with frequency distributions and abstract stimulus materials (e.g. Evans, Handley, and Over, 2003), the participants generally judge $P(\text{if } p, \text{ then } q)$ as $P(q|p)$. A minority judge $P(\text{if } p, \text{ then } q)$ as $P(p,q)$ and judgments as $P(\neg p \text{ or } q)$ rarely occur.

In Evans, Handley, Neilens, and Over (2007) and Oberauer *et al.* (2007) individual variation in the conjunctive and the conditional probability responses was investigated. In Evans *et al.* (2007) it was found that 58% of the participants provided a conditional probability response and that 38% provided a conjunctive probability response (with the rest remaining unclassified), and that the participants providing the conditional probability response generally exhibited a higher level of cognitive ability. In Oberauer *et al.* (2007) it was found that 70% of the participants could be classified as providing conditional probability responses with 18% of the participants providing conjunctive probability responses.

The minority conjunctive-probability-response is puzzling since it seems to indicate a failure to engage with hypothetical thought, as Evans and Over (2004) point out. But it is compatible with mental model theory, which assumes that there are various levels of depth in the processing of indicative conditionals, where only creating a mental model of the A&C cell counts as the shallowest level of processing (Byrne and Johnson-Laird, 2002). Because Over *et al.* (2007) did not find a tendency to produce a conjunctive probability response pattern with realistic stimulus materials, Evans *et al.* (2007) speculate that the conjunctive response pattern may arise due to a combination of abstract stimulus materials and lower cognitive ability.

“Defective Truth Tables”

It has long been known that there is a tendency to treat the false antecedent cells in a truth table task as irrelevant for the truth value of indicative conditionals, when the participants are given a ternary response option. At a time when the material implication was the accepted model, this response pattern was dubbed ‘the defective truth table’. Later it has been reinterpreted as indicative of the de Finetti truth table (Evans and Over, 2004).

In the following, Baratgin, Over, and Politzer (2013) will be discussed as an illustration of the empirical evidence corroborating the de Finetti truth table. In their study, Baratgin *et al.* (2013) not only investigate the truth evaluations of indicative conditionals (‘If A then C’) but also of conditional bets (‘I bet that if A then C’). In the Bayesian tradition, subjective degrees of beliefs as measured by probability assignments are operationalized by betting quotients, where the participants have to make decisions about their relative preferences of betting on events described by pair of propositions. In the Bayesian literature on indicative conditionals going back to Ramsey and de Finetti, the indicative conditional is often compared to conditional bets. This comparison is meant to convey the idea that just as a conditional bet is rendered void, if the antecedent turns out to be false, so indicative conditionals possess an indeterminate truth value in the false antecedent cells of their truth tables.

Or as Baratgin *et al.* (2013: 309) put it: “When A turns out to be false, no indicative assertion or bet is made”.

Using deductive instructions where A and C were treated as certain, Politzer, Over, and Baratgin (2010) tested the 2x2 de Finetti table (see Table 1) both for indicative conditionals and conditional bets by asking the participants to make ternary truth evaluations, or win/lose evaluations, for every cell of the truth table. The results of their study were found to support Table 1. However, often in the so-called new Paradigm of psychology of reasoning, probabilistic instructions (where the premises are treated as uncertain and the response format is one of degrees of belief) are preferred over deductive instructions (where the premises are treated as certain and the participants are asked to make categorical judgments, Singmann and Klauer, 2011). For this reason, Baratgin *et al.* (2013) decided to replicate Politzer *et al.*'s (2010) results for the relation between indicative conditionals and conditional bets using the following 3x3 de Finetti table (see Table 3), where A and C are allowed to be uncertain:

Table 3. De Finetti Truth Table

A	C	If A, then C
T	T	T
T	U	U
T	⊥	⊥
U	T	U
U	U	U
U	⊥	U
⊥	T	U
⊥	U	U
⊥	⊥	U

Note. ‘U’ = Uncertain ‘T’ = True, ‘⊥’ = False

To introduce the uncertainty manipulation, the participants were presented with photos of two opaque boxes containing black/white and round/square chips. Occasionally, a chip from the top box would drop into the bottom box and the participants were presented with a photograph of how the chips looked as they

passed the empty space between the two boxes. These photographs could in turn be taken by cameras with different filters. As a result, under one condition the black/white chips might look grey (in which case the color of the chip would be indeterminate) or the shape might be indiscernible. Consequently, the participants could either be presented with information that conclusively confirm which color or shape the chips had while they passed from the top box to the bottom box (corresponding to the T and \perp cells of Table 3). Alternatively, the participants might be left in a state of uncertainty with respect to these attributes due to the blurring of the filters (corresponding to the U cells of Table 3).

The participants were instructed that two children would be playing with a game containing these two boxes, where one of the children would pick a chip at random from the top box and let it drop down into the lower box and the second child had to guess which chip it was. As part of this game, the second child would then make conditional assertions (e.g. "If the chip is round, then it is white") or conditional bets (e.g. "I bet that if the chip is round, then it is white"). The task of the participants was to give a ternary response about whether the conditional assertion was certainly true/false or neither nor, or the conditional bet had been certainly won/lost or neither nor, for each of the nine truth table cells in Table 3.

The results showed that there were no significant differences between the participants' responses to indicative conditionals and conditional bets and that when the 2x2 cells were considered for indicative conditionals, 38.6% agreed with the de Finetti truth table, 35.6% responded according to the conjunction, and 13.9% responded according to the material implication. Considering the remaining five cells of the 3x3 tables, more than half of the participants were classified as following the de Finetti truth table. The rest were classified according to other three-valued truth tables known in the literature. Again these results provide support for the de Finetti 2x2 and 3x3 truth tables and for the parallel between indicative conditionals and conditional bets, which in turn corroborates the Suppositional Theory of conditionals.

Inferentialism

In the chapters to come we will go more into details with Inferentialism, or the relevance approach to conditionals, as it was called in Skovgaard-Olsen (2016).

Inferentialism holds that indicative conditionals express inferential relations, or reason relations (defined in Chapter 2). On the strong reading, Inferentialism makes reason relations part of the truth conditions of indicative conditionals ("truth conditional Inferentialism") and rejects the validity of 'AAC \models if A, then C' (Douven, 2015; Krzyżanowska, 2015). Truth-conditional Inferentialism rejects the validity of this argument scheme, because the indicative conditional is viewed as expressing a reason relation and the mere truth of A and C does not ensure that they are inferentially connected. Rejecting the validity of the and-to-inference is a distinguishing feature of this approach, since and-to-if inferences are trivially valid on the material implication account and on accounts of conditionals that are inspired by the Ramsey test (Douven, 2015).

In Chapter 4, truth-conditional Inferentialism is tested directly by introducing a relevance manipulation into the truth table task. In Chapters 2-3, a weaker probabilistic implementation of Inferentialism is investigated by measuring the participants' probability assignments to indicative conditionals and their reasoning on the uncertain-and-to-if inference task to investigate whether the participants view indicative conditionals as defective when they fail to express a reason relation.

Putting the Two Theories to the Test

To contrast Inferentialism and the Suppositional Theory experimentally, the following paradigms were selected: (1) assessments of P(if A, then C) and the classical truth table task, because they are among the most frequently cited sources of evidence for the Suppositional Theory in the literature, and (2) the uncertain-and-to-if inference, because and-to-if inferences have been identified as diagnostic for deciding between Inferentialism and the Suppositional Theory.

The general idea behind these experiments is to test how robust the Suppositional Theory is under extreme conditions. To make an analogy: for a researcher interested in examining rationality, it makes sense to study cases of irrationality to investigate the boundary conditions and limitations of human rationality. For a researcher interested in studying text comprehension, and how the participants make sense of a string of connected passages, it is a valuable research strategy to study cases where participants are challenged and experience difficulties in deciphering the meaning of concatenated strings of words. Similarly, for researchers interested in studying relevance, and our use of conditionals to express reason relations, it is a sound research strategy to systematically investigate cases where relevance and reason relations break down. In Chapters 2-4, a series of experiments are conducted that did exactly this. Because most published studies prior to their publication tended to use stimulus materials where the antecedent is probability raising for the consequent for realistic stimulus materials (positive relevance), a central motive behind conducting these experiments was to create conditions that allow for violations of positive relevance. In each case, the goal was to investigate whether the patterns we find for positive relevance generalize to the pathological cases, where the antecedent lowers the probability of the consequent (negative relevance), or leaves it unchanged (irrelevance).

In Chapter 2, an experiment is reported that employed this research strategy for investigations of $P(\text{if } A, \text{ then } C)$ and $\text{Acc}(\text{if } A, \text{ then } C)$. In Chapter 3 an experiment is reported that employed it for the uncertain and-to-if inference task. And, finally, in Chapter 4 experiments are reported that employed it for the truth evaluation, probability and acceptability evaluations for conditionals and related connectives.

References

- Adams, E. W. (1975). *The Logic of Conditionals*. Dordrecht: D. Reidel.
- Arlo-Costa, Horacio (2007). The Logic of Conditionals. *The Stanford Encyclopedia of Philosophy* (spring 2016 Edition), Edward N. Zalta (ed.). URL = <http://plato.stanford.edu/archives/fall2016/entries/logic-conditionals/>.
- Baratgin, J., Over, D. E., Politzer, G. (2013). Uncertainty and the de Finetti tables. *Thinking & Reasoning*, 19(3), 308-28.
- Bennett, J. (2003). *A Philosophical Guide to Conditionals*. Oxford: Oxford University Press.
- Bradley, R. (2007). A Defence of the Ramsey Test. *Mind*, 116 (461), 1-21.
- Douven, I. (2015). *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. (in press). *Conditionals and inferential connections*.
- Edgington, D. (1995). On Conditionals. *Mind*, 104, 235-327.
- (2014). Indicative Conditionals. *The Stanford Encyclopedia of Philosophy* (Winter 2014 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2014/entries/conditionals/>.
- Égré, P. and Cozic, M. (2016). Conditionals. In: Aloni, M. and Dekker, P. (ed.), *The Cambridge Handbook of Formal Semantics*. Cambridge: Cambridge University Press, 490-524.
- Evans, J. St. B. T. (2007). *Hypothetical Thinking: Dual Processes in Reasoning and Judgment*. New York: Psychology Press.
- Evans, J. St. B. T., Handley, S. J., Neilens, H., and Over, D. E. (2007). Thinking about conditionals: A study of individual differences. *Memory & Cognition*, 35(7), 1772-84.

- Evans, J. St. B. T., Handley, S. J., and Over, D. E. (2003). Conditionals and conditional probabilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 321-35.
- Evans, J. St. B. T. & Over, D. (2004). *If*. Oxford: Oxford University Press.
- Johnson-Laird, P. N. (2008). *How we Reason*. Oxford: Oxford University Press.
- Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review*, 109, 646–678.
- Johnson-Laird, P. N., Khemlani, S. S., & Goodwin, G. P. (2015). Logic, probability, and Human reasoning. *Trends in Cognitive Science*, 19, 201–214.
- 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>>.
- Nickerson, R. S. (2015). *Conditional Reasoning. The Unruly Syntactics, Semantics, Thematics, and Pragmatics of “if”*. Oxford: Oxford University Press.
- Oaksford, M. and Chater, Nick (2007). *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*. Oxford: Oxford University Press.
- Oaksford, M. and Chater, N. (eds.) (2010). *Cognition and Conditionals: Probability and Logic in Human Reasoning*. Oxford: Oxford University Press.
- Oberauer, K., Geiger, S. M., Fischer, K., and Weidenfeld, A. (2007). Two meanings of „if“? Individual differences in the interpretation of conditionals. *The Quarterly Journal of Experimental Psychology*, 60(6), 790-819.
- Oberauer, K. and Wilhelm, O. (2003). The meaning(s) of conditionals—conditional probabilities, mental models, and personal utilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 680-93.
- 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. and Cruz, N. (forthcoming). Probabilistic accounts of conditional reasoning. To appear in: Linden J. Ball and Valerie A. Thompson (Eds.), *International Handbook of Thinking and Reasoning*. Hove, Sussex: Psychology Press.
- 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.
- Over, D. E., and Evans, J. S. B. T. (2003). The Probability of Conditionals: The Psychological Evidence. *Mind & Language*, 18(4), 340-58.
- Politzer, G., Over, D. E., and Baratgin, J. (2010). Betting on conditionals. *Thinking & Reasoning*, 16, 172-97.
- Ramsey, F.P. (1929). General Propositions and Causality. In: H. A. Mellor (eds.), *F. Ramsey: Philosophical Papers*, Cambridge: Cambridge University Press, 1990.
- Rott, H. (2017). Preservation and Postulation: Lessons from the New Debate on the Ramsey Test. *Mind*.
- Singmann, H. and Klauer, K. C. (2011). Deductive and inductive conditional inferences: Two modes of reasoning. *Thinking & Reasoning*, 17(3), 247-81.
- Skovgaard-Olsen, N. (2015). Ranking theory and conditional reasoning. *Cognitive Science*, 1-33.
- (2016). Motivating the Relevance Approach to Conditionals. *Mind & Language*, 31(5), 555-79.
- Spohn, W. (2013). A ranking-theoretic approach to conditionals. *Cognitive Science*, 37, 1074–1106.

2. THE RELEVANCE EFFECT AND CONDITIONALS³

Niels Skovgaard-Olsen

University of Konstanz and Albert-Ludwigs-Universität Freiburg

Henrik Singmann

University of Zürich

Karl Christoph Klauer

Albert-Ludwigs-Universität Freiburg

Skovgaard-Olsen, N., Singmann, H., Klauer, K. C. (2016).

The Relevance Effect and Conditionals. *Cognition*, 150, 26-36.

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/>

³ 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.

Abstract

More than a decade of research has found strong evidence for $P(\text{if } A, \text{ then } C) = P(C|A)$ (“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.

Keywords: Indicative conditionals, the New Paradigm, relevance, reasons, concessive conditionals, the Equation.

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: $P(\text{if } A, \text{ then } C) = P(C|A)$ for all probability distributions where $P(A) > 0$ and ‘If A, then C’ expresses a proposition.

ADAMS’ THESIS: $\text{Acc}(\text{if } A, \text{ then } C) = P(C|A)$ 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’.⁴

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: Ch. 5; Woods, 1997: Ch. 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):

Table 1. De Finetti Truth Table

A	C	If A, then C
T	T	T
T	⊥	⊥
⊥	T	void
⊥	⊥	void

THE EQUATION: $P(\text{if } A, \text{ then } C) = P(C|A)$, but ‘If A, then C’ does not express a classical proposition. (Bennett, 2003; Evans & Over, 2004; Oaksford & Chater, 2007)

⁴ 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.

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: Ch. 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 the consequent should somehow be connected or relevant for one another as one aspect of the conditionals’ meaning (for references see Skovgaard-Olsen, 2016; Krzyżanowska, 2015; Douven, 2015b). 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.

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).⁵ 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:

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

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 ‘ $\neg A \vee C$ ’ 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:

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

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

⁵ 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.

(2002; Byrne and 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 $0 < P(\text{premise}) < 1$ (Bennett, 2003, p. 139; Evans & Over, 2004; Oaksford & Chater, 2007; Pfeifer & Kleiter, 2011). However, as argued in Skovgaard-Olsen (2016), 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.⁶

P(If A, then C) and Relevance

The next surprise is that until quite recently,⁷ 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, Hadjichristidis, Evans, Handley, & Sloman, 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).

$$\text{THE } \Delta P \text{ RULE: } \Delta p = P(C|A) - P(C|\neg A)$$

As $P(C|A)$ is already occupied as a predictor of $P(\text{if } A, \text{ then } C)$ by the Equation, Over *et al.* (2007) and Singmann *et al.* (2014) try to obtain an orthogonal predictor for the relevance approach by using $P(C|\neg A)$. The evidence clearly favored $P(C|A)$ as a predictor. 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*:⁸

⁶ However, it should be noted that Over & 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 *et al.* (2007).

⁷ An exception is Douven, Elqayam, Singmann, Over, and Wijnbergen-Huitink (forthcoming). 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.

⁸ In these definitions we follow Spohn (2012: Ch. 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

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 $P(\text{if } A, \text{ then } C)$ 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 $P(\text{if } A, \text{ then } C)$ across systematic manipulations along the relevance dimension.

A further issue is that it is not entirely obvious what the relationship between $P(\text{if } A, \text{ then } C)$ and relevance should be on the relevance approach to conditionals. There are more options available than simply considering Δp as a predictor of $P(\text{if } A, \text{ then } C)$. Indeed, in Douven (2015a) and Olsen (2014: Ch. 3; see also Skovgaard-Olsen, 2015) it is suggested that an alternative could also provide a solution to the unsolved problem of where the participant's 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):

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.

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: Ch. 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 $P(\text{if } A, \text{ then } C)$ or $\text{Acc}(\text{if } A, \text{ then } C)$ the participants evaluate whether A is a sufficient reason for C. Applying the explication of the reason relation given in Spohn (2012: Ch. 6), this requires 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 $P(\text{if } A, \text{ then } C)$ and $\text{Acc}(\text{if } A, \text{ then } C)$.

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 $P(\text{if } A, C)$ or $\text{Acc}(\text{if } A, C)$ (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 and strength of the reason relation.

The Default and Penalty Hypothesis can be motivated by the observation that we use conditionals to display and discuss the inferential relations we are prepared to use in arguments (Fogelin, 1967, Brandom, 2010: 44–8, 104). 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, 2016). 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 $P(\text{If } A, \text{ then } C)$ 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: Ch. 3) suggestions, inferential relations and predictor relationships, respectively, play a role in determining $P(C|A)$, which in turn is used in determining $P(\text{if } A, \text{ then } C)$. According to the Default and Penalty Hypothesis, upon processing the antecedent with realistic materials⁹ and the conditional form, the participants will by default assume the positive relevance of the antecedent for the consequent. Hence, as

⁹ 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 issue 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 and Kratzer, 1998: Ch. 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 and Over, 2004). (We owe parts of this argument to discussions with Seth Yalcin.)

long as we are primarily investigating positive relevance stimulus material then participants should jump directly to the second step and evaluate $P(\text{if } A, \text{ then } C)$ solely on the basis of $P(C|A)$. But 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.

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.¹⁰ 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.¹¹ In contrast, studies investigating $P(\text{If } A, \text{ then } C)$ 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 $P(\text{If } A, \text{ then } C)$, 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

¹⁰ 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.

¹¹ 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.

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 $P(\text{If } A, \text{ then } C)$ without relying on *ad hoc* principles.

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 $P(\text{if } A, \text{ then } C)$. 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 (2016), 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 $P(C|A) = \text{low}$ 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. $P(C|A) \neq \text{low}$). 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. $P(C) = \text{high}$ and $P(C|A) \neq \text{low}$). 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, substitution of 'whether or 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.

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.

Method

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.Crowdfunder.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 seconds or in more than 3600 seconds, 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.

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 $P(A) = \text{high}$ and $P(C) = \text{low}$ 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.

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 roleplaying 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 725 participants reported in Skovgaard-Olsen *et al.* (draft) showed that the average Δp was .32 for the positive relevance conditions, -.27 for the negative relevance conditions, and -.01 for our irrelevance conditions.¹²

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, $P(A) = \text{low}$, $P(C) = \text{high}$) for the scenario text above:

¹² [The published version is here being corrected. The full sample of 725 participants was used to validate the stimulus materials and not just the sample of 495 participants after the exclusion.]

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 $P(\text{if } A, \text{ then } C)$, $\text{Acc}(\text{If } A, \text{ then } C)$, $P(\text{Even if } A, \text{ then still } C)$, or $\text{Acc}(\text{Even if } A, \text{ then still } C)$, 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. $\text{Acc}(\text{If } A, \text{ then } C)$ or $\text{Acc}(\text{Even if } A, \text{ then still } C)$), 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/>.

Results

Figures 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 $P(\text{if } A, \text{ then } C)$, $\text{Acc}(\text{if } A, \text{ then } C)$, $P(\text{Even if } A, \text{ then still } C)$, or $\text{Acc}(\text{Even if } A, \text{ then still } C)$) 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.

Indicative Conditionals

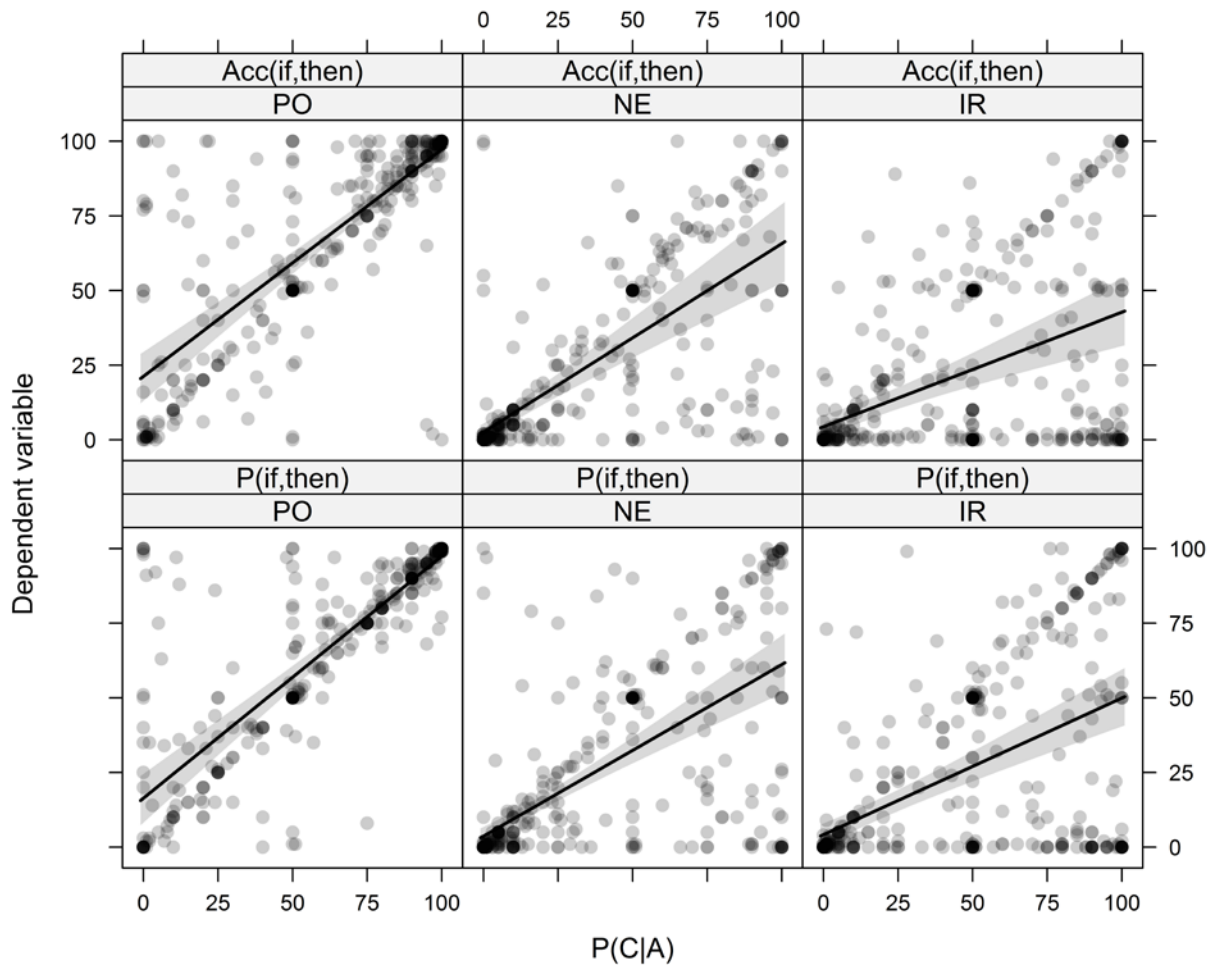


Figure 1. Raw data values (plotted with 80% transparency) and LMM estimated linear effect of $P(C|A)$ as a predictor on $\text{Acc}(\text{if } A, \text{ then } C)$ (upper row) and $P(\text{if } A, \text{ then } C)$ (lower row) across relevance manipulations (PO = left column, NE = center column, IR =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

Figure 1 seems to support our first hypothesis; for indicative conditionals the relevance condition seemed to affect the results but the mode of evaluation ($P(\text{if } A, \text{ then } C)$ vs. $\text{Acc}(\text{if } A, \text{ then } C)$) 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 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 ($P(\text{if } A, \text{ then } C)$ and $\text{Acc}(\text{if } A, \text{ then } C)$). 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]), $t(20.05) = 3.27$, $p_H = .008$ ¹³, 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

¹³ 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").

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 Figure 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 Figure 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,¹⁴ 0.98 in the PO condition, -0.95 in the NE condition, and -1.11 in the IR condition. Ensuring the distribution of random effects is not an artifact of the hierarchical modeling approach; since it shrinks extreme estimates towards the mean estimate, we also estimated separate regressions for

¹⁴ “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).

each individual and relevance condition. This analysis essentially showed the same pattern of results (see supplementary materials).

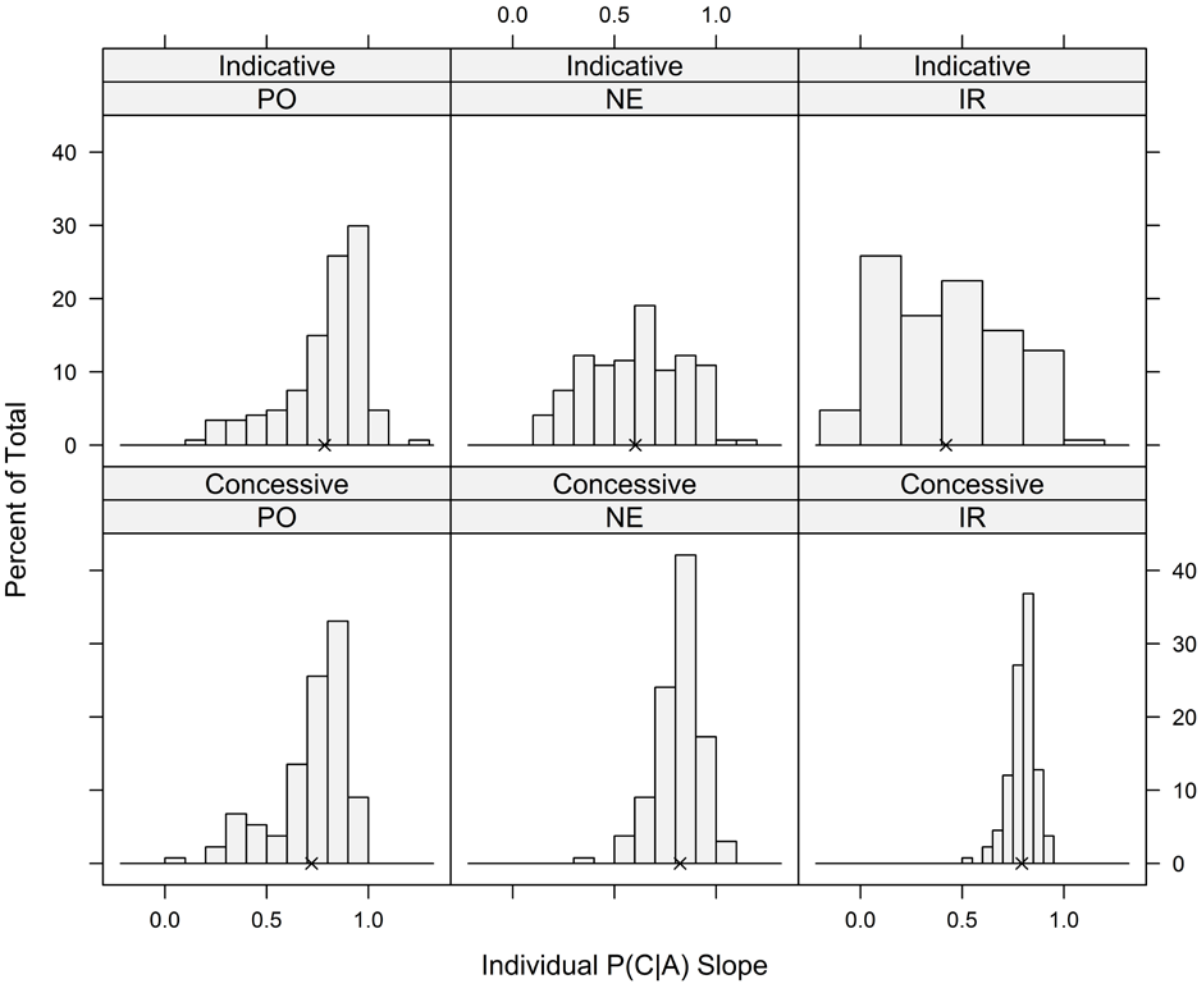


Figure 2. Individual slope estimates for the effect of conditional probability $P(C|A)$ on the dependent variable across conditions. These estimates are derived from the random effects terms of the LMM. In each plot each participant provided one slope estimate. The x denotes the fixed effects estimate.

Concessive Conditionals

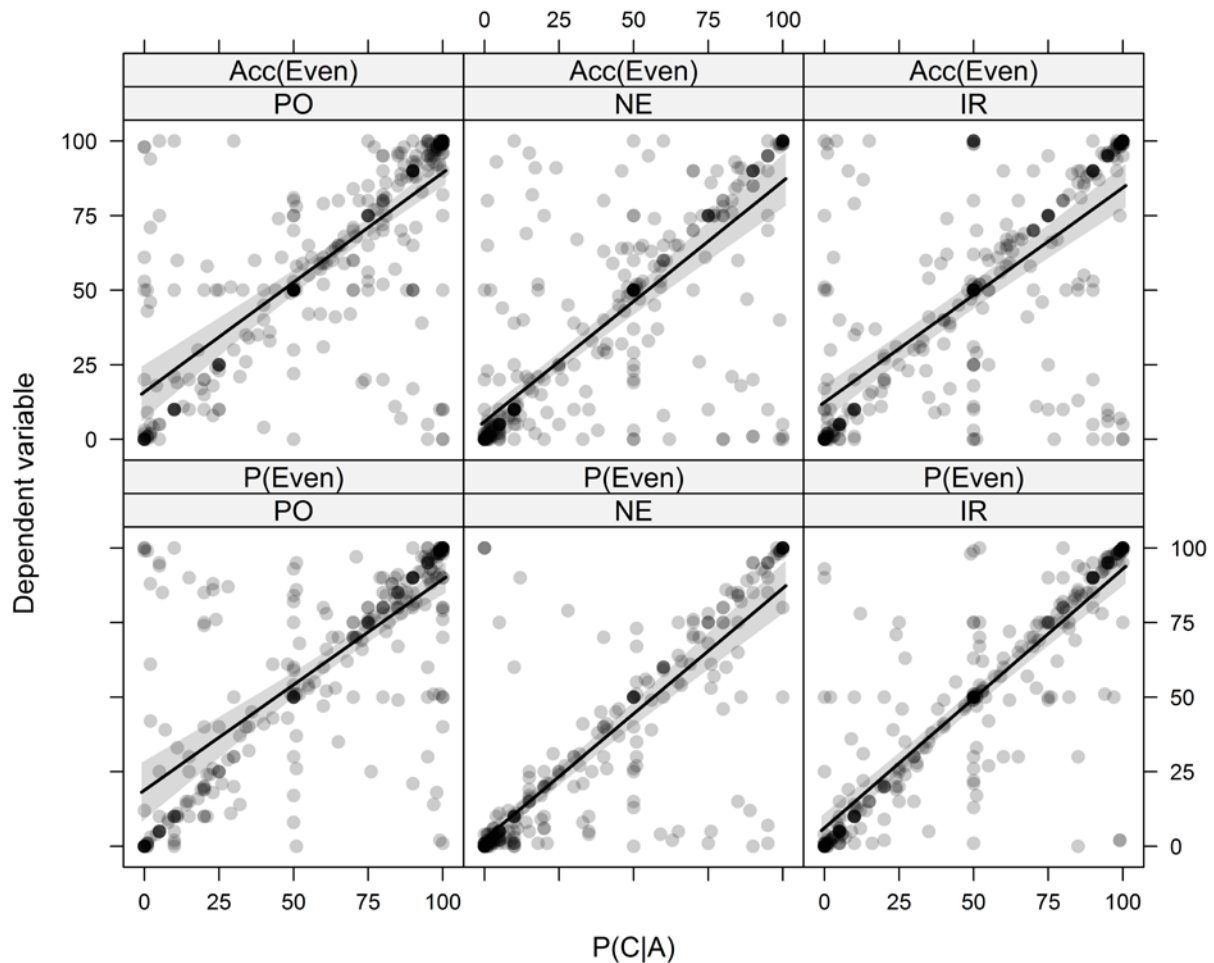


Figure 3. Raw data values (plotted with 80% transparency) and LMM estimated linear effect of $P(C|A)$ as a predictor on $\text{Acc}(\text{Even if } A, \text{ then still } C)$ (upper row) and $P(\text{Even if } A, \text{ then still } C)$ (lower row) across relevance manipulations (PO = left column, NE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

Inspection of Figure 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, $|t| < 1.8$, $p_H > .21$.¹⁵ The distribution of individual conditional probability effects (Figure 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).

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 $P(C|A) = \text{high}$. Yet, when

¹⁵ 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.

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 Figure 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.¹⁶ 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 $P(\text{If } A, \text{ then } C) = P(C|A)$ (Over & Evans, 2003; Evans & Over, 2004). 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

¹⁶ 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.

proponents of the mental model theory from using the same dialectical strategy to account for the evidence for $P(\text{If } A, \text{ then } C) = P(C|A)$ 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 continues to be deeply controversial in both the philosophical and linguistic literature on purely theoretical grounds (Bach, 1997, Birner, 2013: Ch. 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 $P(\text{If } A, \text{ then } C)$ 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

$P(\text{Even if } A, \text{ then still } C)$ than of $P(\text{If } A, \text{ then } C)$ 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 $\text{Acc}(\text{Even if } A, \text{ then still } C)$ 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 $P(\text{If } A, \text{ then } C) = P(C|A)$ 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:

GENERALIZED EQUATION: $P(\text{if } \phi, \text{ then } \psi | \chi) = P(\psi | \phi, \chi)$, for any ψ, ϕ, χ such that $P(\phi, \chi) > 0$

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.

Conclusion

More than a decade of research has offered the Equation ($P(\text{If } A, \text{ then } C) = P(C|A)$) 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 $P(\text{If } A, \text{ then } C)$. 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.

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. *J. Mem. Lang.*, 59(4), 390–412. doi: 10.1016/j.jml.2007.12.005.

- Bach, K. (1997). The Semantics-Pragmatics Distinction: What It Is and Why It Matters. *Linguistische Berichte*, 8, 33–50. doi: 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), 308–28. doi: 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. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (in press). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*. URL = <http://arxiv.org/abs/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. doi: 10.1093/acprof:oso/9780199542871.001.0001
- Byrne, R.M.J. and Johnson-Laird, P. N. (2009). ‘If’ and the Problems of Conditional Reasoning. *Trends in Cognitive Sciences*, 13(7), 282–7. doi: 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. doi: 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://doi.org/10.1037/1082-989X.2.3.292>
- Douven, I. (2010). The Pragmatics of Belief. *Journal of Pragmatics* 42 (1), 35–47. doi: [10.1016/j.pragma.2009.05.025](http://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. doi: 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. (forthcoming). Conditionals and Inferential Connections.
- Douven, I. and Verbrugge, S. (2010). The Adams Family. *Cognition* 117(3), 302–18. doi: 10.1016/j.cognition.2010.08.015
- Douven, I. and Verbrugge, S. (2012). Indicatives, Concessives, and Evidential Support. *Thinking and Reasoning* 18 (4), 480–99. doi: 10.1080/13546783.2012.716009
- Douven, I. and Verbrugge, S. (2013). The Probabilities of Conditionals Revisited. *Cognitive Science* 37(4). doi: 10.1111/cogs.12025
- Evans, J. St. B.T. and 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. and Kratzer, A. (1998). *Semantics in Generative Grammar*. Malden & Oxford: Blackwell Publishers.
- Johnson-Laird, P.N. and Byrne, R.M.J. (2002). Conditionals: A Theory of Meaning, Pragmatics, and Inference. *Psychological Review*. 109, 646–678. doi: 10.1037//0033-295X.109.4.646
- Johnson-Laird, P.N. Khemlani, S.S., and Goodwin, G. P. (2015). Logic, Probability, and Human Reasoning. *Trends in Cognitive Science* 19, 201–214.
doi: <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.
URL = <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 and Reasoning*, 13 (4), 340–69. doi: 10.1080/13546780601035794
- Olsen, N.S. (2014). *Making Ranking Theory Useful for Psychology of Reasoning*. PhD dissertation, University of Konstanz.
URL = <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), 340–58. doi: 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. doi: 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), 329–45.
doi: 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* (p. 145–66). Hove: Psychology Press.
- Politzer, G., Over, D. & Baratgin, J. (2010). Betting on Conditionals. *Thinking & Reasoning*, 16 (3), 172–97. doi: 10.1080/13546783.2010.504581
- Reips, U. D. (2002). Standards for Internet-Based Experimenting. *Experimental Psychology*, 49 (4), 243–256. doi: 10.1027//1618-3169.49.4.243
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Singmann, H., Klauer, K. C., and Over, D. (2014). New Normative Standards of Conditional Reasoning and the Dual-Source Model. *Front. Psychol.* 5:316. doi: 10.3389/fpsyg.2014.00316
- Skovgaard-Olsen, N. (2015). Ranking Theory and Conditional Reasoning. *Cognitive Science*. doi: 10.1111/cogs.12267

- Skovgaard-Olsen, N. (2016). Motivating the Relevance Approach to Conditionals. *Mind & Language*.
- Skovgaard-Olsen, N., Singmann, H., and 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. doi: 10.1111/cogs.12057
- Woods, M. (1997). *Conditionals*. Oxford: Oxford University Press. [Edited by Wiggins, D., Commentary by: Edgington, D.]

Appendix:

Details of the LMM Analysis

All analyses were performed using lme4 (Bates, Maechler, Bolker, & Walker, in press) 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 et al. (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 Figure 2). For an analysis of the effect of the priors see the supplementary materials.

Supplementary Materials

The Influence of the Prior Manipulation

Besides the two between-participants factors (conditionals and mode of evaluation) we had two within-participants factors: 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). Combining these two within-participants factors lead to the 12 items on which each participant worked.

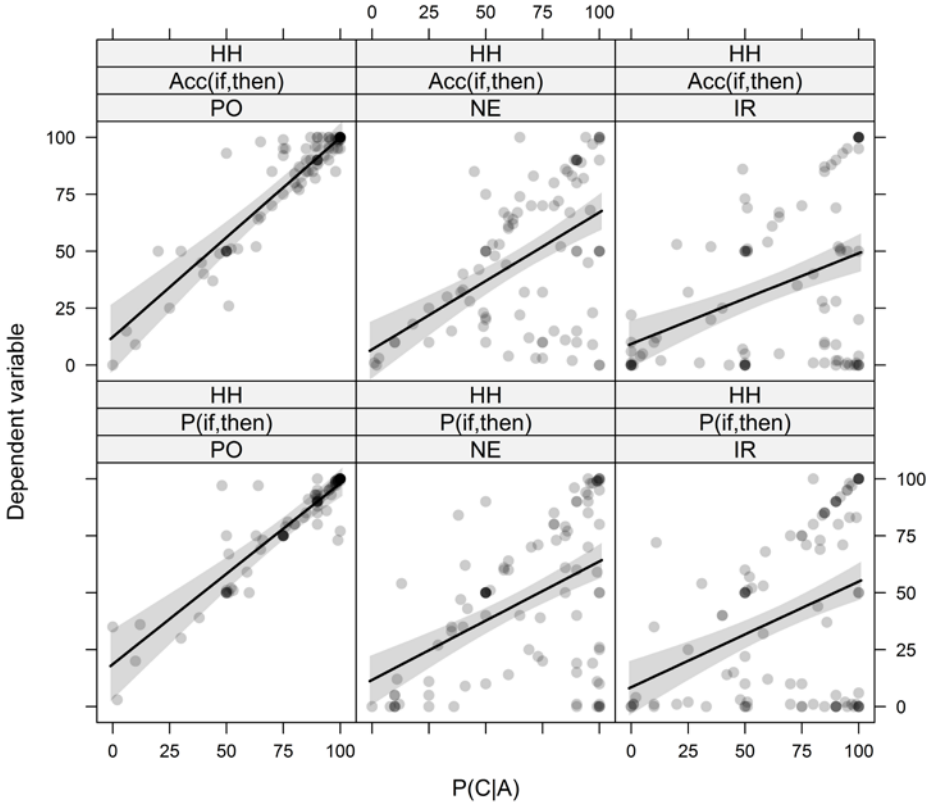
Indicative Conditionals by Prior Manipulation

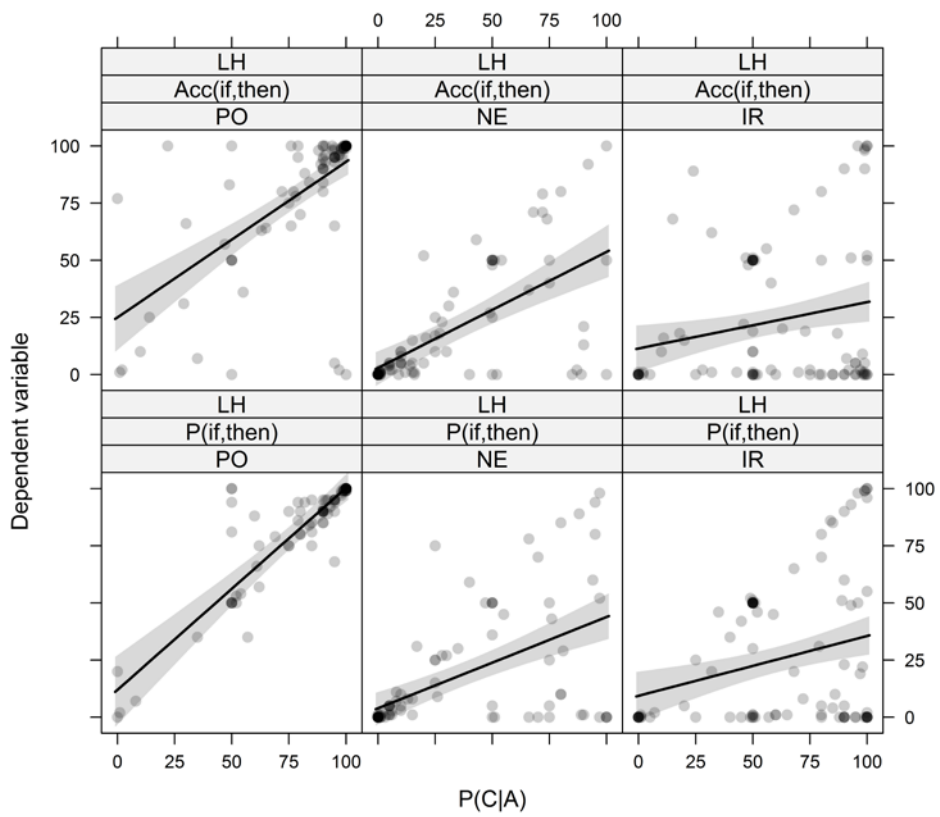
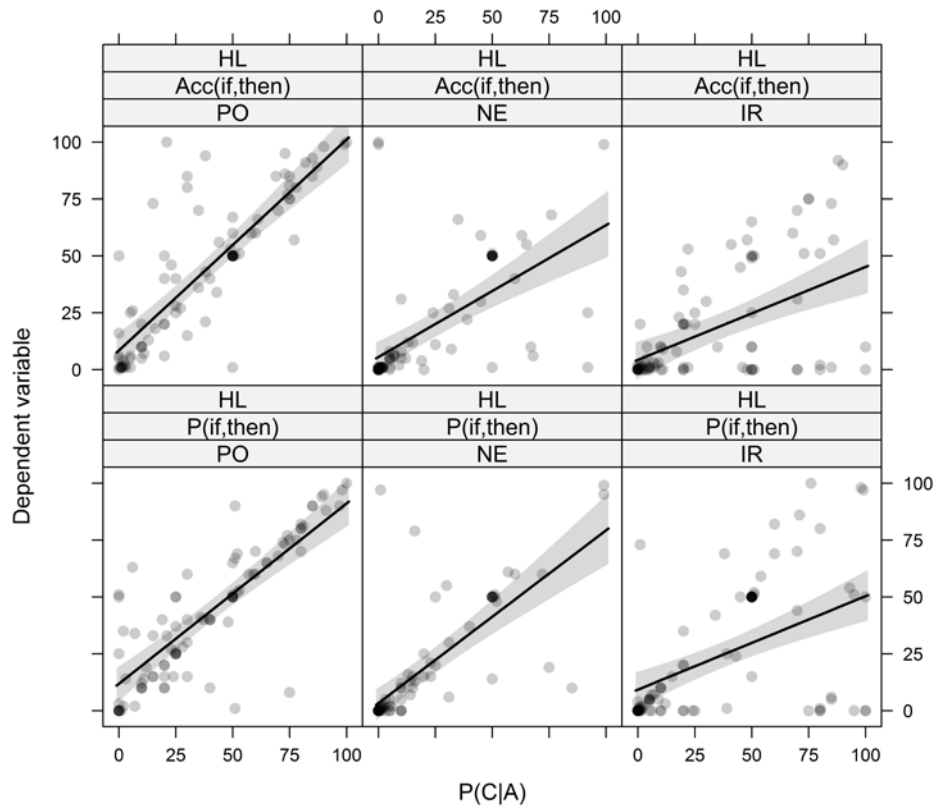
The following plots show the data for the indicative conditionals separated by prior manipulation. The order is HH ($P(A)$ = high and $P(C)$ = high), HL ($P(A)$ = high and $P(C)$ = low), LH ($P(A)$ = low and $P(C)$ = high), and LL ($P(A)$ = low and $P(C)$ = low). As can be seen when comparing HH and LH with HL and LL, manipulating the prior of the consequent achieved the intended goal of producing a spread in the conditional probability. For HH and LH most of the mass is on the right side of the scale (i.e., near 100), whereas for LH and LL most of the mass is near the left end of the scale (i.e. near 0). Interestingly, this pattern does not seem to hold completely consistently. For the LH and the NE relevance condition most of the data points remained on the left side and for the LL conditionals and the PO relevance condition the data still showed a relatively uniform spread. This shows that while in general the prior manipulation worked, participants' estimates of the conditional probabilities were not unaffected by the relevance condition.

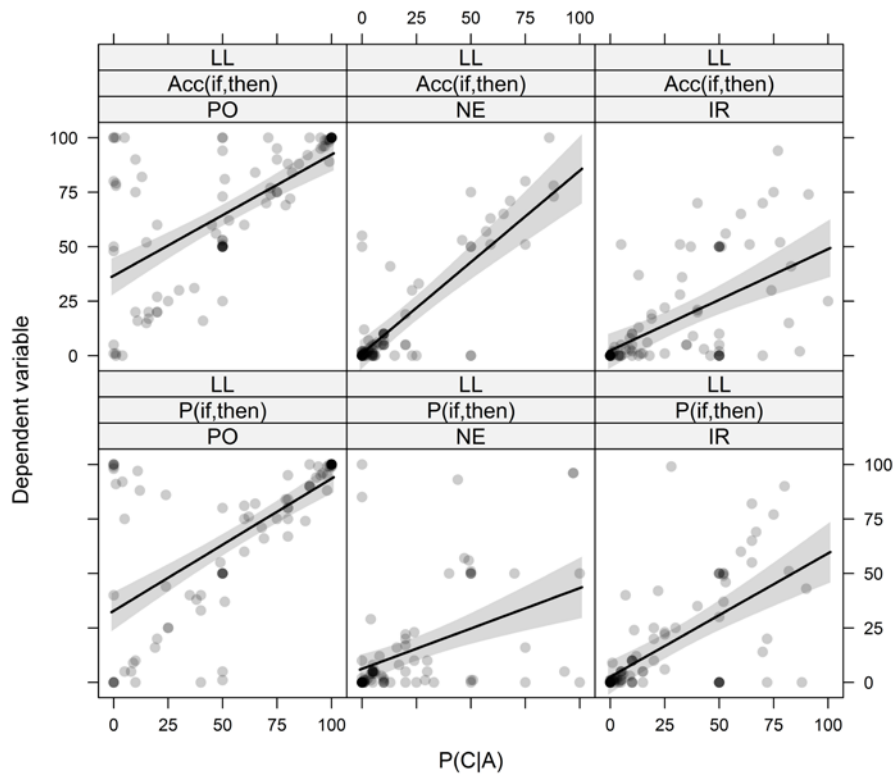
Given the reduced spread in each of the sub-plots the precision with which the individual slopes were estimated was obviously reduced compared to the main analysis. If all data points were on the same value of the independent variable (i.e., the

conditional probability), the estimated slope would be 0. Consequently, the few data points that did not have the same value of independent variable as most others have an unusually and unjustifiably high influence on the estimate of the slope (so-called *influential observations*). Nevertheless, the pattern was surprisingly robust. With the exception of the LL prior the estimated slopes always followed the order PO > NE > IR. For the LL prior, the estimates of PO and NE were almost identical. This shows that the main pattern holds across the prior manipulation and they did not systematically affect the results. The following tables gives the estimated slopes by condition aggregated across conditional type:

Condition	HH	HL	LH	LL
PO	0.83	0.86	0.78	0.58
NE	0.56	0.67	0.45	0.60
IR	0.43	0.41	0.23	0.52





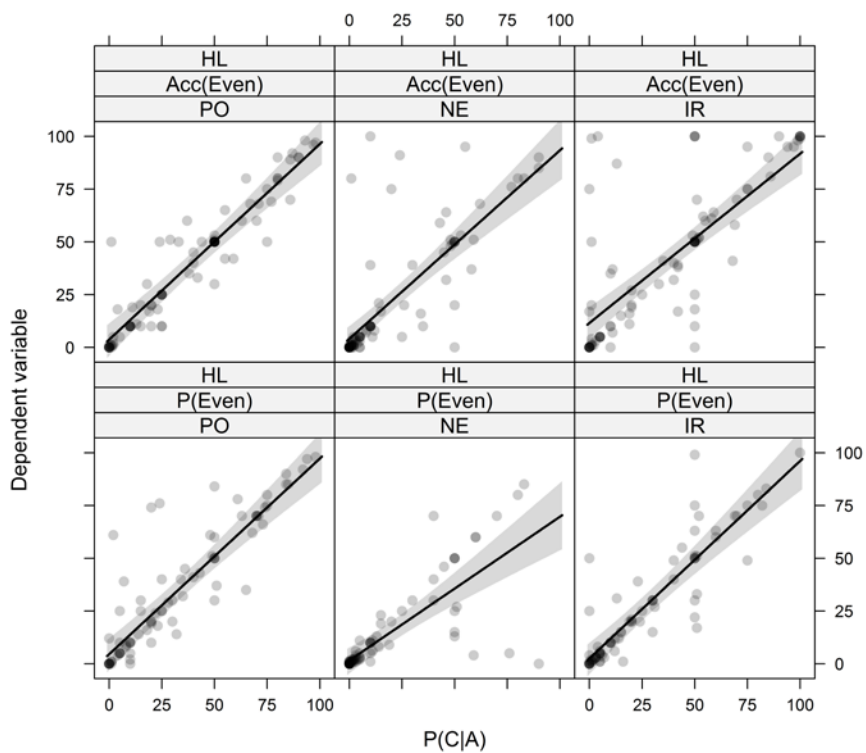
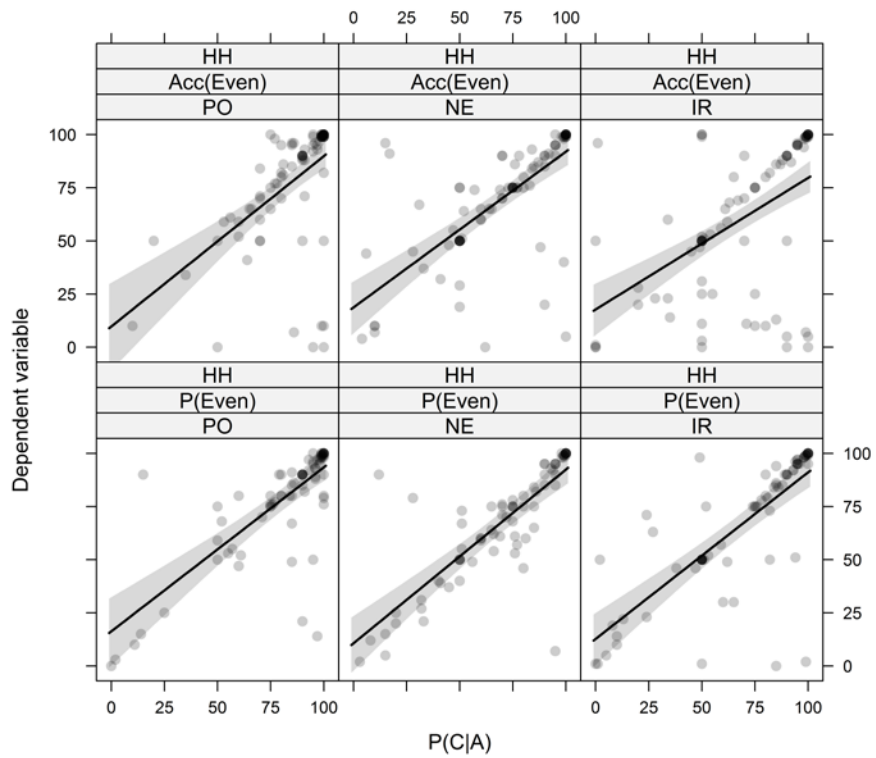


Concessive Conditionals by Prior Manipulation

For the concessive conditionals the following plots show the data by prior manipulation; the order is again HH, HL, LH, and LL. As before, the prior of the consequent (i.e., the second letter) seems to have a strong effect on where on the y-axis most of the data mass was located. For HH and LH, most data points were on the right side of the scale and for LH and LL most of the data points were on the left side of the scale. Also replicating the findings from the indicative conditionals, the only real outliers of this pattern seemed to be LH for NE and LL for PO.

As for the indicative conditionals, the pattern obtained for the full data set was also mostly replicated for each prior manipulation. For HH and HL the pattern was perfectly replicated despite the lowered spread (although there was some imprecision in the estimates at those parts of the scale for which there was little data). For LH and LL there was more variance in the estimated slopes but this was again due to some influential outliers: in the PO and NE conditions most data was so lumped at the ends

of the scale that just a few outliers were enough to drag the slope away from 1. There did not seem to be any systematic deviation from the overall pattern.



Analysis of Individual Regressions

Our hierarchical modeling approach (LMM) ensured that the individual estimates were distributed nicely around the mean slope estimate for each condition via shrinkage of the individual parameters. This allowed us to perform a joint analysis of all participants while simultaneously controlling for random participant and item (i.e., scenario) variability. However, this parameter shrinkage might not be completely desired at this point as it might mask a bimodal distribution of the slopes in the IR condition. Consequently, we also estimated individual regressions for each participant and condition, which are displayed in the following figure. In total we estimated 1044 individual regressions (348 x 3) of which 25 slopes were above 1.5 (max = 6.4), 2 were below -1.5 (min = -2.12), and 13 could not be estimated (as the conditional probabilities were constant).

For the indicative conditionals the figure shows a pattern very similar to the LMM estimates, but also hinted at a bimodal distribution in (at least) the IR condition with one peak around 0 and one peak around 1. The median slope estimate was 0.96 in the PO condition, 0.60 in the NE condition, and 0.29 in the IR condition confirming the pattern of the LMM analysis (the mean estimates were close to the LMM means, as shown when comparing the next figure to Figure 2 in the main text).

For the concessive conditionals, the median estimates from the individual regressions were 0.91 (PO), 0.94 (NE) and 0.93 (IR).

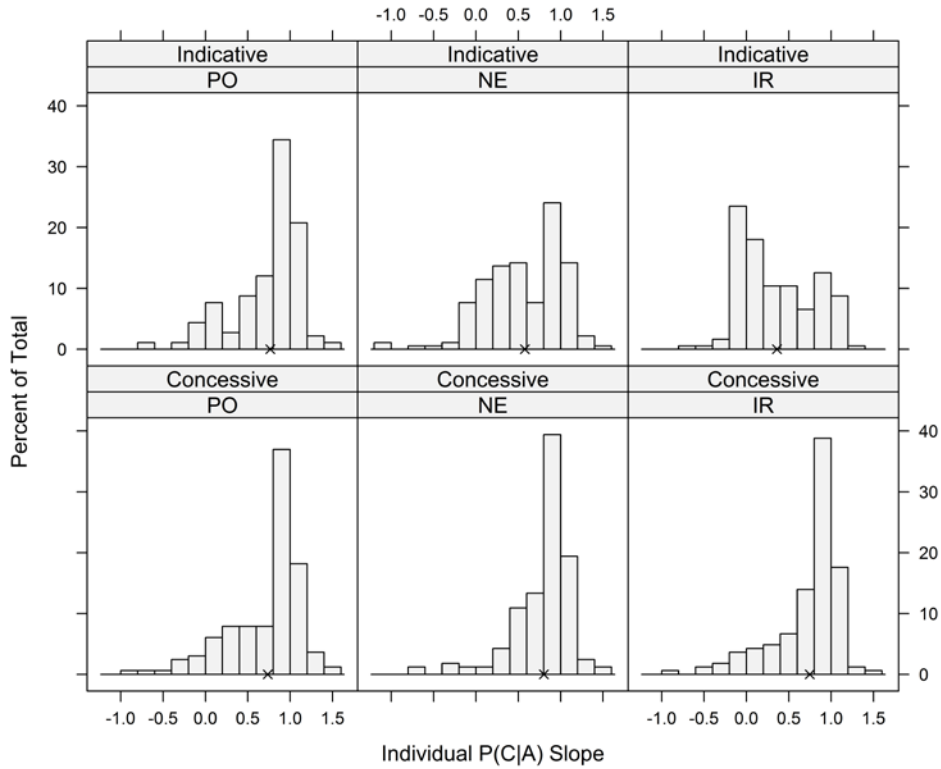


Figure Slopes. Individual slope estimates for the effect of conditional probability $P(C|A)$ on the dependent variable across conditions. These estimates are derived from individual regressions per relevance condition based on four data points each. In each plot each participant provided one slope estimate. We excluded estimates above 1.5 and below -1.5. The x denotes the mean of the displayed estimates.

3. RELEVANCE AND REASON RELATIONS¹⁷

Niels Skovgaard-Olsen

University of Konstanz and Albert-Ludwigs-Universität Freiburg

Henrik Singmann

University of Zürich

Karl Christoph Klauer

Albert-Ludwigs-Universität Freiburg

Skovgaard-Olsen, N., Singmann, H., Klauer, K. C. (2016).

Relevance and Reason Relations. *Cognitive Science*.

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/fdbq2/>

¹⁷ We are very grateful for discussions with David Over, Igor Douven, Vincenzo Crupi, Nicole Cruz, Wolfgang Spohn, Karolina Krzyżanowska, Peter Collins, Ulrike Hahn, and the audiences of talks at the annual meeting of New Frameworks of Rationality, the What-If group in Konstanz, the Department of Social Psychology and Methodology in Freiburg, and the ‘Working Group in the History and Philosophy of Logic, Mathematics, and Science’ in UC Berkeley. Moreover, careful comments by the reviewers substantially improved the manuscript.

Abstract

The present paper examines precursors and consequents of perceived relevance of a proposition A for a proposition C. In Experiment 1, we test Spohn's (2012, ch. 6) assumption that $\Delta P = P(C|A) - P(C|\sim A)$ is a good predictor of ratings of perceived relevance and reason relations, and we examine whether it is a better predictor than the difference measure $(P(C|A) - P(C))$. In Experiment 2, we examine the effects of relevance on probabilistic coherence in Cruz, Baratgin, Oaksford, and Over's (2015) uncertain "and-to-if" inferences. The results suggest that ΔP predicts perceived relevance and reason relations better than the difference measure and that participants are either less probabilistically coherent in "and-to-if" inferences than initially assumed or that they do not follow $P(\text{if } A, \text{ then } C) = P(C|A)$ ("the Equation"). Results are discussed in light of recent results suggesting that the Equation may not hold under conditions of irrelevance or negative relevance.

Keywords: Relevance, reason relations, and-to-if-inferences, conditionals, probabilistic coherence, the Equation

Introduction

Although the *reason relation* plays a central role in a number of philosophical discussions, a precise explication of this concept is usually absent (e.g., Brandom, 1994; McDowell, 1994; Brewer, 2002; Reisner & Steglich-Petersen, 2011). In Spohn (2012, ch. 6), a precise account has, however, been given in terms of the difference that one proposition, A, makes in the degree of belief of another proposition, C, which draws on the literature on confirmation measures:¹⁸

A is a reason *for* C iff $P(C|A) > P(C|\sim A)$ [1]

A is a reason *against* C iff $P(C|A) < P(C|\sim A)$ [2]

At the same time, the notion of epistemic *relevance* is explicated by stating that A is *positively relevant to* C iff [1] holds, *negatively relevant* iff [2] holds, and *irrelevant* iff $P(C|A) = P(C|\sim A)$. One of the general advantages of having such a formal account is that it enables one to investigate the formal properties of reason relations and relevance and to formulate a taxonomy of reason relations (see Spohn, 2012, section 6.2), which has repercussions for their application to philosophy and psychology (Spohn, 2013; Skovgaard-Olsen, 2015).

The first goal of the present study is to test the following prediction attributed to Spohn (2012): there is both a high correlation between ΔP (i.e. $P(C|A) - P(C|\sim A)$) and perceived relevance and between ΔP and ratings of reason relations. As ΔP is only one among a whole family of confirmation measures we contrast it with the difference measure ($P(C|A) - P(C)$), which is another popular alternative (Tentori, Crupi, Bonini, and Osherson, 2007; Douven and Verbrugge, 2012). Formally, $P(C|A) > P(C|\sim A)$ entails $P(C|A) > P(C)$.¹⁹ However, the degree of relevance as measured by $P(C|A) - P(C|\sim A)$

¹⁸ Yet it should be noted that there is large problem of the unification of theoretical reasons and practical reasons raised by the contributions in Reisner and Steglich-Petersen (2011), which Spohn's account does not yet tackle, and that there are predecessors for analyzing epistemic relevance in the way Spohn does in the literature (see Falk and Bar-Hillel, 1983; Walton, 2004, ch. 4).

¹⁹ In the special case where $P(A) = 1$, $P(C|A) = P(C)$ but $P(C|\sim A)$ is undefined.

need not match the degree of relevance as measured by $P(C|A) - P(C)$. This raises the empirical issue of which of the two best describes the degree of perceived relevance and the perceived strength of the reason relation of the participants.²⁰

In Experiment 2 we turn to the effects of relevance on the probabilistic coherence of the participants in the uncertain and-to-if inference (i.e. inferring ‘if A then C’ from ‘A and C’).

Experiment 1

Method

Participants

A total of 725 people from the USA, UK, and Australia completed the experiment, which was launched over the Internet (via www.Crowdfunder.com) to obtain a large and demographically diverse sample. Participants were paid a small amount of money for their participation.

The following exclusion criteria were used: not having English as native language (33 participants), completing the experiment in less than 240 seconds or in more than 5400 seconds (43 participants), failing to answer two simple SAT comprehension questions correctly in a warm-up phase (214 participants), providing answers outside the range of 0% to 100% (three participants), and answering ‘not serious at all’ to the question how serious they would take their participation at the beginning of the study (zero participants). Since some of these exclusion criteria were overlapping, the final sample consisted of 475 participants. Mean age was 38.91 years, ranging from 18 to 73, 55.8 % indicated that the highest level of education that they had completed was an undergraduate degree or higher.

²⁰ Actually, Spohn (2012: ch. 6)’s preference for the delta-p measure over the difference measure is grounded in the different behavior of their ranking theoretic analogues. Although it would indeed be attractive to investigate psychological applications of ranking theory, the present study takes the more conservative, probabilistic route.

Design

The experiment implemented a mixed design with three factors that determined the content and relationship of the antecedent, A, and consequent, C, of a conditional ‘If A then C’. There were two factors that varied within participants: relevance (with three levels: positive relevance (PO), negative relevance (NE), irrelevance (IR)), and priors (with four levels: HH, HL, LH, LL, meaning, for example, that $P(A) = \text{low}$ and $P(C) = \text{high}$ for LH). One further factor varied between participants: type of irrelevance (with two levels labelled ‘same content’ and ‘different content’). Participants were randomly assigned to one of the two irrelevance conditions for each scenario implementing a conceptual distinction between whether A is topically relevant or irrelevant for C. As this factor did not affect any of the results reported here we do not discuss it any further (see supplementary materials, section 3) and only use the different content irrelevance condition in Experiment 2.

Materials and Procedure

We created 18 different scenarios (see supplemental materials for full list) for each of which we constructed 16 conditions according to our design (i.e., 4 conditions for PO [i.e., HH, HL, LH, LL], 4 conditions for NE, 4 conditions for IR-same content, and 4 conditions for IR-different content; note again that the two IR conditions were collapsed for the analysis as they did not differ). Each participant worked on one randomly selected (without replacement) scenario for each of the 12 within-subjects conditions such that each participant saw a different scenario for each condition.²¹ Following the recommendations of Reips (2002) to reduce dropout rates, we presented two SAT comprehension questions as an initial high hurdle in a warm-up phase (in addition to using them for excluding participants).

The experiment was split into twelve blocks, one for each within-subjects condition. The order of the blocks was randomized anew for each participant and

²¹ The supplementary materials contain details on how 12 scenarios were selected for future experimentation on the basis of the 18 scenarios we created.

there were no breaks between the blocks. Within each block, participants were presented with two pages. The scenario text was placed at the top of each page. One participant might thus see the following scenario text:

Julia has gained some weight during her holiday in Egypt, and now wishes to lose 5 kilos. She is very determined to make lifestyle changes. She is not obese by any means. Yet it is unlikely that she will end up looking like a model — nor is it her goal. Most would characterise her as being within the normal range.

The 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. So to introduce the 12 within-subjects conditions for the scenario text above we, *inter alia*, exploited the fact that participants would assume that Julia's beginning to exercise would raise the probability of her losing weight (PO), lower the probability of her gaining weight (NE), and that a sentence describing the present weather conditions of the location where Julia spent her holiday would be irrelevant for whether or not 'Julia will lose weight' by exercising after returning from the holiday (IR).

On the first page of each block, the scenario text was followed by two questions presented in random order that measured the prior probability of the two sentences:

Please rate the probability of the following statement on a scale from 0 to 100%:

[Julia begins to exercise/ Julia will gain weight]

On the second page, the same scenario text was followed by four questions presented in random order. The first two questions measured the conditional probability of the consequent given the antecedent, $P(C|A)$, and its negation, $P(C|\sim A)$. To illustrate using the NE-LL condition (= negative relevance, $P(A) = \text{low}$, $P(C) = \text{low}$) for the scenario above:

Suppose Julia has weight loss surgery.

Under this assumption, how probable is it that the following sentence is true on a scale from 0 to 100%:

Julia will gain weight.

The third question, the relevance rating, asked the participants to rate the extent to which the antecedent was relevant for the consequent on a five point scale ranging from 'irrelevant' to 'highly relevant'. The fourth question, the reason relation scale, asked the participants to rate the extent to which the antecedent was a reason for/against the consequent on a five point scale ranging from 'a strong reason against', 'a reason against', 'neutral', and 'a reason for' to 'a strong reason for'. For each question, participants gave their response by entering a number into a specified field. The full list of scenarios, the raw data, the data preparation script, and the analysis script for both Experiment 1 and 2 can all be found at: <https://osf.io/fdbq2/>.

Results and Discussion

We performed a manipulation check (see supplementary materials) and prepared the data for the analysis. Perceived relevance was initially measured in an undirected way, because it was assumed that participants would not be sensitive to the theoretical distinction between positive and negative relevance. To obtain a directed perceived relevance rating, we combined the directional information of the reason relation scale with the relevance rating to generate a directional relevance scale ranging from -4 (strongly negatively relevant) to +4 (strongly positively relevant). If participants indicated that A was a reason *against* C on the reason relation scale, their assessment of how relevant A was for C was interpreted as *negative* relevance. If the participants indicated that A was a reason *for* C, their assessment of how relevant A was for C was interpreted as *positive* relevance. If the participants indicated that A was neutral in relation to C, their assessment was interpreted as that A was *irrelevant* for C. The reason relation scale was coded on a scale from -2 (strong reason against) to +2

(strong reason for). ΔP and the difference measure were calculated from the conditional probability questions and the prior of the consequent.

As the data had replicates both on the level of the participant (each participant provided one response for each of the 12 within-participant conditions) and on the level of the scenarios (each scenario could appear in each relevance condition across participants) we employed a linear mixed model (LMM) analysis with crossed random effects for participants and scenarios for the analysis (Baayen, Davidson, & Bates, 2008). We estimated one LMM with directional relevance scale as dependent variable and one LMM with reason relation as dependent variable. Each LMM had fixed effects for ΔP as well as the difference measure. The random effects structures were “maximal” (Barr, Levy, Scheepers, & Tily, 2013): random intercepts for participants and contents with by-participant and by-content random slopes for both fixed effects, and correlations among all by-participant and among all by-content random terms. Fixed effects were evaluated via the Kenward-Roger approximation (via afex; Singmann, Bolker, Westfall, & Aust, 2016). The LMMs did not include effects for the prior manipulations, which were primarily introduced to ensure that our results generalize to the whole spectrum of sentences describing likely and unlikely events.

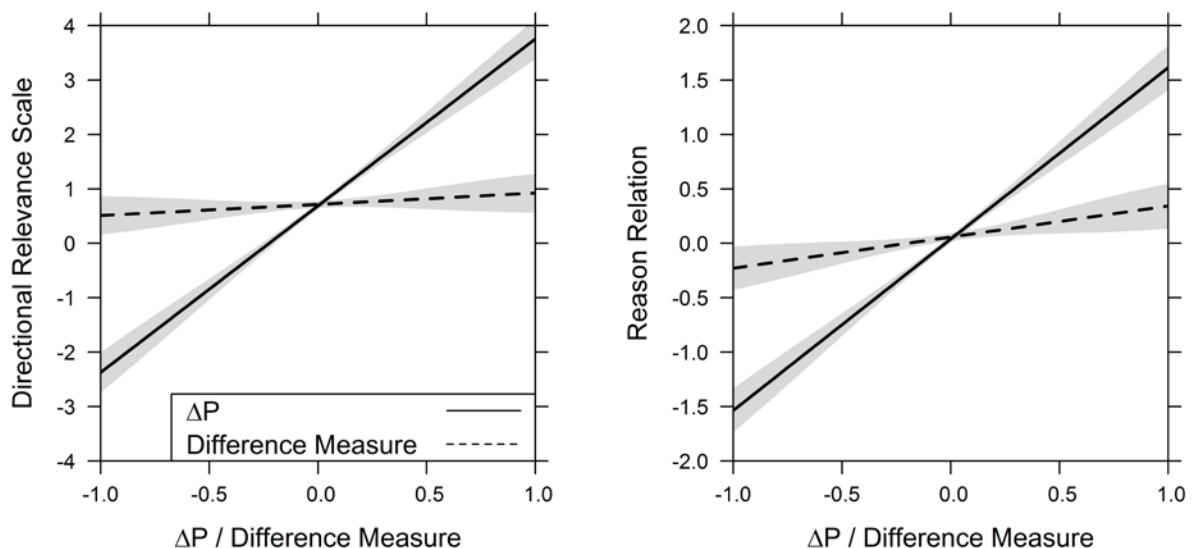


Fig. 1. LMM estimates of fixed effects for Experiment 1. In the left panel the directional relevant scale is the dependent variable and in the right panel reason

relation is the dependent variable. Error bands show 95% confidence intervals from the LMM.

Fig. 1 displays the estimated effects from both models which clearly show that the effects of ΔP are considerably stronger than the effects of the difference measure (dashed lines). Furthermore, for the model with relevance scale as dependent variable, ΔP was a significant predictor, $F(1, 20.94) = 269.07, p < .0001$, but the difference measure failed to be, $F(1, 18.21) = 1.30, p = .27$. This indicates that in contrast to the difference measure, only ΔP could explain unique variance. For the model with reason relation as dependent variable both $\Delta P, F(1, 18.99) = 232.35, p < .0001$, as well as the difference measure, $F(1, 17.26) = 7.72, p = .01$, were significant predictors.²² Although initially purely philosophically motivated, it turns out that the explication of relevance and reason relations of Spohn (2012) in terms of ΔP is descriptive of the assessments of our participants. See supplementary materials (section 2) for a comparison with a further confirmation measure.²³

Experiment 2

The last 10-15 years of research on conditionals within the psychology of reasoning have been marked by the emergence of a *New Paradigm* characterized by a shift from models based on classical logic to probabilistic competence models (Elqayam & Over, 2013). Within the New Paradigm there is a widespread endorsement of *the Equation*, $P(\text{if } A, \text{ then } C) = P(C|A)$ (Evans & Over, 2004; Oaksford & Chater, 2007; Baratgin, Over,

²² Note that the zero-order correlations of the difference measure with both dependent variables were highly significant ($r > .39$). But its effect was reduced in the joint LMM due to the high covariance with ΔP ($r = .73$), which was itself more strongly correlated with both dependent variables ($r > .53$).

²³ We tested a further confirmation measure, Keynes and Horwich's logged-ratio measure (Tentori *et al.*, 2007). Unfortunately, this measure introduces the problem of extreme ($-\infty$) or undefined values for 24% of our observations (e.g., when the denominator is 0). When analyzing the reduced sample, no unique variance was accounted for by the logged-ratio measure and again only ΔP was a significant predictor (see section 2, supplementary materials). Other proposed confirmation measures contain variables not collected in this study (such as the likelihood) and could therefore not be applied to our data.

and Politzer, 2013; Pfeifer, 2013). In addition to direct evidence stemming from investigations of the probability of the conditionals, and evidence from the truth table task (Over & Evans, 2003), it has been suggested that evidence from uncertain and-to-if inferences supports this view (Cruz, Baratgin, Oaksford, and Over, 2015). Cruz *et al.* (2015, p. 3) use their results from uncertain *and-to-if inferences* to make an argument in favor of the Equation based on the following line of thought: “If people’s judgments are highly incoherent for one interpretation [of the conditional], and yet highly coherent for another, there is an argument in favor of the interpretation that renders their judgments coherent”. Since it was found that the Equation was better able to make the participants’ responses coherent than the material conditional,²⁴ they interpret their results as providing strong evidence in favor of the Equation.

In Skovgaard-Olsen, Singmann, and Klauer (2016), we found that the evidence for the Equation was qualified once $P(\text{if } A, \text{ then } C)$ was evaluated across three relevance levels, where relevance was defined as described above.²⁵ While there was an almost perfect relationship between $P(\text{if } A, \text{ then } C)$ and $P(C|A)$ in the positive relevance (PO) condition, this relationship was markedly weaker in the negative relevance (NE) and even weaker in the irrelevance (IR) condition. Moreover, the results showed that $P(C|A)$ is a much better predictor of $P(\text{Even if } A, \text{ then still } C)$ across relevance levels. The second goal of the present study is therefore to test whether introducing the same relevance manipulation to and-to-if inferences leads participants to perceive a defect in the conditionals in the NE and IR conditions, which should make them more reluctant to infer the conclusion under these conditions.

Following our earlier findings (Skovgaard-Olsen *et al.*, 2016), we hypothesize that the results of Cruz *et al.* (2015) are similarly affected by a relevance manipulation. More specifically, in line with Skovgaard-Olsen *et al.* (2016) we hypothesize that for

²⁴ The material conditional (\supset) has a truth table that is logically equivalent to $\neg A \vee C$ and for this reason, Cruz *et al.* (2015) attribute the prediction that $P(\text{if } A, \text{ then } C) = P(\neg A \vee C)$ to this theory.

²⁵ For results on introducing the relevance manipulation into the truth table task see Skovgaard-Olsen, Kellen, Krahl, and Klauer (in review).

indicative conditionals we replicate their findings in the PO condition, but not in the NE or IR conditions. In contrast, for the concessive (i.e., *even-if*) conditionals, the level of probabilistic coherence of the participants was not expected to drop in the NE/IR conditions as compared to the PO condition. Hence, we test whether the participant's degree of probabilistic consistency drops under manipulations of negative relevance and irrelevance for indicative conditionals, when $P(\text{if } A, \text{ then } C)$ and $P(C|A)$ are equated.

Method

Participants

The present experiment was part of Skovgaard-Olsen *et al.* (2016); consequently we analyzed the same 348 participants reported there.²⁶ However, the data for this specific task are reported here for the first time. Data were collected over the Internet.

Design

Experiment 2 implemented a mixed design with the same 12 within-participant conditions as Experiment 1. In addition, the type of conditional was varied between participants (with two levels: indicative ('if A, then C'), concessive ('Even if A, then still C')).

Materials and Procedure

Prior to Experiment 2, we selected 12 scenarios from the set of 18 scenarios for which all within-subjects condition were most precisely realized (see supplementary

²⁶ In contrast to the other task in Skovgaard-Olsen *et al.* (2016), the present task did not involve differentiating between assessing the probability and acceptability of the respective sentences as two modes of evaluation. For this reason, the two groups evaluating probabilities and acceptabilities separated in Skovgaard-Olsen *et al.* (2016) are analyzed together below.

materials). The 12 within-participants conditions were randomly assigned to 12 different scenarios for each participant anew.²⁷ Moreover, in contrast to Experiment 1, the participants reported probabilities using sliders ranging from 0% to 100%. Aside from this, Experiment 2 was designed following the schema of Experiment 1.

Within each of the 12 within-participants conditions, the participants were presented with three pages, which had a randomly chosen scenario text at the top. On the first page of the experiment, the scenario text was followed by two questions presented in random order. The first measured the conditional probability of the consequent given the antecedent using the same question format as in Experiment 1. The second question measured the probability of the conjunction of the antecedent and the consequent, which was used to measure the probability of the premise of an inference task on the third page.

On the second page, the participants evaluated either the acceptability or the probability of conditionals in a task reported in Skovgaard-Olsen *et al.* (2016). On the third page, the participants were presented with a short argument, whose premise was the conjunction, and a conditional as the conclusion. The participants were here reminded of the probability that they had assigned to the conjunction on the first page and asked to assess the probability of the conditional on its basis. Thus, one participant might see the following question on page three:

In the following you will be presented with a short argument.

Premise: Julia starts to exercise AND Julia will gain weight.

(You have estimated the probability of the premise as: 13%)

Based on the premise and its probability, please indicate how much confidence you have in the following conclusion:

Conclusion: Therefore, IF Julia starts to exercise, THEN Julia will gain weight.

²⁷ See also: <https://osf.io/j4swp/>.

Results and Discussion

We estimated probabilistic coherence across relevance manipulations in the and-to-if inference following Cruz *et al.* (2015).²⁸ This entailed comparing the observed coherence rates against a chance coherence rate.²⁹ As shown in Table 1, the descriptive data seemed to confirm our predictions. For indicative conditionals participants' probabilistic coherence was above chance levels only for PO while for the concessive conditionals participants' probabilistic coherence was above chance levels for all relevance conditions. Table 1 also shows whether participants' probability evaluations conformed to $P(C|A) \geq P(A,C)$ independent of the uncertain and-to-if inference task (i.e., both responses from the first page of each within-subject condition). Participants reliably conform to this inequality in $\approx 78\%$ of the cases ($\approx 19\%$ above chance) across relevance levels with 77% in PO, 81% in NE, and 76% in IR. In contrast, the participants' conformity to $P(\text{Conclusion}) \geq P(A,C)$ varied markedly across relevance levels with 87% in PO, 66% in NE, and 54% in IR. Given this apparent discrepancy between the effects of our relevance factor on the conformity to these two inequalities, we decided to analyze the effect of relevance on the conformity to both inequalities together while correcting for chance.

²⁸ Cruz *et al.* (2015) argued that – given the truth of the Equation, where $P(\text{if } A, \text{ then } C)$ is interpreted as $P(C|A)$ – participants have to respond with $P(\text{if } A, \text{ then } C) \geq P(A\&C)$ to be probabilistically coherent. From $P(A\&C) = P(C|A) * P(A)$ and $0 \leq P(A) \leq 1$ it follows that $P(C|A) \geq P(A\&C)$.

²⁹ Assuming that a response produced by any other process or a random response has an equal chance of falling on any point of the response scale, the probability of selecting a response greater than $P(A,C)$ amounts to $1 - P(A,C)$.

Table 1. Frequency of probabilistically coherent and-to-if inference (and corresponding percentages).

	P(conclusion) \geq P(A,C)			P(C A) \geq P(A,C)	
	True	Chance	Δ	True	Δ
P(conclusion) = P(If A, C)	1511 (69%)	59%	9%	1713 (78%)	19%
PO	634 (87%)	45%	41%	565 (77%)	32%
NE	481 (66%)	71%	-5%	591 (81%)	10%
IR	396 (54%)	62%	-8%	557 (76%)	14%
P(conclusion) = P(Even if)	1516 (77%)	59%	18%	1557 (79%)	20%
PO	515 (78%)	47%	31%	508 (77%)	30%
NE	482 (73%)	70%	3%	518 (79%)	9%
IR	519 (79%)	60%	19%	531 (81%)	20%

Note. “True” gives raw probabilistic coherence, “Chance” gives probabilistic coherence based on uniform responses, and “ Δ ” their difference. Value for conformity to $P(C|A) \geq P(A,C)$ are given in the two rightmost columns.

The statistical analysis of these data followed Singmann, Klauer, and Over’s (2014; see also Evans, Thompson, & Over, 2015). We first coded coherent/conforming responses in which either $P(\text{Conclusion})$ or $P(C|A)$ was at least as large as that of the premise with ‘1’ and incoherent responses/non-conforming with ‘0’. To implement the chance baseline, we subtracted 1 minus the probability of the premise from this value. We then estimated a LMM with this chance corrected violation score (in which values above 0 indicate coherent/conforming responding above chance) as dependent variable and relevance condition as well as type of probabilistic measure (coherence versus conformity) and their interaction as independent variables separately for the indicative and concessive conditional groups. We thus had two LMMs in total; one for each type of conditional group (i.e., indicative and concessive). We again estimated crossed random effects for participants and scenarios with maximum random slopes (i.e., by-participant and by-scenario random slopes for all fixed effects plus correlations among the slopes).

For the indicative conditionals the statistical analysis confirmed our prediction that relevance affects probabilistic coherence. It also affected probabilistic conformity, but to a lesser degree. All effects of the LMM were significant (including the intercept), most importantly the interaction of relevance condition and type of probabilistic measure, $F(2, 15.15) = 25.15, p < .0001$. It indicated that coherence was only above

chance for PO ($\beta = 0.42$, 95% CI [0.33, 0.50]), but not for NE ($\beta = -0.05$, 95% CI [-0.11, 0.00]) and IR ($\beta = -0.07$, 95% CI [-0.14, -0.00]). In contrast, conformity was above chance for all three conditions (smallest β for NE = 0.09, 95% CI [0.05, 0.14]). Note also that for both types, PO was larger than NE and IR (all $ps < .0001$), while the latter two did not differ from each other ($ps > .43$).

For the concessive conditionals we found both a significant intercept indicating general above chance responses, $\beta = 0.19$, 95% CI [0.15, 0.23], $F(1, 21.37) = 78.39$, $p < .0001$, and an effect of relevance condition, $F(2, 13.05) = 18.75$, $p = .0001$, but no further effects (all remaining $p > .22$) indicating that type of probabilistic measure had no effect for the concessive conditionals. For the main effect of relevance, all three relevance conditions differed significantly from each other, all $p < .004$, but coherence and conformity were significantly above chance in each case ($\beta_{NE} = 0.06$, 95% CI [0.01, 0.12], $\beta_{IR} = 0.19$, 95% CI [0.14, 0.25], and $\beta_{PO} = 0.30$, 95% CI [0.23, 0.37]).

As Table 1 indicates, the proportions of coherent or conforming responses were around 78% across relevance conditions in all cases except for probabilistic coherence for indicative conditionals. Table 1 indicates that all further differences in the statistical analysis are solely driven by different sizes of the chance intervals. This finding thus corroborates the result from Skovgaard-Olsen *et al.* (2016) that $P(\text{if } A, \text{ then } C) \neq P(C|A)$ for negative relevance or irrelevance but that $P(\text{Even if } A, \text{ then still } C) = P(C|A)$ across all relevance levels.

Our results extend Cruz *et al.* (2015), who found that participants were probabilistically coherent above chance levels overall. However, they employed stimulus material inspired by the Linda problem from Tversky and Kahneman's (1983) work on the conjunction fallacy, which implements only the PO condition for the indicative conditional at one specific priors level ('If Linda votes in the municipal elections, then she votes for the Socialist Party').³⁰ In contrast, our results are based on all permutations of the relevance and priors levels for both indicative and concessive conditionals.

³⁰ Nicole Cruz (personal communication, 20.05.2016).

Interestingly, Tentori, Crupi, and Russo (2013) showed that the prevalence of the conjunction fallacy also depends on whether or not the information the participants are given in the scenario is positively relevant for the second conjunct. Tentori *et al.* (2013) interpret their results as showing that the participants committing the conjunction fallacy tend to substitute a sound estimation of confirmation relations for the intended probability assignment. This introduces the possibility that a similar cognitive mechanism may be implemented in the participants' lack of conformity to $P(\text{conclusion}) \geq P(\text{premise})$ above chance levels in the NE and IR conditions.

General Discussion

In this study we manipulated relevance and prior probabilities using a new cluster of scenarios (see supplementary materials). Experiment 1 presented evidence for high agreement between ΔP and ratings of perceived relevance and reason relations and suggests that ΔP is a better predictor than the difference measure. Follow up studies might contrast ΔP with further confirmation measures (see Tentori *et al.*, 2007; Douven and Verbrugge, 2012). Interestingly, when removing extreme ($-\infty$) and undefined values, ΔP correlates to a very high degree, $r = .96$, with the following log odds ratio for our data set:

$$\tau(C|A) - \tau(C|\bar{A}) \approx \ln \left(\frac{\frac{P(Y=1|X=1)}{P(Y=0|X=1)}}{\frac{P(Y=1|X=0)}{P(Y=0|X=0)}} \right)$$

The logged odds ratio measure thus accounts for pretty much the same variance as ΔP . The log odds ratio measure is a more direct approximation of Spohn's (2012: ch. 6) ranking-theoretic explication of the reason relation (i.e. $\tau(C|A) - \tau(C|\bar{A}) > 0$) in probability theory, and it was therefore used as a relevance parameter in the logistic regression model of the conditional inference task put forward in Skovgaard-Olsen (2015).

In their relevance theory, Wilson and Sperber (2004; see also Sperber, Cara, Girotto, 1995) propose that maximization of relevance is a general principle

structuring both cognition and communication. Their account introduces an economic aspect to assessments of relevance; the cost of processing information decreases the perceived relevance, whereas the gain in cognitive effects increases the perceived relevance. In principle, it is possible to combine this idea with Spohn's (2012) theory as the latter gives us a precise notion of cognitive effect in terms of difference-making in degrees of belief, and the precise formal principles guiding belief revision, whereas Sperber and Wilson's theory introduces a focus on processing costs.

In Experiment 2, we examined the role of relevance for the uncertain and-to-if inference task presented in Cruz *et al.* (2015). It was found that the participants perform above chance levels for PO, but below chance levels for NE and IR, thus qualifying Cruz *et al.*'s (2015) results. This result is hard to reconcile with probabilistic approaches to the semantics of conditionals that equate $P(\text{if } A, \text{ then } C)$ with $P(C|A)$ (Evans & Over, 2004; Oaksford & Chater, 2007; Pfeifer, 2013). In fact, it presents these theories with a dilemma: either $P(\text{if } A, \text{ then } C) = P(C|A)$ does not hold across relevance manipulations, or the participants are less probabilistically coherent than initially seemed to be the case.

The Equation ($P(\text{if } A, \text{ then } C) = P(C|A)$) here acts as an auxiliary assumption that implies $P(\text{conclusion}) \geq P(\text{premise})$, if the participants are to be probabilistically coherent. Throughout the last 10-15 years, the Equation has been supported time after time (see Douven, 2015: ch. 3-4). However, relevance levels dramatically moderate this relationship (Skovgaard-Olsen *et al.*, 2016). Accordingly, we suggested that conditionals that violate the default assumption of positive relevance (e.g. 'If the sun is shining in Egypt, then Julia will lose weight') are viewed as defective and penalized in their probability ratings. Based on these results, we suspect that the culprit that makes it appear that the participants are probabilistically incoherent in the NE and IR condition is the Equation. On the alternative outlined in Skovgaard-Olsen *et al.* (2016), the participants rely on a heuristic for assessing reason relations when evaluating $P(\text{if } A, \text{ then } C)$, which introduces no normative requirement that $P(\text{conclusion}) \geq P(\text{premise})$ for the NE and IR conditions, where there is no strong

relationship between $P(\text{if } A, \text{ then } C)$ and $P(C|A)$. The participants' lack of conformity to $P(\text{conclusion}) \geq P(\text{premise})$ above chance levels in the NE and IR conditions is in other words explained by the perceived defect of these conditionals owing to their violation of the expectation that A is a reason for C .

This interpretation is supported by the observation that participants conformed to $P(C|A) \geq P(A,C)$ in their probability evaluations independently of the uncertain and-to-if inference task in $\approx 78\%$ of the cases both in the group with indicative and with concessive conditionals across relevance conditions. Hence, the below chance level performance in the uncertain and-to-if inference task for the NE and IR conditions does not appear to reflect a general failure to conform to $P(C|A) \geq P(A,C)$ across relevance levels.

Aside from the Equation, which introduces the normative requirement that $P(\text{conclusion}) \geq P(\text{premise})$, because $P(\text{Conclusion})$ is treated as $P(C|A)$ and it is a requirement of probability theory that $P(C|A) \geq P(A,C)$, other semantics of conditionals are also committed to $P(\text{conclusion}) \geq P(\text{premise})$ in the and-to-if inference. The reason is that several conditional logics treat ' $A \wedge C \models \text{if } A, \text{ then } C$ ' as a valid argument schema (Arlo-Costa, 2007), and given $P(B) \geq P(A)$, whenever $A \models B$, it holds that $P(\text{if } A, \text{ then } C) \geq P(A \wedge C)$.

One example is Lewis' semantics of counterfactuals, and there is already discussion about whether a weakening of the system should be allowed, which avoids treating ' $A \wedge C \models \text{if } A \text{ were, then } C \text{ would have been}$ ' as a theorem (Kutschera, 1974). It remains to be seen, whether there are differences in the participants' conformity to $P(\text{Conclusion}) \geq P(A \wedge C)$ for indicative and counterfactual conditionals across relevance conditions.

However, it is clear that other theories aside from those endorsing the Equation are faced by explanatory challenges by our results. As Douven (2015: section 2.1-2.2) points out, ' $A \wedge C \models \text{if } A, \text{ then } C$ ' is valid for semantics of indicative conditionals such as the material conditional, Stalnaker's possible worlds semantics, and three-valued truth tables like the de Finetti table. Yet it is rejected by Inferentialism, which holds that it is

part of the truth conditions of indicative conditionals that there is an inferential relation connecting A and C.

References

- Arlo-Costa, Horacio (2007). The Logic of Conditionals. *The Stanford Encyclopedia of Philosophy* (spring 2016 Edition), Edward N. Zalta (ed.). URL = <http://plato.stanford.edu/archives/fall2016/entries/logic-conditionals/>.
- Baratgin, J., Over, D. E., & Politzer, G. (2013). Uncertainty and the de Finetti tables. *Thinking & Reasoning*, 19(3), 308-28. doi: 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. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*. URL = <http://arxiv.org/abs/1406.5823>
- Brandom, B. (1994). *Making it Explicit*. Cambridge, Mass.: Harvard University Press.
- Brewer, B. (2002). *Perception and Reason*. Oxford: Oxford University Press.
- Cruz, N., Baratgin, J., Oaksford, M. and Over, D.E. (2015). Bayesian reasoning with ifs and ands and ors. *Front. Psychol.*, 6, 192. doi: 10.3389/fpsyg.2015.00192
- Douven, I. (2015). *The Epistemology of Indicative Conditionals. Formal and Empirical Approaches*. Cambridge: Cambridge University Press.
- Douven, I. & Verbrugge, S. (2012). Indicatives, concessives, and evidential support. *Thinking & Reasoning*, 18(4), 480-99. doi: [10.1080/13546783.2012.716009](https://doi.org/10.1080/13546783.2012.716009)
- Elqayam, S. and Over, D. E. (2013). New paradigm psychology of reasoning: An introduction to the special issue. *Thinking & Reasoning*, 19:3-4, 249-265.
- Evans, J. St. B. T. and Over, D. (2004). *If*. Oxford: Oxford University Press.
- Evans, J. S. B.T., Thompson, V. A., and Over, D. E. (2015). Uncertain deduction and conditional reasoning. *Front. Psychol.*, 6: 398. doi: 10.3389/fpsyg.2015.00398

- Falk, R. & Bar-Hillel, M. (1983). Probabilistic dependency between events. *Two-Year College Mathematics Journal*, 14, 240-7.
- Kutschera, F. (1974). Indicative Conditionals. *Theoretical Linguistics*, 1, 257-69.
- McDowell, J. (1994). *Mind and World*. Cambridge, Mass.: Harvard University Press.
- Oaksford, M. & Chater, N. (2007). *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*. Oxford: Oxford University Press.
- Over, D. and Evans, J. St. B. T. (2003). The Probability of Conditionals: The Psychological Evidence. *Mind and Language*, 18 (4), 340-58. doi: 10.1111/1468-0017.00231
- Pfeifer, N. (2013). The new psychology of reasoning: A mental probability logical perspective. *Thinking & Reasoning* 19(3-4), 329-45. doi: 10.1080/13546783.2013.838189
- R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Singmann, H., Bolker, J. Westfall, S. Højsgaard, J. Fox, M. Lawrence, et al. (2016). afex: Analysis of Factorial Experiments. R package version 0.13-145, Available via <http://cran.rproject.org/package=afex>.
- Singmann, H., Klauer, K. C., and Over, D. (2014). New normative standards of conditional reasoning and the dual-source model. *Front. Psychol.* 5:316. doi: 10.3389/fpsyg.2014.00316
- Skovgaard-Olsen, N. (2015). Ranking Theory and Conditional Reasoning. *Cognitive Science*. doi: 10.1111/cogs.12267
- Skovgaard-Olsen, N., Kellen, D., Krahl, H., and Klauer, K. C. (in review). Relevance differently affects the truth, acceptability, and probability evaluations of 'and', 'but', 'therefore', and 'if then'.
- Skovgaard-Olsen, N., Singmann, H., and Klauer, K. C. (2016). The Relevance Effect and Conditionals. *Cognition*, 150, 26-36. [doi:10.1016/j.cognition.2015.12.017](https://doi.org/10.1016/j.cognition.2015.12.017)
- Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task.

Cognition, 57, 31-95.

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. doi: 10.1111/cogs.12057

Tentori, K., Crupi, V., Bonini, N, and Osherson, D. (2007). Comparison of confirmation measures. *Cognition*, 103, 107-119.

Tentori, K., Crupi, V., & Russo, S. (2013). On the determinants of the conjunction fallacy: Probability versus inductive confirmation. *Journal of Experimental Psychology: General*, 142(1), 235–255. doi:10.1037/a0028770

Tversky, A., and Kahneman, D. (1983). Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. *Psychol. Rev.* 90, 293–315. doi: 10.1037/0033-295X.90.4.293

Walton, D. (2004). *Relevance in Argumentation*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Wilson, D. and Sperber, D. (2004). Relevance Theory. In Horn, L.R. & Ward, G. (eds.), *The Handbook of Pragmatics*. Oxford: Blackwell, 607-632.

Supplementary Materials:

Manipulation Check and Selection of Scenarios

We first performed a manipulation check to ensure that the numbers the participants provided could be interpreted as probabilities satisfying the axioms of the probability calculus. To this end, the law of total probability, $P(C) = \sum_{i=1}^n P(C|A_i)P(A_i)$, was applied to the measurements of $P(A)$, $P(C|A)$, and $P(C|\bar{A})$ to calculate an ideal value that $P(C)$ should take if the participants were probabilistically consistent. This calculated value for $P(C)$ was then subtracted from the actual value of $P(C)$ supplied by the participants to form a probabilistic consistency scale using the following formula: $1 - |P(C) - [P(C|A) \cdot P(A) + P(C|\bar{A}) \cdot (1 - P(A))]|$. This measure takes on values smaller or equal to one, where a value of one indicates perfect probabilistic consistency. Fig. 1 shows the distribution of mean consistency values for the participants in a boxplot and reveals that participants are surprisingly probabilistically consistent with 75% of the distribution having probabilistic consistency rates of almost .9. Given these results we were confident that participants' responses could be interpreted as probabilities.

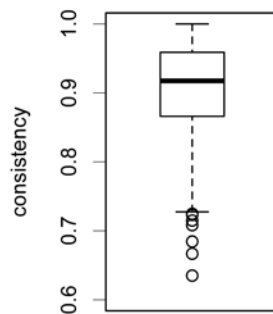


Fig. 1. Probabilistic consistency ratings of the participants based on applying the law of total probability to the probabilities they provided.

Experiment 1: Correlation Matrices

Table 1 displays the inter-correlation of the four variables of Experiment 1 and shows that, as expected, all correlations were highly significant. One can also see that, as hypothesized, ΔP seemed to be a better predictor for both relevance and the reason relation than the difference measure.

Table 1. Correlation Matrix and Descriptive Statistics for Measures Obtained in Experiment 1

	Reason relation	ΔP	Difference Measure	Mean (SD)
Relevance	.81	.54	.40	0.73 (2.29)
Reason Relation		.58	.46	0.06 (1.19)
ΔP			.73	0.01 (0.41)
Difference Measure				-0.01 (0.34)

Note. All correlations are highly significant, $p < .0001$. Given the non-independence of data points within participants and within contents, these p -values should, however, be read with caution. The ranges for the variables are: directional relevance from -4 to 4, reason relation from -2 to 2, ΔP from -1 to 1.

To appropriately test this hypothesis it is important to consider that the data has replicates both on the level of the participant (since each participant provided one response for each of the 12 within-participant conditions) and on the level of the scenarios (as each scenario could appear in each relevance condition across participants). Due to this dependency structure with conditions repeated within participants and scenarios, standard statistical procedure such as correlation cannot be used. For this reason, a linear mixed model was used in the paper for the analysis.

Out of the six confirmation measures mentioned in Tentori *et al.* (2007), our design only allowed us to test the Keynes and Horwich's ratio measure, $\log(P(C|A)/P(C))$ in addition to the difference measure (which is also listed there). Unfortunately, this measure introduces the problem of extreme ($-\infty$) or undefined values for 24% of our observations. Furthermore, it correlates highly with the difference measure for the reduced sample, $r = .89$:

Table 2. Correlation Matrix and Descriptive Statistics for Measures Obtained in Experiment 1

	Reason relation	ΔP	Difference Measure	Ratio	Mean (SD)
Relevance	.83	.54	.40	.37	0.73 (2.29)
Reason Relation		.57	.43	.40	0.06 (1.19)
ΔP			.69	.61	0.01 (0.41)
Difference Measure				.89	-0.01 (0.34)
Ratio					-0.01 (0.80)

Note. All correlations are highly significant, $p < .0001$. Given the non-independence of data points within participants and within contents, these p -values should, however, be read with caution. The ranges for the variables are: directional relevance from -4 to 4, reason relation from -2 to 2, ΔP from -1 to 1.

When adding the ratio measure to our LMM model for Experiment 1, the results indicate that it accounts for no unique variance on its own for either perceived relevance, $F(1, 24.00) = 2.63, p = .12$, or perceived reason relation as DV, $F(1, 24.57) = 2.21, p = .15$. Indeed, it remains the case that of these three predictors, only ΔP accounts for unique variance for perceived relevance, $F(1, 25.14) = 235.84, p < .0001$, and perceived reason relation, $F(1, 22.16) = 216.87, p < .0001$.

Selection of the Scenarios

For the selection of the scenarios, the full sample of 725 participants was used without applying our exclusion criteria, and as there were no significant differences between the IR_S and IR_D conditions, the difference between them was collapsed for the analysis.

The distinction between these two ways of implementing the irrelevance category was initially introduced in an attempt to implement the notion of ‘topical relevance’ from relatedness logic (Iseminger, 1986; Walton, 2004: ch. 4), which treats two propositions as relevant if they share a subject matter and as irrelevant if they don’t. The way we operationalized this requirement was that two propositions are judged to be relevant, if they concern the same context/content and irrelevant if they didn’t. Accordingly, if Stephen is going on a date, then we assumed that if two propositions (A, C) both concern preparations for the dating situation then they will

share a subject matter, whereas a proposition concerning what Stephen's neighbor likes to eat (B) concerns a different subject matter. Under this assumption, A and C are topically relevant to each other, whereas A and B are topically irrelevant to each other. However, as there were no significant differences between the IR_S and IR_D conditions, the difference between them was collapsed for the analysis, and the IR_D conditions of our stimulus materials were selected for Experiment 2.

To prevent scenario content from becoming a nuisance variable only complete scenarios were selected so that we could ensure that all experimental conditions were represented within each scenario. For each experimental condition, the outputs of the following three equations were z-transformed and the average was taken. This average was used to calculate the 30th percentile with the largest distance from *optimal*: (1) $(\overline{\Delta P} - \textit{optimal})^2$, (2) $(\overline{P(A)} - \textit{optimal})^2$, and (3) $(\overline{P(C)} - \textit{optimal})^2$. To illustrate, for the experimental condition IRHH, the optimal value of ΔP would be 0 and the optimal values of P(A) and P(C) would be 100. So to ensure that our selected scenarios were able to implement this experimental condition, the distance of the average ΔP , P(A), and P(C) from these optimal values was used as a selection criterion.

For each scenario, the frequency of its experimental conditions lying within the 30th percentile of the worse experimental conditions was counted. The 30th percentile with the largest number of bad experimental conditions was then used to exclude six complete scenarios. That is to say, scenarios with five or more counts of worse experimental conditions were excluded. The mean frequency of the worse experimental manipulations for excluded scenarios was 5.83, and the mean frequency of worse manipulations for included scenarios was 2.6. In one case, a choice had to be made between two scenarios that both had 5 worse manipulations using boxplots.

In Table 3, summary statistics is shown for the excluded {3, 4, 6, 9, 15, 18} and included scenarios {1, 2, 5, 7, 8, 10, 11, 12, 13, 14, 16, 17}. With ΔP values of almost 0 on average for the IR conditions and the NE and PO conditions differing with ΔP values above $|.25|$ from the IR conditions in the expected directions, the relevance manipulations were successfully implemented. Moreover, with high and low prior

manipulations differing on average with $|.20|$ or more from the midpoint of the scale, the prior manipulations were also successfully implemented.

Table 3. Summary statistics of selected scenarios.

	Included	Excluded
PO ΔP mean	.32	.22
NE ΔP mean	-.27	-.21
IR ΔP mean	-.01	.020
Mean high P(A)	.70	.63
Mean low P(A)	.15	.15
Mean high P(C)	.77	.70
Mean low P(C)	.27	.30

However, because complete scenarios were selected some outliers had to be accepted in particular scenarios, which are still in need of further improvement.

References

- Iseminger, G. (1986). Relatedness Logic and Entailment. *The Journal of Non-Classical Logic*, 3, 5-23.
- Walton, D. (2004). *Relevance in Argumentation*. Mahwah, N.J.: Lawrence Erlbaum Associates.

4. RELEVANCE DIFFERENTLY AFFECTS THE TRUTH, ACCEPTABILITY, AND PROBABILITY EVALUATIONS OF ‘AND’, ‘BUT’, ‘THEREFORE’, AND ‘IF THEN’³¹

Niels Skovgaard-Olsen

University of Konstanz and Albert-Ludwigs-Universität Freiburg

David Kellen

Syracuse University

Hannes Krahl

University of Chemnitz

Karl Christoph Klauer

Albert-Ludwigs-Universität Freiburg

Skovgaard-Olsen, N., Kellen, D. Krahl, H., Klauer, K. C. (*forthcoming*).

Relevance differently affects the truth, acceptability, and probability
Evaluations of ‘And’, ‘But’, ‘Therefore’, and ‘If Then’. (*Thinking & Reasoning*).

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).

³¹ We are very grateful for discussions with Igor Douven, Wolfgang Spohn, Karolina Krzyżanowska, Peter Collins, Ulrike Hahn, Maria Biezma, Ulrike Hahn, and the audiences of talks at the annual meeting of New Frameworks of Rationality, and the What-If group in Konstanz. Moreover, careful comments by the reviewers substantially improved the manuscript.

Abstract

In this study we investigate the influence of reason relation readings of indicative conditionals and 'and', 'but', and 'therefore' sentences on various cognitive assessments. According to the Frege-Grice tradition, a dissociation is expected. Specifically, differences in the reason-relation reading of these sentences should affect participants' evaluations of their acceptability but not of their truth value. In two experiments we tested this assumption by introducing a relevance manipulation into the truth-table task as well as in other tasks assessing the participants' acceptability and probability evaluations. Across the two experiments a strong dissociation was found. The reason relation reading of all four sentences strongly affected their probability and acceptability evaluations, but hardly affected their respective truth evaluations. Implications of this result for recent work on indicative conditionals are discussed.

Keywords: Relevance, truth conditions, indicative conditionals, probability, acceptability, conjunctions

Introduction

The goal of the present paper is to investigate experimentally which cognitive assessments are affected by salient reason-relation readings of conjoined sentences. Following the tradition in linguistics, philosophy and psychology, we will focus on truth, probability, and acceptability evaluations (Garmut, 1991; McCawley, 1993; Nickerson, 2015). For our target sentences, we have chosen indicative conditionals (e.g. ‘if it rains, then Sally’s birthday party will be ruined’) and ‘But’, ‘And’, and ‘Therefore’ sentences (henceforth referred to as ABT sentences). As we will explain in detail below, a focused comparison between these sentences allows us to investigate which aspects of their meaning are influenced by their reason relation reading.

The standard work of reference in this context is Grice’s (1989) seminal work. In that book, Grice was concerned with enforcing *Modified Occam’s Razor* as a methodological principle, according to which “*Senses are not to be multiplied beyond necessity*” (p. 47). Grice’s central point was that when modeling the meaning of natural language content, we can keep the logic simple by distinguishing between the truth-conditional content of sentences and the different *implicatures* that enrich the truth-conditional content of such sentences. In particular, Grice (1989) was concerned about things such as preventing that:

- 1) we reject the truth table of the material implication (\supset) as a model of the truth conditional content of indicative conditionals (see Table 1 below) based on the fact that natural language conditionals have a prominent reason relation reading, according to which the antecedent is a reason *for* the consequent,
- 2) we reject the truth table of the logical conjunction (\wedge) based on the fact that natural language conjunctions have readings that indicate temporal succession and causal relations, and

- 3) we reject the logical conjunction as a model of the truth conditional content of sentences involving 'but' and 'therefore' based on the salient reason relation reading of natural language sentences in which they occur.

A key distinction proposed by Grice (1989) was *conversational* and *conventional implicatures*. These twin notions provided the means to explain how these various sentences could have a meaning that goes beyond their truth conditional content. The notion of a conversational implicature was called upon to explain the pragmatic phenomenon that we often use words to convey a meaning that differs from what we literally say. For instance, when Julia responds that "I don't like parties" to the question "Are you going to the party tonight?", her interlocutor can infer that the message that Julia intends to convey is that she won't go to the party tonight (Blome-Tillmann, 2013). Grice's (1989) project was to reconstruct such pragmatic inferences to the intended meaning rationally, based on maxims of informative, truthful, relevant, and clear communication that implement the goal of cooperative discourse.

In addition, Grice (1989) also introduced the category of conventional implicatures to account for the cases where it is not a feature of the utterance of a sentence in a particular context that invites an interpretation of the intended meaning that differs from the truth conditional content of the sentence. In particular, Grice thought that it was part of the conventional meaning of 'therefore' to introduce a consequence relation (e.g. 'He is an Englishman; he is, *therefore*, brave'), which goes beyond its truth conditions as modeled by the logical conjunction.

Another example is that sentences such as 'she was poor *but* honest' have been thought to express a contrast between being poor and honest since Frege (1892), which is not to be found in its truth-conditional equivalent 'she was poor *and* honest'. Often the treatment of 'but' is left with this observation in the philosophical literature. In fact, the analysis of 'but' is a rich topic in linguistics, where at least four different readings are dissected. However, attempts have been undertaken to subsume these various readings under the prominent denial of expectation reading, where an implicit

or explicit assumption is denied by the second clause (Iten, 2000, Chap. 5; Blackmore, 2004, Chap. 4), which is closely related to a reason relation reading.

Common to both conversational and conventional implicatures is the expectation that their content will not affect the truth-conditional content of the sentences in which they occur (Potts, 2015). Conversational implicatures and conventional implicatures differ in that only conversational implicatures can be cancelled (e.g. adding the qualification “I don’t like parties, but I would be happy to come tonight” cancels the conversational implicature that Julia won’t attend the party tonight). Furthermore, only conventional implicatures can be detached, or removed, by uttering a different sentence with the same truth-conditional content in the context of utterance (e.g. uttering “she was poor *and* honest” instead of “she was poor *but* honest”). Finally, only conversational implicatures can be reconstructed on the basis of Grice’s maxims of conversation (Blome-Tillmann, 2013).

It is not uncommon to find references to semantic and pragmatic modulation in the psychological literature (Johnson-Laird and Byrne, 2002), or to invoke Grice’s theory to explain pragmatic effects (Nickerson, 2015). Given the prominent role that Grice’s thought continues to have on theorizing about natural language, it is important that its central claims are subjected to empirical tests and do not just figure as *ad hoc* explanations that are invoked to explain divergent results when convenient.

One natural assumption is that meaning components classified as conversational or conventional implicatures influence the acceptability/assertability assessments of the participants, but they should not affect their truth evaluations. This predicts a dissociation between the influence of factors relating to conversational and conventional implicatures on acceptability/assertability³² and truth-value assessments. Although probability assignments were not originally discussed in this context, we would expect them to be vary in tandem with the acceptability judgments.

As noted above, the ABT sentences have been thought to be truth conditionally equivalent in the Frege/Grice tradition. In addition, therefore-sentences have a salient

³² In this paper we will not distinguish the two.

reason relation reading, whereby ϕ is a reason *for* ψ in “Nick forgot about her birthday (ϕ), therefore he didn’t buy Sally a present (ψ)”. In contrast, ϕ appears to be a reason *against* ψ in “Nick dislikes Sally (ϕ), but he attended her birthday party (ψ)”. Finally, and-sentences seem to fall somewhere in between and can be used as a baseline for comparing these two extremes.

In light of these differences in their reason relation readings despite their assumed truth conditional equivalence, comparing the ABT sentences on their truth, probability, and acceptability evaluations will give us important clues about whether the abovementioned dissociation can be found. If the differences in their reason relation reading are attributable to implicatures, then these differences should only show up in their acceptability evaluations and probability evaluations. Yet their truth evaluations should be unaffected. If, on the contrary, the assumption is wrong that the ABT sentences only differ in their implicatures, then we should see evidence of their different reason relations readings affecting their truth evaluations.

There is a long tradition in the psychology of reasoning of investigating truth conditions by presenting participants with the cells of truth tables in the *truth-table task* (for reviews, see Manktelow, 2012; Nickerson, 2015). However, in order to address the question of what range of cognitive evaluations the reason-relation reading affects, a measurement tool is needed to capture the presence and absence of specific reason relations and to be able vary it orthogonally to the presence or absence of other psychological factors of interest.

Reason Relations

For the purpose of this paper, we will rely on Spohn’s (2012, Chap. 6) explication of reason relations. According to Spohn (2012: ch. 6), the reason relation and the notion of epistemic relevance can be explicated as follows:

THE ΔP RULE: $\Delta p = P(\psi \mid \phi) - P(\psi \mid \neg\phi)$

POSITIVE RELEVANCE/ ϕ IS A REASON FOR ψ : $\Delta p > 0$

IRRELEVANCE: $\Delta p = 0$

NEGATIVE RELEVANCE/ ϕ IS A REASON AGAINST ψ : $\Delta p < 0$

The underlying intuition here is that relevant information changes the probability of the propositions that it concerns. The probabilistic change is here explicated through a comparison between conditioning on the information and conditioning on its negation. When such a comparison shows that ϕ increases the probability of ψ , then ϕ is said to be a *reason for* ψ . When it shows that ϕ decreases the probability of ψ , then ϕ is said to be a *reason against* ψ . Using these explications, we specify Grice's conventional implicature hypothesis as follows: sentences containing ' ϕ therefore ψ ' differ from sentences containing the connective ' ϕ and ψ ' in suggesting that ϕ is a *reason for* ψ ($\Delta p > 0$), and sentences containing ' ϕ but ψ ' differ from the latter in suggesting that ϕ is a *reason against* ψ ($\Delta p < 0$).

In Skovgaard-Olsen, Singmann, and Klauer (2016b), participants' perceived reason relations and perceived relevance were investigated for a large range of everyday contexts, and empirical support for the explication of the reason relation above could be obtained. The stimulus materials used in the present study contains 12 scenarios with 12 different conditions, implementing all permutations of positive relevance (PO), negative relevance (NE), irrelevance (IR) and high (H) and low (L) prior probability within each scenario. For Δp values ranging from -1 to 1, a pretest with 725 participants showed that the average Δp was .32 for the *positive relevance conditions*, -.27 for the *negative relevance conditions*, -.01 for the *irrelevance conditions*. Moreover, the pretest found that the average prior probability was ca. 70% for the high probability items and less than 30% for the low probability items.

The scenarios used were designed to trigger the participants' stereotypical assumptions about basic causal, functional, or behavioral information to implement the above relevance categories. For instance, one scenario text runs as follows:

Scott was just out playing with his friends in the snow. He has now gone inside but is still freezing and takes a bath. As both he and his clothes are very dirty, he is likely to make a mess in the process, which he knows his mother dislikes.

As verified in Skovgaard-Olsen *et al.* (2016b), the two sentences ‘Scott turns on the warm water’ and ‘Scott will be warm soon’ both have a high prior probability, with the first sentence raising the probability of the second (PO HH). Moreover, the sentence ‘Scott turns on the cold water’ has a low prior probability and it lowers the probability of ‘Scott will be warm soon’ (NE LH). Finally, in the absence of further information it seems reasonable to assume that ‘Scott’s friends are roughly the same age as him’ has a high prior probability. Yet this sentence leaves the probability of the previous sentences unchanged and is therefore irrelevant for them. Now if ‘ ϕ therefore ψ ’ expresses positive relevance, then the sentence ‘Scott turns on the warm water *therefore* Scott will be warm soon’ should sound fine, and if ‘ ϕ but ψ ’ expresses negative relevance then the sentence ‘Scott turns on the warm water *but* Scott will be warm soon’ should sound rather strange. Conversely, the negative relevance version should sound better with ‘but’ (‘Scott turns on the cold water *but* Scott will be warm soon’) yet strange with ‘therefore’ (‘Scott turns on the cold water *therefore* Scott will be warm soon’).

One area in which these explications have already been applied is in the empirical research on conditionals to which we now turn. As we shall see, indicative conditionals are another type of sentences which have a salient reason-relation reading, and there is currently a large theoretical interest in this field in diagnosing for which types of cognitive assessments the reason-relation reading plays a role. Accordingly, we will be interested in whether the same kind of dissociations predicted above with respect to the ABT sentences can be found for the cognitive evaluations of indicative conditionals.

Reason Relations and Indicative Conditionals

Throughout the last 10-15 years, a new paradigm has been introduced to the psychology of reasoning (Elqayam & Over, 2013) with researchers turning to probabilistic competence models of reasoning and drawing on Bayesian formal epistemology (Pfeifer & Douven, 2014). In the study of conditionals, this paradigm-

shift is reflected in the widespread endorsement of “the Equation”, the Ramsey Test, and the de Finetti truth table (see Table 1 below).

THE EQUATION: $P(\text{if } \phi, \text{ then } \psi) = P(\psi \mid \phi)$ (Bennett, 2003; Evans & Over, 2004; Oaksford & Chater, 2007).

THE RAMSEY TEST: Instead of calculating conditional probabilities by means of the ratio $P(\phi \text{ and } \psi) / P(\phi)$, the participants are conjectured to evaluate conditional probabilities on the basis of the Ramsey test. The Ramsey test is a mental algorithm that temporarily adds the antecedent to the participant’s stock of beliefs, makes changes to preserve consistency, and evaluates the consequent under its supposition (Evans & Over, 2004; Oaksford & Chater, 2007; Pfeifer, 2013).

Table 1. Truth Tables of the Indicative Conditional

Truth-Conditional Inferentialism				
	TT	T⊥	⊥T	⊥⊥
PO	T	T	T	T
NE	⊥	⊥	⊥	⊥
IR	⊥	⊥	⊥	⊥
Material Implication Account				
PO	T	⊥	T	T
NE	T	⊥	T	T
IR	T	⊥	T	T
De Finetti Table				
PO	T	⊥	void	void
NE	T	⊥	void	void
IR	T	⊥	void	void

Note. ‘T’ = true; ‘⊥’ = false; PO = positive relevance; NE = negative relevance; IR = irrelevance.

As discussed by Elqayam and Over (2013), there are two direct sources of evidence supporting this approach to conditionals. First, it has long been known that the participants tend to judge the false antecedent cells of the truth table (⊥T, ⊥⊥) to be ‘irrelevant’ to the truth or falsity of the indicative conditional in the truth-table task (an effect known as “the defective truth table”). This defective truth table effect has been interpreted as direct evidence in favor of the *de Finetti truth table* (see Table 1), because it is hard to reconcile with accounts based on the material implication, which

would require that the conditional is treated as true in the false antecedent cells (Over & Evans, 2003). Second, direct investigations of the probability of indicative conditionals have repeatedly supported the Equation. For example, $P(\psi \mid \phi)$ turns out to be a much better predictor of $P(\text{If } \phi, \text{ then } \psi)$ than $P(\neg\phi \vee \psi)$, the probability under the material implication interpretation (Over, Hadjichristidis, Evans, Handley, & Sloman, 2007; Douven, 2015b, Chapters 3 and 4).

Despite this empirical support for The Equation, recent results have shown, however, that the relationship between $P(\text{if } \phi, \text{ then } \psi)$ and $P(\psi \mid \phi)$ is moderated by the relevance manipulation reviewed above. Employing the above-described stimulus material, Skovgaard-Olsen *et al.* (2016a) showed that the marginal means of assessments of $P(\text{if } \phi, \text{ then } \psi)$ were judged to be substantially lower in NE and IR than in PO and that the slopes of the regression lines employing $P(\psi \mid \phi)$ as a predictor of $P(\text{if } \phi, \text{ then } \psi)$ were steeper in the PO condition. These results corroborate the idea that there is something defective about indicative conditionals that violate the default assumption of positive relevance. Thus whereas the sentence 'If Scott turns on the warm water, then Scott will be warm soon' (PO HH) has a high probability in the scenario above, sentences like 'If Scott's friends are roughly the same age as him, then Scott will turn on the warm water' (IR HH) and 'If Scott makes an effort to be tidy, then the bathroom will be dirtier than before he took his bath' (NE HH) have a much lower probability than what would be expected based on $P(\psi \mid \phi)$ alone.

In earlier empirical work (Oberauer, Weidenfeld, & Fischer, 2007; Over, Hadjichristidis, Evans, Handley, and Sloman, 2007; Singmann, Klauer, & Over, 2014), the Δp rule was investigated in relation to probability evaluations of conditionals, with little or no evidence for a relationship being found. But the crucial difference from Skovgaard-Olsen *et al.*'s (2016a) original study consists in that these other studies did not systematically vary Δp to be negative, equal to zero and positive for realistic stimulus materials. In addition, Skovgaard-Olsen *et al.* (2016a) did not assume that $P(\text{if } \phi, \text{ then } \psi)$ would be directly predicted by either Δp or $P(\psi \mid \neg\phi)$ as its proxy. Instead a heuristic for judging $P(\text{if } \phi, \text{ then } \psi)$ based on a reason relation assessment was

formulated according to which the presence of negative relevance or irrelevance would make the participants apply a penalty to $P(\text{if } \phi, \text{ then } \psi)$ based on the conditionals' perceived defect in expressing a reason relation.

In philosophy, it has long been argued that indicative conditionals without connection between the antecedent and the consequent (e.g. 'If Copenhagen is in Denmark, then H. C. Anderson is dead') are defective (Spohn, 2013; Olsen, 2014; Douven, 2015a; Krzyżanowska, 2015). This intuition has not, however, been integrated into contemporary treatments of conditionals in the psychology of reasoning based on either the classical treatment of conditionals as the material implication (' \supset '), nor by probabilistic accounts based on the de Finetti truth table and the Ramsey test (Skovgaard-Olsen, Singmann, and Klauer, 2016a).

As an alternative, Krzyżanowska, Wenmackers, and Douven (2014) and Krzyżanowska (2015) argued that there should be an inferential relation (or reason relation) between the antecedent and the consequent as part of the truth conditions of indicative conditionals. In Krzyżanowska (2015, p. 62), the presence of either an indicative, abductive, or deductive inferential relation between the antecedent and the consequent is made part of the truth conditions of indicative conditionals. Moreover, they allow for auxiliary assumptions coming from the participants' background knowledge to play a role in determining the inferential relation and add further qualifications that need not concern us here. Based on this line of reasoning, *truth-conditional inferentialism* predicts that the modal truth value assignments should follow the first rows in Table 1.

One thing to note about the 'True' prediction for the $T\perp$ cell is the following:³³ according to Krzyżanowska (2015: Chap.. 3), the inferential relation between ϕ and ψ admits exceptions (i.e. the $T\perp$ cells) via inductive or abductive consequence relations.

³³ Both Karolina Krzyżanowska and Igor Douven have confirmed that Table 1 is a reasonable explication of the definition cited above (personal communication, February, 2016). However, Karolina Krzyżanowska did express some doubts about her earlier proposed definition and expressed concerns that it is less clear whether truth conditional inferentialism is really committed to predicting True for the $T\perp$ cell in the positive relevance condition.

Inductive consequence relations would be instances of probability-raising based on purely frequentist information. Abductive consequence relations would be instances of probability raising based on explanatory considerations (e.g. causal structures and theoretical assumptions). Most likely, the stimulus material reviewed above and further specified in Table 2 below instantiates the abductive consequence relation. But crucially, it is not based on the deductive consequence relation, which does not admit of exceptions (i.e. the $T\perp$ cells).

As pointed out by Douven (2015a; 2015b), a main distinguishing feature of truth-conditional inferentialism is the rejection of the *and-to-if* inferences (i.e. inferring 'if ϕ , then ψ ' from ' ϕ and ψ '), which is valid under the other main semantics of conditionals. This inference is not supported by truth-conditional inferentialism, given ϕ is not a reason for ψ in the TT cell for the IR and NE condition.

Note that whereas the Frege-Grice tradition is based on ignoring relevance differences between sentences in truth value evaluations, truth-conditional inferentialism stands out by making predictions that directly vary with the levels of the relevance factor. In contrast, both the *material implication* and the *de Finetti truth tables* follow the Frege-Grice tradition in assigning truth conditions to indicative conditionals that are invariant across the different levels of the relevance factor (see Table 1).

For both the ABT sentences and indicative conditionals, the main focus of our experiments is to investigate whether evidence for the hypothesized dissociation can be found that the reason relation readings of these sentences affect the acceptability and probability evaluations of these sentences, but not their truth evaluations. In order to investigate these issues, Experiment 1 sets out to introduce the relevance manipulation into the truth-table task and Experiment 2 introduces the relevance manipulation into tasks that elicit the probability and acceptability judgments of our four target sentences.

Experiment 1

As discussed above, it is part of the dissociation predicted by the Frege-Grice tradition that the truth table of the logical conjunction ' \wedge ' fits the truth tables ' ϕ and ψ ', ' ϕ but ψ ', and ' ϕ therefore ψ ', and that no effect of the relevance condition can be found for these three sentences (a hypothesis we denote as ' H_{0_ABT} '). Similarly, the material implication and de Finetti truth tables predict that no effect of the relevance condition should be found for 'if ϕ , then ψ ' (a hypothesis we denote as ' H_{0_IF} ') and that the participants' truth evaluations fit their respective truth tables (see Table 1). In contrast, truth-conditional inferentialism predicts that a relevance effect on the truth evaluations of 'if ϕ , then ψ ' can be found, and that its respective truth table accurately describes participants' modal responses.

Method

Participants

The experiment was conducted over the Internet to obtain a large and demographically diverse sample. A total of 752 people completed the experiment. The participants were sampled through Mechanical Turk from USA, UK, and Australia and were paid a small amount of money for their participation.

The following exclusion criteria were used: 1) not having English as native language, 2) failing to answer two simple SAT comprehension questions correctly in a warm-up phase, 3) completing the task in less than 160 seconds or in more than 3600 seconds, and 4) answering 'not serious at all' to the question of how serious the participant would take his or her participation at the beginning of the study. The final sample consisted of 557 participants. Mean age was 38 years, ranging from 18 to 75 years; 36 % of the participants were male; 72 % indicated that the highest level of education that they had completed was an undergraduate degree or higher. The demographic measures of the participants differed only minimally before and after the exclusion.

Design

The experiment involved a within-subject design. Specifically, there were three factors that were varied within participants: 1) sentence, with four levels: ‘...and...’, ‘...but...’, ‘...therefore...’, ‘if..., then...’; 2) relevance, with three levels: PO, NE, and IR; and 3) priors, with four levels: HH, HL, LH, and LL (e.g. LH indicates that $P(\phi)$ = low and $P(\psi)$ = high). The prior manipulation had the goal of ensuring that the participants’ truth evaluations were representative across different combinations of prior probabilities of the sentences in question.

Materials

The twelve within-participants conditions crossing the factors relevance and priors were randomly assigned to twelve different scenarios for each participant. Within each relevance level, each participant saw all four sentence levels randomly distributed across the priors manipulation. One participant might thus see the sentences ‘...and...’, ‘...but...’, ‘...therefore...’, ‘if..., then...’ in the PO level in the HH, LH, LL, HL prior levels, whereas the next would see them in a different permutation of the priors factor.

With minor adjustments, the twelve scenarios used in this study were obtained from a large pre-study (Skovgaard-Olsen *et al.*, 2016b).³⁴ From each of the twelve selected scenarios we were able to construe all twelve within-participant conditions. Consequently, mapping of the condition to each possible scenario was completely randomized for each participant anew.

³⁴ The minor adjustments in question concern slight formulation changes to a few of the sentences and changing the temporal structure of all the sentences for the present experiment. Whereas the sentences in Skovgaard-Olsen *et al.* (2016b) had the temporal form of ‘if ϕ occurs, then ψ will occur’, their temporal form in the present study was ‘if ϕ is now happening, then ψ will occur’ (or: ‘ ϕ is now happening *and/but/therefore* ψ will occur’). The latter temporal form was introduced in the present study to allow for the introduction of a reversal condition to test for violations of commutativity, which was later dropped prior to launching the experiment, however.

To better illustrate these differences, Table 2 contains all of the experimental conditions for the ‘Scott scenario’ presented in the Introduction here illustrated using the connective ‘And’.

Table 2. Stimulus Materials, Scott Scenario illustrated with And-Sentences

	PO	NE	IR
HH	Scott is now turning on the warm water AND he will be warm soon.	Scott is now making an effort to be tidy AND the bathroom will be dirtier than before he took his bath.	Scott’s friends are now also going home to take a bath AND Scott will turn on the warm water.
HL	Scott is now making an effort to be tidy AND the bathroom will be just as clean as before he took his bath.	Scott is now turning on the warm water AND he will soon start to freeze even more.	Scott’s friends are now also going home to take a bath AND Scott will turn on the cold water.
LH	Scott is now bathing in a hot spring AND he will be warm soon.	Scott is now turning on the cold water AND he will be warm soon.	Scott’s friends are now participating in the Winter Olympics AND Scott will turn on the hot water.
LL	Scott is now turning on the cold water AND he will soon start to freeze even more.	Scott is now bathing in a hot spring AND he will soon start to freeze even more.	Scott’s friends are now participating in the Winter Olympics AND Scott will turn on the cold water.

Note. PO = positive relevance; NE = negative relevance; IR = irrelevance.

The complete list of stimulus materials, R-scripts, and raw data will be available on the *Open Science Framework*: <https://osf.io/yder9/>.

Modified Truth Table Task

For each of the twelve priors×relevance within-participants conditions stemming from our experimental design, the participants were presented with two pages. The first page featured a modified version of the truth-table task. In typical implementations of the truth-table task participants are asked to evaluate the truth value of a conditional statement on the basis of an outcome statement describing a cell in the truth table (TT, TF, FT, FF) with either binary or ternary response options (Schroyens, 2010; Nickerson, 2015, pp. 38). In our modified version, the participants were asked for each trial to evaluate the truth value of a randomly chosen sentence from our four target

sentences (' ϕ and ψ ', ' ϕ but ψ ', ' ϕ therefore ψ ', 'if ϕ , then ψ ') on the basis of two randomly chosen truth table cells.

Since none of the truth tables reviewed in the introduction holds that speaker intentions and the Gricean maxims should play a role for the truth evaluation of the target sentences, we decided to test the participants' truth evaluations under the relevance manipulation in a situation in which they would not have to worry about the speaker intentions behind uttering the strange IR items. To achieve this the participants were instructed that they should consider the target sentences as output produced by a computer program in the development phase in response to the scenario texts as input. A computer program in the development phase does not have any communicative intentions when producing odd sentences. Hence, calibrating its output based on truth values should increase a focus on truth evaluations of the sentences produced solely based on their content. The combination of naturalistic stimulus materials with our computer calibration task meant that the participants were encouraged to set aside concerns about speaker intentions behind the presented assertions and encouraged to use their background knowledge, underlying our manipulation of relevance, in evaluating their content. To illustrate, one participant might have seen the following scenario text:

INPUT: Scott was just out playing with his friends in the snow. He has now gone inside but is still freezing and takes a bath. As both he and his clothes are very dirty, he is likely to make a mess in the process, which he knows his mother dislikes.

with the following PO HH sentence presented as output produced by the computer:

Computer output: Scott is now turning on the warm water BUT he will be warm soon.

To help the participants organize the information, the output sentences were distinguished by a different font, as illustrated above. Following the output sentences, two randomly chosen truth-table cells were presented as continuations of the scenarios which occurred after the computer produced its output:

Continuation: Scott turned on the warm water. He did become warm.

On the basis of this continuation, the computer output turned out to be:

True False Neither true nor false

As shown below, the task of the participants was then to help us calibrate the output sentences of the computer by evaluating separately for each continuation, whether the output sentences were 'True' (T), 'False' (F), or 'Neither true nor false' (NN) given the continuations of the scenario. The exact wording of the instruction was as follows:

On each of 13 pages, you will read, in order, a short text describing a scenario, a sentence, and two different continuations of the scenario. For each case, we ask you to imagine that a computer has been given the scenario text as input and that it produced the sentence as output. The computer is still in the development phase and we need you to help us calibrate its output sentences. For each of the two possible continuations of the scenario, your task is to evaluate whether the sentence produced by the computer turned out to be 'True', 'False', or 'Neither true nor false' by the way the scenario developed.

Before continuing to our 12 experimental conditions, the participants first saw a practice trial, which was later discarded in our the analysis.

On the second page, the participants were instructed to evaluate how confident they were in their responses on a scale from 0 % to 100 %.

Procedures

To reduce the dropout rate once the proper experiment started, participants first went through three pages that: 1) stated our academic affiliations, 2) asked for personalized information (that was not paired with the participants' other responses, however), 3) posed two SAT comprehension questions in a warm-up phase, and 4) presented a seriousness check emphasizing the importance of careful responses for the scientific utility of the results (Reips, 2002). After a practice trial and a repetition of

the instructions, the experiment itself began with the presentation of the twelve within-participants conditions. Their order was randomized anew for each participant.

Results

The observed response frequencies were analyzed with multinomial processing tree models (MPT; Riefer & Batchelder, 1988), a well-known class of models that provides a convenient testbed for hypothesis concerning categorical data. We will evaluate the MPT models' absolute performance via the G^2 statistic (Read & Cressie, 1988) and their relative performance with the Fisher Information Approximation (FIA; Grünwald, 2007). FIA is a model-selection statistic that penalizes models according to their functional flexibility and improves upon traditional statistics such as AIC and BIC. Further details on the models and the analyses are provided in the Appendix. But it is worth to highlight here that the models assumed that individual's responses are stochastic in the sense that they can fail to reflect their true judgments with some probability. When specifying the different hypotheses, we relied on the most lenient stochastic specification, which only imposes the constraint that the preferred response option should be the modal response. For example, in the case of the TT cell, the stochastic implementation of the material implication account then predicts that

$$P(T) \geq P(\perp), P(NN)$$

The reason behind the adoption of this specific stochastic specification is the diagnostic power associated to its failure, as any theory that fails to succeed under these minimal constraints should be seriously questioned.

As can be seen from Figure 1 (right upper panel), aside from the $\perp\perp$ cell, there does not appear to be much of a relevance effect for the truth evaluations of the indicative conditional. Indeed, for the true antecedent cells (TT, T \perp), there appears to be an absolute majority for T and \perp respectively across the different relevance levels, which is shared by all of the *ABT sentences* (' ϕ and ψ ', ' ϕ but ψ ', ' ϕ therefore ψ '). It is

only for the false antecedent cells ($\perp T$, $\perp \perp$) that the indicative conditional seems to stand out from the ABT sentences. For the ABT sentences, there is an absolute majority of \perp responses for the false antecedent cases, whereas there is a mixed response for the indicative conditional with large differences across the relevance levels for the $\perp \perp$ cell with the indicative conditional.

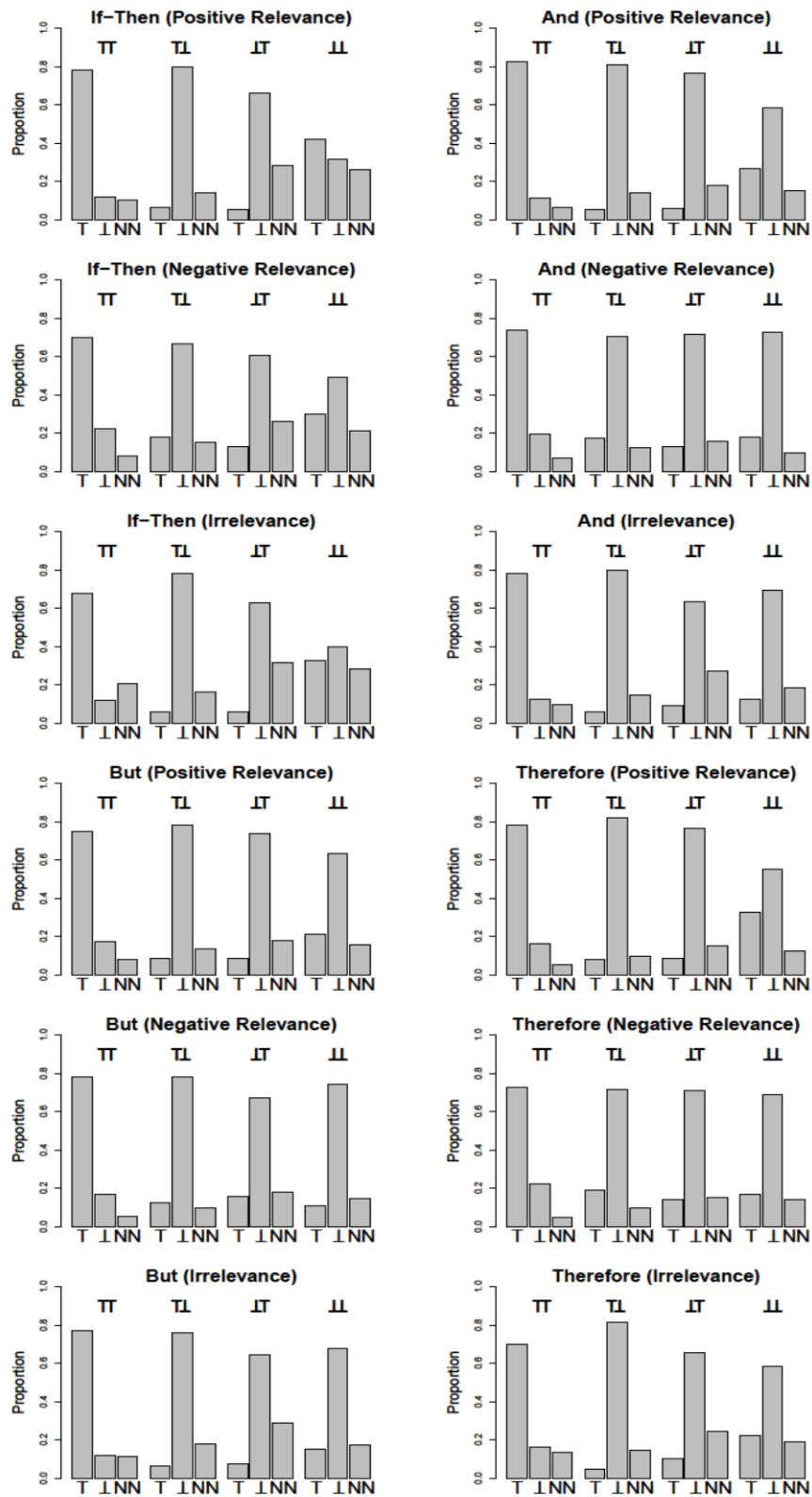


Figure 1. Truth evaluations across sentences, relevance condition, and truth-table cells. Participants could respond 'True' (T), 'False' (\perp), or 'Neither True nor False' (NN).

The ABT Sentences

We first tested whether response distributions differed across sentences and relevance levels. Our general approach for testing different hypothesis was the following: We fitted a set of constrained MPT models representing different hypotheses (e.g., response distributions do not differ across sentence levels), and compared their performance with an unconstrained model ($M_{\text{saturated}}$) that fits the data perfectly using one free parameter per degree of freedom provided by the data. An hypothesis (instantiated by a constrained model) is said to be rejected when it performs worse than the unconstrained model and/or any of the competing alternative hypotheses (even after taking differences in flexibility into account via FIA).

For the ABT sentences, we considered four models: 1) $M_{\text{saturated}}$, which imposes no constraints whatsoever, 2) M_{sentence} , which assumes that response probabilities are the same across sentences, but allows for differences across the different levels of the relevance factor, 3) $M_{\text{relevance}}$, which assumes no differences across the relevance manipulation, but allows responses to differ across sentences, and 4) M_{full} , which assumes no differences across both relevance and sentence levels. As shown in Table 3, the preferred model for the ABT sentences in terms of FIA was clearly M_{sentence} . Note that FIA differences larger than 3.40 already indicate a very strong preference for the winning model (see Kellen, Klauer, & Bröder, 2013). In absolute terms, all models were rejected under a significance level of $\alpha = .05$. However, it should be noted that when large samples are used, any minor deviation from model predictions can lead to a statistically-significant misfit, and having a p -value of .02 for M_{sentence} can be considered satisfactory. In fact, in structural-equation modelling, which faces a similar problem of large samples in the interpretation of χ^2 tests, a ratio χ^2/df between 0 and 2 is considered to indicate a good fit (Schermelleh-Engel, Moosbrugger, & Müller, 2003), and here $G^2/\text{df} = 1.49$.

Table 3. Model-Comparison Results: ABT

Model	G^2	df	p	Δ FIA
$M_{\text{saturated}}$	0	0	1	86.04
M_{sentence}	71.40	48	.02	0
$M_{\text{relevance}}$	237.51	48	.00	83.05
M_{full}	271.76	64	.00	50.84

Note. df = degrees of freedom; G^2 = goodness of fit; p = p-value; Δ FIA = difference between the model's FIA and the FIA from the best-performing model.

ABT Sentences and Indicative Conditionals

When comparing the ABT sentences with the indicative conditionals ('If-Then' sentences in Figure 1), we separated the true antecedent truth table cells from the false antecedent cells. In both cases, we considered three models: 1) $M_{\text{saturated}}$, 2) M_{same} , which assumes that responses to the ABT sentences are equal to the ones given to their indicative-conditional counterparts, and 3) $M_{\text{different}}$, which assumes the same responses rates across the ABT sentences, but allows them to differ from the responses to the indicative conditionals. In the case of the true antecedent cells, M_{same} was the preferred model (see Table 4). In contrast, $M_{\text{different}}$ was preferred in the case of the false antecedent cells (see Table 5).

Table 4. Model-Comparison Results: TT, TF, ABT + IF

Model	G^2	df	p	Δ FIA
$M_{\text{saturated}}$	0	0	1	64.12
M_{same}	57.30	36	.01	0
$M_{\text{different}}$	31.84	24	.13	19.16

Note. df = degrees of freedom; G^2 = goodness of fit; p = p-value; Δ FIA = difference between the model's FIA and the FIA from the best-performing model.

Table 5. Model-Comparison Results: FT, FF, ABT + IF

Model	G^2	df	p	Δ FIA
$M_{\text{saturated}}$	0	0	1	40.98
M_{same}	243.73	36	.00	70.03
$M_{\text{different}}$	39.55	24	.02	0

Note. df = degrees of freedom; G^2 = goodness of fit; p = p-value; Δ FIA = difference between the model's FIA and the FIA from the best-performing model.

Indicative Conditionals and Truth Tables

We now turn to an evaluation of the fit of truth tables in Table 1 for the indicative conditionals. As previously discussed, we will allow the models to predict that the expected responses constitute at least a relative majority (a very lenient requirement). The results reported in Table 6 show that none of the models was able to accurately characterize the individuals' responses. Overall, the modal responses indicate a slight tendency to judge indicative conditionals as true whenever the antecedent and the consequent have the same truth status (in accordance with the truth table of the material bi-conditional, which is true in the TT and $\perp\perp$ cells and false otherwise). This pattern is corroborated by the set of studies gathered by Schroyens (2010), which involved abstract stimulus materials with explicit negations and the option to respond that the truth table cell is 'irrelevant' for the truth value of the conditional (see Table 7).

Table 6. Goodness-of-Fit Results

Truth Conditional Inferentialism				
	TT	T \perp	\perp T	$\perp\perp$
PO	0.00	193.88 (\perp)	171.94 (\perp)	0.00
NE	73.12 (T)	0.00	0.00	0.00
IR	107.41 (T)	0.00	0.00	0.00
Material Implication Account				
PO	0.00	0.00	171.94 (\perp)	0.00
NE	0.00	0.00	96.58 (\perp)	12.64 (\perp)
IR	0.00	0.00	147.56 (\perp)	2.09 (\perp)
De Finetti Table				
PO	0.00	0.00	43.44 (\perp)	10.70 (T)
NE	0.00	0.00	40.47 (\perp)	30.92 (\perp)
IR	0.00	0.00	28.31 (\perp)	5.87 (\perp)

Note. Values correspond to the G^2 statistic. The symbols in parentheses (T, \perp , N) indicate the observed modal response in case the model prediction failed. All values above 2.71 are rejected ($p < .05$) according to the most conservative $\bar{\chi}^2$ distribution (Self & Liang, 1987).

Table 7. Data from Schroyens' (2010) Meta-Analysis

	TT	T⊥	⊥T	⊥⊥
T	100%	0%	5%	64%
⊥	0%	100%	77%	0%
Irrelevant	0%	0%	18%	36%

Note. Values correspond to percentage of studies (out of a subset of 22 studies with abstract stimulus material and 'irrelevant' as third response option) in Schroyens' (2010) meta-analysis in which a specific response was modal.

Confidence Ratings

All the confidence ratings were in the interval [76%, 81%]. For the 'Therefore'-sentences, the participants were more confident in the PO (mean = 80.28, SD = 19.54) condition than in the NE (mean = 78.27, SD = 21.85), $V = 60142$, $p_H < .05$, $r = -.11$,³⁵ and IR (mean = 75.97, SD = 22.60) conditions, $V = 66528$, $p_H < .0001$, $r = -.21$. But these were small effects and the participants continued to remain highly confident in the truth values they provided even when these conflicted with the reason-relation readings of the 'Therefore'-sentences. For the 'But'-sentences, the participants were no more confident in the NE (mean = 78.48, SD = 21.60) condition than they were in the PO (mean = 78.14, SD = 21.89) condition, $V = 54333$, $p_H = .53$, $r = -.027$. For the indicative conditionals, the participants were more confident in the PO (mean = 78.74, SD = 20.65) condition than they were in the IR (mean = 76.01, SD = 22.53) condition, $V = 61801$, $p_H < .05$, $r = -.12$, but no more confident in the PO condition than they were in the NE (mean = 76.42, SD = 21.90) condition, $V = 63538$, $p_H = .074$, $r = -.089$. Again, these were small effects and the participants continued to remain highly confident in the truth values they provided even when these conflicted with the reason-relation readings of indicative conditionals.

³⁵ We controlled for the family-wise error rate using the Bonferroni-Holm correction (indicated by the index "H").

Discussion

The results of Experiment 1 indicate that there are significant effects of relevance on response probabilities in truth evaluations for each of the ABT sentences. But these effects are minor judging from Figure 1. It is possible that some of the small differences observed are due to some degree of heterogeneity at level of the individuals as well as at level of the stimulus material (e.g., Rouder et al., 2008). Moreover, the model comparison in Table 3 shows that the preferred model assumes no difference between these three sentence types, nor in terms of how they are affected by the relevance manipulation. Accordingly, the results do not support the idea that conventional implicatures (“therefore” conveying PO; “but” conveying NE) affect truth-table evaluations. Instead a common truth-table semantics of the ABT emerged in line with H_{O_ABT} , as derived from the Frege-Grice tradition, and effects of relevance on their shared truth table were small.

This finding was also supported by the high confidence ratings across conditions. Even in conditions where conflicts between truth evaluations based on the logical conjunction and the reason relation reading of the ABT sentences were induced, the participants indicated that they had a confidence of 76% or higher in their truth value judgments.

In addition, the results indicate that none of the truth tables for indicative conditionals outlined in Table 1 are able to capture the patterns in the data for indicative conditionals, even when applying the most lenient test (i.e. that the deterministic truth tables only have to predict the relative, rather than absolute, majority responses). As shown in Table 6, the de Finetti table and the material implication account are much better suited to capture the modal responses in the true antecedent cases than truth-conditional inferentialism.³⁶ In comparison, truth-

³⁶ As noted in Footnote 33, Karolina Krzyżanowska has in discussion expressed doubts about whether the truth conditional inferentialism is really committed to predicting ‘True’ for the T⊥ cell in the positive relevance condition—not least due to the context-sensitive interpretation of the indicative conditional voiced in Krzyżanowska *et al.* (2014). However,

conditional inferentialism does a much better job in the false antecedent cases, where the de Finetti table and the material implication account encountered difficulties. Indeed, although the defective truth table effect is often cited in the literature as strong evidence in favor of the de Finetti truth table (Over & Evans, 2003), neither the results reported in Table 6 with realistic stimulus material nor the results from the 22 studies with abstract stimulus materials reported in Table 7 support the assumption that ‘Neither true nor false’ is the modal response in the false antecedent cases.

Interestingly, Figure 1 does indicate the presence of a relevance effect on the truth evaluations of the indicative conditional in the $\perp\perp$ cell, which is compatible with the truth table for truth-conditional inferentialism (see Table 1).

However, the gross failure of the predictions in Table 1 to account for the data of Experiment 1 suggests the possibility of some kind of within-subject variation with respect to the truth tables the participants rely on. Indeed, it is possible that a mixture of truth tables (*inter alia*, the bi-conditional table and the de Finetti table) would have to be invoked to account for our results. However, the present data suggest that it would have to be a small proportion of the individuals that follow the de Finetti table. The proportion of ‘NN’ responses in the false-antecedent cells for ‘If Then’ sentences hardly exceeds 33%, suggesting that no more than a third of the individuals in these cells of the experimental design were able to respond in accordance with the de Finetti table. Moreover, such a mixture account is not able to account for the relevance effect found for the $\perp\perp$ cell, captured by truth-conditional inferentialism. Follow-up studies better suited for testing individual variation that have the participants fill out all four truth table cells for each sentence (perhaps multiple times in order to estimate response-error probabilities) would be needed to investigate this possibility.

even when taking this point into account, it is not clear to how the theory could be adjusted in order to successfully accommodate an absolute majority of ‘True’ responses in the TT cell for the NE and IR conditions.

Experiment 2

Given that no difference was found among the ABT sentences in Experiment 1, we wanted to see in Experiment 2 whether a dissociation between these sentences occurs when they are evaluated in the context of a probability-judgment task and an acceptability-ranking task. According to the reason-relation reading, ‘ ϕ but ψ ’ expresses that ϕ lowers the probability of ψ ($\Delta p < 0$), and ‘ ϕ therefore ψ ’ expresses that ϕ raises the probability of ψ ($\Delta p > 0$). In contrast, ‘ ϕ and ψ ’ can suggest that ϕ raises the probability of ψ , but according to its reading as the logical conjunction ‘ ϕ & ψ ’, ϕ need not affect the probability of ψ at all.

Hence, we expected that when presented with the TT cell, the acceptability ratings would accord with the following pattern, where ‘ ϕ and ψ ’ acts as a baseline:

(NE)	ϕ but $\psi > \phi$ and $\psi > \phi$ therefore ψ	$(b > a > t)$
(PO)	ϕ therefore $\psi > \phi$ and $\psi > \phi$ but ψ	$(t > a > b)$
(IR)	ϕ and ψ / ϕ but $\psi > \phi$ therefore ψ	$(b/a > t)$

Moreover, on the assumption that ‘ ϕ but ψ ’ expresses that ϕ is assumed to be a sufficient reason *against* ψ , and that ‘ ϕ therefore ψ ’ expresses that ϕ is assumed to be a sufficient reason *for* ψ , we would expect $P(\psi | \phi) = \text{high/low}$ to act as a moderator variable. That is to say, we expect the pattern $b > a > t$ to be more frequent in NE when $P(\psi | \phi) = \text{low}$ compared to $P(\psi | \phi) = \text{high}$, and $t > a > b$ to be more frequent in PO when $P(\psi | \phi) = \text{high}$ as compared to $P(\psi | \phi) = \text{low}$.

As a manipulation check, we tested whether the effect of the relevance manipulation on $P(\psi | \phi)$ as a predictor of $P(\text{if } \phi, \text{ then } \psi)$ from Skovgaard-Olsen *et al.* (2016a) replicates despite the procedural change that our conditionals had the form of ‘if ϕ is *now happening*, then ψ will occur’ as opposed to ‘if ϕ is *occurs*, then ψ will occur’ (see Footnote 34). In addition, we tested whether a similar moderation of $P(\phi \& \psi)$ as a predictor of $P(\phi \text{ but } \psi)$, $P(\phi \text{ and } \psi)$, and $P(\phi \text{ therefore } \psi)$ could be found for our ABT sentences with the expectation that the marginal means would be higher in the NE condition compared to the PO condition for $P(\phi \text{ but } \psi)$ and that the marginal

means of $P(\phi \text{ therefore } \psi)$ would be higher in the PO condition compared to the NE and IR conditions.

Method

Participants

Like Experiment 1, the experiment was conducted over the Internet. A total of 805 people completed the experiment. The participants were sampled through the Internet platform Mechanical Turk from USA, UK, and Australia and were paid a small amount of money for their participation. The same exclusion criteria were applied as in Experiment 1. The final sample thus consisted of 593 participants. Mean age was 39 years, ranging from 18 to 80 years; 32% of the participants were male; 73% indicated that the highest level of education that they had completed was an undergraduate degree or higher. The demographic measures of the participants differed only minimally before and after the exclusion.

Design

Experiment 2 had the same experimental design as Experiment 1 for the probability task. In contrast, the acceptability task only differed by presenting the participants with three levels of the sentence factor ('...and...', '...but...', '...therefore...').

Materials and Procedure

The procedures were the same as in Experiment 1 unless otherwise stated. For each of the twelve priors×relevance within-participants conditions, the participants were presented with four pages. The first page featured only the scenario text, which participants read had been supplied as input to a computer program in the development phase (following the instructions from Experiment 1). The second page asked the participants both to evaluate the probability of the antecedent (e.g. 'Scott is now turning on the warm water') and of the consequent (e.g. 'Scott will be warm

soon') conditional on the antecedent on a slider with a scale from 0 to 100%. The instruction for evaluating the conditional probability was as follows:

Suppose Scott is now turning on the warm water.

Under this assumption, how probable is it that the following sentence is true on a scale from 0 to 100%:

Scott will be warm soon.

The third page asked the participants to evaluate the probability of a randomly chosen member of our four target sentences (' ϕ and ψ ', ' ϕ but ψ ', ' ϕ therefore ψ ', 'if ϕ , then ψ '). Continuing with our example of a PO HH condition from the Scott scenario, one participant might be asked to evaluate the probability of the 'but' sentence as follows:

Could you please rate the probability that the following sentence is true on a scale from 0 to 100 %:

Scott is now turning on the warm water, BUT Scott will be warm soon.

On page four the participants were presented with the acceptability task. Inspired by the task of evaluating the categorical acceptability of conditionals in Douven & Verbrugge (2012), we introduced the novel task of rank-ordering the acceptability of the ABT sentences given the TT cell with the computer program calibration instructions from Experiment 1. That is to say, the participants were presented with the scenario text, which they had been instructed to regard as input to the computer program. They were then presented with the TT cell as a continuation of the scenario, which took place after the computer program had produced its output sentences. The task was to evaluate which of the three ABT sentences was most acceptable in light of the continuation of the scenario. Continuing with the example from above:

Continuation: *Scott turned on the warm water. Scott did get warm.*

Please order the OUTPUT in terms of how acceptable they are in light of the continuations of the scenarios. Click on the most acceptable output for rank 1, the

second most acceptable for rank 2, and the third most acceptable for rank 3. Note that the responses can be deselected.

Output: Scott is now turning on the warm water, BUT Scott will be warm soon.

Output: Scott is now turning on the warm water AND Scott will be warm soon.

Output: Scott is now turning on the warm water THEREFORE Scott will be warm soon.

As in Experiment 1, the computer output sentences were distinguished by a different font to help the participants organize the information. Finally, the participants were asked to indicate whether they agree with the statement that at least one of the output sentences was acceptable given the continuation.

The instructions presented after the practice trial followed Skovgaard-Olsen *et al.* (2016a) in giving the following explication of how ‘acceptable’ was meant to be understood:

Please note that 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 and their continuations.

Results

Acceptability

We excluded rank orders for which participants found none of the output sentences to be categorically acceptable (24%).³⁷ Table 8 reports the rank order of the sentences

³⁷ Note that the ranking distributions did not differ qualitatively, when including rankings for which participants found none of the sentences to be acceptable. However, the increase of $b > a > t$ in the NE condition when $P(\psi | \phi) < .50$ was no longer significant ($p = .06$).

from the acceptability task. Overall, the results matched our predictions, with the rank order $t > a > b$ occurring most often in PO, $b > a > t$ in NE, and $a > b > t$ in IR. Another prediction corroborated by the data was that in PO, $t > a > b$ occurred less often when the participants judged $P(\psi | \phi)$ to be low. Indeed, the proportion of $t > a > b$ ranks was larger (66%) when $P(\psi | \phi) \geq .50$ than when $P(\psi | \phi) < .50$ (53%), a difference that was found to be statistically significant ($\Delta G^2 = 25.51, p < .001, \Delta FIA = 10.29$). Conversely, we predicted that in NE, $b > a > t$ would occur more often when $P(\psi | \phi)$ was judged to be low rather than high, a difference that was also found in the data (74% versus 66%; $\Delta G^2 = 9.85, p = .001, \Delta FIA = 2.44$).

Table 8. Percentage of rank orders in the acceptability task

	ϕ/ψ Priors	$a > b > t$	$a > t > b$	$b > a > t$	$t > a > b$	$b > t > a$	$t > b > a$
PO	HH	3%	31%	1%	62%	1%	2%
	HL	5%	28%	5%	54%	4%	4%
	LH	1%	23%	1%	71%	2%	3%
	LL	4%	19%	5%	64%	3%	4%
NE	HH	8%	4%	73%	5%	9%	1%
	HL	6%	4%	74%	5%	8%	3%
	LH	10%	10%	66%	7%	6%	1%
	LL	7%	3%	74%	4%	10%	3%
IR	HH	49%	35%	6%	8%	1%	1%
	HL	52%	19%	23%	3%	1%	1%
	LH	57%	19%	15%	4%	2%	2%
	LL	53%	22%	14%	7%	2%	2%

Note. ϕ = the antecedent (or first conjunct); ψ = the consequent (or second conjunct). The operator $>$ denotes “more acceptable than”. ‘a’ = ‘and’, ‘b’ = ‘but’, and ‘t’ = ‘therefore’.

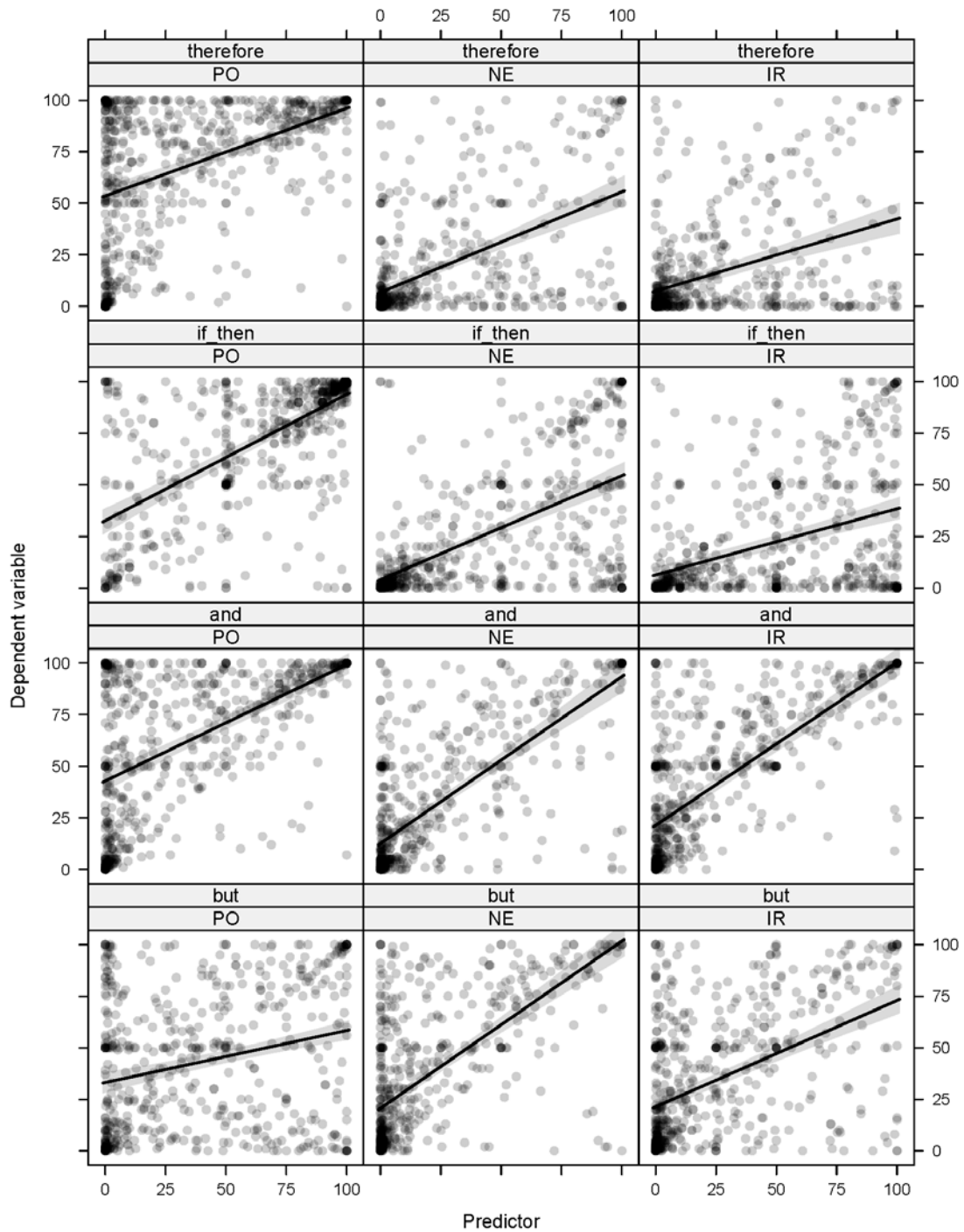
The interactions between sentence levels and relevance levels in determining the preferred rank orders shown in bold in Table 8 are of such magnitude that a statistical analysis is superfluous; their prominence is too severe to leave any doubt.

Probability Judgments

According to Figure 2 (see panels in the second row from the top), the results from Skovgaard-Olsen *et al.* (2016a) for $P(\psi | \phi)$ as a predictor of $P(\text{if } \phi, \text{ then } \psi)$ across relevance levels appear to be replicated. In comparison to Figure 1 in Skovgaard-Olsen *et al.* (2016a), the intercept is slightly larger and less points fall on the diagonal in the

present Figure 2. But the overall tendency is the same: There is a stronger relationship between $P(\psi | \phi)$ and $P(\text{if } \phi, \text{ then } \psi)$ for the PO condition than for the IR condition in particular and the marginal mean (i.e. the overall level of the set of points on the y axis) of $P(\text{if } \phi, \text{ then } \psi)$ is higher in the PO condition than in the NE and IR conditions across the scale of $P(\psi | \phi)$. For the IR and NE conditions there is a substantial portion of data points assigning probabilities almost equal to zero to 'if ϕ , then ψ ' across variations in $P(\psi | \phi)$. This same tendency is visible in the probability evaluations in Figure 2 (top row) for ' ϕ therefore ψ ' with $P(\phi \& \psi)$ as a predictor; only this time the differences in intercept and marginal means between PO and the IR and NE conditions appear to be even more pronounced. In Figure 2 (third row from top), it moreover seems that $P(\phi \& \psi)$ is a good predictor of $P(\phi \text{ and } \psi)$ across relevance levels, with a higher intercept in the PO condition. Finally, $P(\phi \& \psi)$ appears to best predict $P(\phi \text{ but } \psi)$ for the NE condition. In contrast, in the PO condition there appears to be only a very weak relationship between $P(\phi \& \psi)$ and $P(\phi \text{ but } \psi)$.

Figure 2. Probability evaluations of the sentence types across relevance levels



Note. $P(\phi \ \& \ \psi)$ is plotted as a predictor of $P(\phi \ \text{therefore/and/but} \ \psi)$ across relevance levels (rows 1, 3, and 4). $P(\psi \ | \ \phi)$ is plotted as a predictor of $P(\text{if} \ \phi, \ \text{then} \ \psi)$ across relevance levels (row 2). Raw data values (plotted with 80% transparency) and LMM linear effect of the predictors on $P(\phi \ \text{therefore} \ \psi)$ (row 1), $P(\text{if} \ \phi, \ \text{then} \ \psi)$ (row 2), $P(\phi \ \text{and} \ \psi)$ (row 3), and $P(\phi \ \text{but} \ \psi)$ (row 4) across relevance manipulations (PO = left column, NE = center column, IR = right column). The confidence band show the 95% confidence region of the effect of the two independent variables, $P(\phi \ \& \ \psi)$ and $P(\psi \ | \ \phi)$.

This pattern was confirmed in an LMM analysis (see Appendix for details). Main effects of $P(\psi | \phi)$ as predictor of $P(\text{if } \phi, \text{ then } \psi)$ and $P(\phi \& \psi)$ as predictor of the ABT sentences were found, $F(1, 18.4) = 1095.21, p < .0001$. Also, a main effect of the relevance conditions was found, $F(2, 12.3) = 99.24, p < .0001$, as well as an interaction between the $P(\psi | \phi)/P(\phi \& \psi)$ predictors and the relevance conditions, $F(2, 13) = 10.20, p < .01$. Main effects for the sentence type, and interactions between sentence type and the other predictors, were also found.

Follow-up analyses for $P(\psi | \phi)$ as a predictor of $P(\text{if } \phi, \text{ then } \psi)$ showed that the slope found in the PO condition ($b = 0.61, 95\text{-CI} = [0.54, 0.69]$), was significantly larger than the slope in the IR condition ($b = 0.32, 95\text{-CI} = [0.26, 0.39]$), $z = 5.99, p_T < .001$, but was not significantly larger than the slope in the NE condition ($b = 0.51, 95\text{-CI} = [0.43, 0.58]$), $z = 2.02, p_T = .68$. The subscript 'T' indicates p-value correction by the Tukey method for comparing a family of 12 estimates for all follow-up analyses.

Follow-up analysis for $P(\phi \& \psi)$ as a predictor of $P(\phi \text{ therefore } \psi)$ showed that the slope in the PO condition ($b = 0.43, 95\text{-CI} = [0.36, 0.49]$) was neither significantly larger than the slope in the IR condition ($b = 0.35, 95\text{-CI} = [0.27, 0.43]$), $z = 1.40, p_T = .97$, nor significantly smaller than the slope in the NE condition ($b = 0.49, 95\text{-CI} = [0.40, 0.58]$), $z = -1.11, p_T = .99$.

Furthermore, follow-up analyses for $P(\phi \& \psi)$ as a predictor of $P(\phi \text{ and } \psi)$ showed that the slope in the PO condition ($b = 0.56, 95\text{-CI} = [0.49, 0.64]$) was both significantly smaller than the slope in the IR condition ($b = 0.79, 95\text{-CI} = [0.71, 0.87]$), $z = -4.30, p_T < .01$, and significantly smaller than the slope in the NE condition ($b = 0.81, 95\text{-CI} = [0.72, 0.90]$), $z = -4.15, p_T < .01$.

Finally, a follow-up analysis for $P(\phi \& \psi)$ as a predictor of $P(\phi \text{ but } \psi)$ showed that the slope in the PO condition ($b = 0.25, 95\text{-CI} = [0.18, 0.32]$) was both significantly smaller than the slope in the IR condition ($b = 0.52, 95\text{-CI} = [0.44, 0.59]$), $z = -5.07, p_T < .0001$, and significantly smaller than the slope in the NE condition ($b = 0.81, 95\text{-CI} = [0.72, 0.91]$), $z = -9.24, p_T < .0001$.

Differences in estimated marginal means for the four target sentences were tested for their statistical significance at the three scale points of the independent variables ($P(\psi | \phi)$ for $P(\text{if } \phi, \text{ then } \psi)$ and $P(\phi \& \psi)$ for $P(\phi \text{ and/but/therefore } \psi)$) in Table 9.

Table 9. Differences in Estimated Marginal Means

Sentence	Relevance	IV = 0%	IV = 50%	IV = 100%
P(if ϕ , then ψ)	PO	32.6%	63.3%	93.9%
	NE	4.0% ***	29.2% ***	54.4% ***
	IR	6.3% ***	22.4% ***	38.4% ***
P(ϕ therefore ψ)	PO	53.6%	79.9%	96.1%
	NE	6.8% ***	31.2% ***	55.6% ***
	IR	7.4% ***	24.9% ***	42.4% ***
P(ϕ and ψ)	PO	42.9%	71.1%	99.3%
	NE	12.6% ***	53.0% ***	93.5%
	IR	21.3% ***	60.8% *	100%
P(ϕ but ψ)	PO	33.2%	45.8%	58.4%
	NE	20.6% *	61.2% ***	100% ***
	IR	21.4%	47.2%	73.0%

Note. Differences for the Estimated Marginal Means for the four sentence levels were tested for their statistical significance across the following scale points of the independent variables (IV = $P(\psi | \phi)$ for $P(\text{if } \phi, \text{ then } \psi)$ and IV = $P(\phi \& \psi)$ for $P(\phi \text{ and/but/therefore } \psi)$): 0%, 50%, and 100%. The pairwise contrasts indicate whether the NE or IR conditions differed significantly from the PO condition using z-ratios and adjusted p -values through Tukey's method for comparing a family of 36 estimates. Signif. codes: '***' .001, '**' .01, '*' .05, '.' .1.

As Table 9 shows, the estimated marginal means of $P(\text{if } \phi, \text{ then } \psi)$ and $P(\phi \text{ therefore } \psi)$ were consistently statistically higher in the PO condition than in the NE and IR condition across the probability scales of their respective independent variables. In contrast, the estimated marginal means of $P(\phi \text{ and } \psi)$ was only significantly higher in the PO condition than in the NE and IR conditions at $P(\phi \& \psi) = 0\%$ and $P(\phi \& \psi) = 50\%$. And finally, the estimated marginal means of $P(\phi \text{ but } \psi)$ was only significantly higher in the PO condition than in the NE condition at $P(\phi \& \psi) = 0\%$. For $P(\phi \& \psi) = 50\%$ and $P(\phi \& \psi) = 100\%$, the estimated marginal means of $P(\phi \text{ but } \psi)$ were significantly higher in the NE condition than in the PO condition.

Discussion

The results obtained in Experiment 2 show an unmistakable pattern. Participants rank ordered the acceptability of the ABT sentences given the continuation of the scenario in the TT cell. We predicted that ‘ ϕ but ψ ’ differs from the ‘ ϕ and ψ ’ baseline in being more preferable in NE, and that ‘ ϕ therefore ψ ’ differs from the ‘ ϕ and ψ ’ baseline in being more preferable in PO. The predicted rank orders clearly dominate the other possible rank orders in the data. Furthermore, $P(\psi \mid \phi)$ moderates the relationship between the rank order acceptabilities and the relevance levels in the manner one would expect, if ‘ ϕ but ψ ’ expresses that ϕ is a sufficient reason against ψ and ‘ ϕ therefore ψ ’ expresses that ϕ is a sufficient reason for ψ .

This basic finding is corroborated by the results from the probability evaluation task outlined in Table 9. Consistent with the reading of ‘if ϕ , then ψ ’ and ‘ ϕ therefore ψ ’ as indicating that ϕ is a reason *for* ψ , the estimated marginal means for $P(\text{if } \phi, \text{ then } \psi)$ and $P(\phi \text{ therefore } \psi)$ in the PO condition were invariantly higher than the estimated marginal means in the NE and IR conditions across the scale of their respective independent variables. Moreover, consistent with the reading of ‘ ϕ but ψ ’ as indicating that ϕ is a reason *against* ψ , the estimated marginal means for $P(\phi \text{ but } \psi)$ was higher in the NE condition than in the PO condition at $P(\phi \ \& \ \psi) = 50\%$ and $P(\phi \ \& \ \psi) = 100\%$.

General Discussion

Given the large sample sizes and high power for detecting even small differences, the finding from Experiment 1 that the (relative) response frequencies of the ABT sentences can be set equal across sentence levels, and that relevance does not interact with the sentence factor, is a strong and in fact surprising result. It suggests that the ABT sentences have exactly the same truth conditions. Taken together, Experiment 1 and 2 show a clear dissociation between evaluations of truth,

probability, and acceptability. Relevance interacts with sentence (and, but, therefore) in the rank-order acceptability task and the probability evaluation task as expected, but does not interact with sentence in the truth-evaluation task. In fact, truth evaluations of the ABT sentences could be set equal across sentences in that task with only minor losses in goodness of fit. The fact that Table 8 and 9 display such strong interaction effects, when the participants are asked to rank order the acceptability and provide probability evaluations, makes the absence of an interaction of relevance and sentence in the truth-table task even more surprising. It suggests that there is a deeply entrenched modularization and little cross-talk between the processes and/or representations tapped by the tasks in Experiments 1 and 2. These findings suggest that the Frege-Grice tradition was right in its assumption that the difference in reason-relation readings of these sentences does not affect their truth evaluations.

Instead, Grice (1989) conjectured that the differences of the ABT sentences would be part of their conventional implicatures. In fact, Experiment 2 found the signature effects of the reason-relation readings in the orderings of the ABT sentences according to acceptability, as also corroborated in the probability judgments. Finally, the fact that we were able to find strong effects of the relevance manipulation in the expected directions for the ABT sentences in Experiment 2 suggests that the absence of such a difference in Experiment 1 is not an artifact of the stimulus material (see Footnote 34) nor of the instructions of the computer-calibration task.

Turning to indicative conditionals, the results from Experiment 1 on the truth evaluations of the indicative conditional present an explanatory challenge, as none of the investigated truth tables were able to account for the patterns found. We found that there is a marked relevance effect on the truth evaluations of the indicative conditional in the $\perp\perp$ cell. This finding challenges the H_{0_IF} assumption underlying the material implication account and the de Finetti truth table, which holds that truth evaluations of the indicative conditional are not affected by relevance manipulations. Moreover, it was shown that in spite of its recent popularity in psychology of reasoning (Baratgin, Politzer, and Over, 2013; Pfeifer, 2013; Elqayam and Over, 2013),

the de Finetti truth table lacks support for its distinguishing feature: The prediction of ‘Neither true nor false’ evaluations in the false antecedent cases did neither find support in the results from Experiment 1 nor in the results of the 22 studies with abstract stimulus material from the Schroyens (2010) meta-analysis presented in Table 7. In contrast, truth-conditional inferentialism correctly rejects H_{0_IF} , but there is little support for its alternative predictions. In particular, truth-conditional inferentialism faced problems accounting for our results on the true antecedent cells, yet it was compatible with the relevance effect found in the $\perp\perp$ cell.

It is remarkable that none of the investigated theories was able to fit the data even under the most lenient stochastic interpretation of their deterministic truth tables, whereby they only had to be able to predict the relative-majority responses of the participants (see the Appendix for further discussion). A stricter interpretation would have imposed the requirement that the absolute majority responses were predicted by the theories, or by admitting relatively small “error” rates (e.g., 10%).

As noted, it is possible that the pattern of truth evaluations of the indicative conditional found across relevance manipulations is best accounted for by a mixture of truth tables. It is up to future experiments more suitably designed to test for individual variation to explore this possibility. However, any account based on a mixture of truth tables is still constrained by the finding in Experiment 2 of a modest relevance effect on the truth evaluations of the indicative conditional.

It should be noted, however, that a different perspective on our results is also possible. A non-truth functional conditional is a conditional whose truth value cannot be determined by the truth values of its parts. As Rescher (2007, p. 43) points out, for any non-truth functional conditional that is logically stronger than ‘ \supset ’, it holds that “we can say nothing about the truth status of $p \rightarrow q$ without a deeper look at the specifics of the matter”. It is therefore possible to interpret our results as indicating that indicative conditionals are non-truth functional with further parameters determining their truth values in the false antecedent cases. If the participants set

these parameters differently, then mixed results, like those found in Figure 1, may be the outcome.

Potential Limitations

When evaluating our results it is important to keep the following points in mind. First, when asking the participants to calibrate the output of a computer program, the aim was to create a situation in which the participants would not naturally begin interpreting communicative intentions behind producing assertions with components that are irrelevant for one another and reconstruct speaker meaning. It turns out that such an approach has a track-record in the literature. For instance, Schwarz, Strack, Hilton, and Naderer (1991) employ a conceptually similar manipulation to set aside Gricean conversational norms (for discussion see Lee, 2006). Moreover, Doran, Ward, Larson, McNabb, and Baker (2012) achieved a similar effect by asking the participants to provide truth value judgments based on adopting the perspective of someone who is only able to understand the literal content of what has been said. Finally, in Wright and Wells (1988) an attempt was undertaken to control for demand characteristics relating to the Gricean ideal of cooperative discourse in the attitude-attribution paradigm by instructing the participants that the set of questionnaire items they were presented with had been randomly selected from a larger pool. Due to this random assignment, the participants were instructed that they could encounter cases in which they had to make judgments on the basis of information that was either irrelevant or insufficient for the task at hand.

In Doran *et al.* (2012) confidence ratings are moreover used as a measure of task complexity, because confidence ratings have been found to be inversely correlated with perceived task complexity. In the truth-table task in Experiment 1, all mean confidence ratings were in the interval [76%, 81%], which indicates that the participants were highly confident of judgments across all conditions. Indeed, the participants continued to remain highly confident of their responses even when

judging truth values based on logical connectives and reason relation constraints would have led to conflicting responses.

For the rank-ordering of acceptability and probability evaluations in Experiment 2 we retained the computer program instruction to discourage participants from looking for some hidden intention for why an irrelevance item had been asserted and to encourage a focus on rating the sentences for their acceptability and probability under different conditions.

Note, finally, that the differences between the findings in Experiment 1 and Experiment 2 cannot merely be attributed to the presence of a forced choice format or dependencies in the rank ordering task, since the probability evaluation task in Experiment 2 produced similar results without having a forced-choice format. Moreover, even if one restricts the attention to the most preferred sentence in each of the relevance conditions, because one is worried that the choices are not independent, a very clear pattern emerges: in the PO condition, the therefore-sentences are by far the most acceptable sentences, in the NE condition, the but-sentences are by far the most acceptable sentences, and in the IR condition the and-sentences are by far the most acceptable sentences. This pattern is in agreement with the reason relation reading of these sentences and it is also one that is mirrored in the probability evaluations, on which our predictions were based.

Conclusion

In the Frege-Grice tradition of applying logic as a model of natural language connectives, it is assumed that the difference in reason-relation readings of the sentences ‘ ϕ and ψ ’, ‘ ϕ but ψ ’, and ‘ ϕ therefore ψ ’ does not affect their truth conditions. Support for this assumption was found in Experiment 1. In contrast, Experiment 2 indicated a dissociation between the effects of relevance on ‘ ϕ and ψ ’, ‘ ϕ but ψ ’, and ‘ ϕ therefore ψ ’ in truth evaluations and in evaluations of their acceptability and probabilities. For, when the participants are asked to rank order the

acceptability of ‘ ϕ and ψ ’, ‘ ϕ but ψ ’, and ‘ ϕ therefore ψ ’ on the basis of the TT cell, the different rank orders predicted by a reason-relation reading of each sentence are strongly preferred. Moreover, their probability evaluations accord with the reason-relation reading. Turning to the indicative conditionals, a relevance effect on truth evaluations was found, and neither the truth tables supplied by material implication account, the popular de Finetti truth table, nor truth-conditional inferentialism were able to account for these results of Experiment 1, even under the most lenient stochastic interpretation of their predictions. Accordingly, the results for the truth evaluations of indicative conditionals across relevance levels from Experiment 1 present an explanatory challenge for further theorizing and empirical work to solve.

References

- 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>.
- Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, 6, 57–86.
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*. URL = <http://arxiv.org/abs/1406.5823>
- Blome-Tillmann, M. (2013). Conventional Implicatures (and How to Spot Them). *Philosophy Compass*, 8/2, 170-85.
- Birnbaum, M. H. (2013). True-and-error models violate independence and yet they are testable. *Judgment and Decision Making*, 8, 717-737.
- Blackmore, D. (2004), *Relevance and Linguistic Meaning: The Semantics and Pragmatics of Discourse Markers*. Cambridge: Cambridge University Press.
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. New York: Springer.

- Doran, R., Ward, G., Larson, M., McNabb, Y., and Baker, R. E. (2012). A Novel Experimental Paradigm for Distinguishing between What is Said and What is Implicated. *Language*, 88(1), 124-54. doi: [10.1353/lan.2012.0008](https://doi.org/10.1353/lan.2012.0008)
- Douven, I. (2015a). *The Epistemology of Indicative Conditionals. Formal and Empirical Approaches*. Cambridge: Cambridge University Press.
- Douven, I. (2015b). How to account for the oddness of missing-link conditionals. *Synthese*, 1-14. doi:10.1007/s11229-015-0756-7
- Douven, I. and Verbrugge, S. (2012). Indicatives, concessives, and evidential support. *Thinking and Reasoning* 18 (4), 480-99. doi: 10.1080/13546783.2012.716009
- Elqayam, S. and Over, D. E. (2013). New paradigm psychology of reasoning: An introduction to the special issue. *Thinking & Reasoning*, 19:3-4, 249-265.
- Erdfelder, E., Auer, T., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models. *Zeitschrift für Psychologie / Journal of Psychology*, 217, 108–124.
- Evans, J. St. B. T. and Over, D. (2004). *If*. Oxford: Oxford University Press.
- Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100, 25-50.
- Garmut, L. T. F. (1991). *Logic, Language, and Meaning, Vol 1*. Chicago: The University of Chicago Press.
- Grice, P. (1989). *Studies in the Way of Words*. Cambridge, MA.: Harvard University Press.
- Grünwald, P. (2007). *The minimum description length principle*. Cambridge, Mass: MIT Press.
- Hilbig, B. E., & Moshagen, M. (2014). Generalized outcome-based strategy classification: Comparing deterministic and probabilistic choice models. *Psychonomic Bulletin & Review*, 21, 1431-1443. doi: [10.3758/s13423-014-0643-0](https://doi.org/10.3758/s13423-014-0643-0)

- Iten, C. B. (2000). *'Non-Truth-Conditional' Meaning, Relevance and Concessives*.
 Doctoral thesis, University of London. URL = <
<http://discovery.ucl.ac.uk/1348747/1/324676.pdf>>
- Johnson-Laird, P. N. and 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>
- Karabatsos, G. (2005). The exchangeable multinomial model as an approach for testing axioms of choice and measurement. *Journal of Mathematical Psychology*, 49, 51-69.
- Kellen, D., Klauer, K. C., & Bröder, A. (2013). Recognition Memory Models and Binary Response ROCs: A Comparison by Minimum Description Length. *Psychonomic Bulletin & Review*, 20, 693-719.
- Klauer, K. C. & Kellen, D. (2011). The flexibility of models of recognition memory: An analysis by the minimum-description length principle. *Journal of Mathematical Psychology*, 55, 430-50.
- Klauer, K. C. & Kellen, D. (2015). The Flexibility of Models of Recognition Memory: The Case of Confidence Ratings. *Journal of Mathematical Psychology*, 67, 8-25.
- Klauer, K. C., Singmann, H., & Kellen, D. (2015). Parametric order constraints in Multinomial Processing Tree Models: An extension of Knapp & Batchelder (2004). *Journal of Mathematical Psychology*, 64, 1-5.
- Krzyżanowska, K., Wenmackers, S. and Douven, I. (2014). Rethinking Gibbard's Riverboat Argument. *Studia Logica*, 102 (4), 771-92.
 doi: 10.1007/s11225-013-9507-2.
- Krzyżanowska, K. (2015). *Between "If" and "Then": Towards an empirically informed philosophy of conditionals*. PhD dissertation, Groningen University.
 URL = <http://karolinakrzyzanowska.com/pdfs/krzyzanowska-phd-final.pdf>
- Lee, C. J. (2006). Gricean Charity: The Gricean Turn in Psychology. *Philosophy of the Social Sciences*, 36, 193-218.
- Luce, R. D. (1995). Four tensions concerning mathematical modeling in psychology.

- Annual Review of Psychology*, 46, 1–26.
- Luce, R. D. (1997). Several unresolved conceptual problems of mathematical psychology. *Journal of Mathematical Psychology*, 41, 79–87.
- Manktelow, K. (2012). *Thinking and Reasoning: An Introduction to the Psychology of Reason, Judgment and Decision Making*. Sussex: Psychology Press.
- McCawley, J. (1993). *Everything that Linguists have Always Wanted to Know About Logic*. Second Edition. Chicago, University of Chicago Press.
- Nickerson, R. S. (2015). *Conditionals and 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.
URL = <http://kops.uni-konstanz.de/handle/123456789/29353>.
- Over, D. and Evans, J. St. B. T. (2003). The Probability of Conditionals: The Psychological Evidence. *Mind and Language*, 18 (4), 340-58.
doi: 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), 329-45. doi: 10.1080/13546783.2013.838189
- Pfeifer, N. and Douven, I. (2014). Formal epistemology and the new paradigm psychology of reasoning. *The Review of Philosophy and Psychology*, 5(2), 199-221. doi: 10.1007/s13164-013-0165-0
- Potts, C. (2015). Presuppositions and implicature. In: Shalom Lappin and Chris Fox (eds.), *The Handbook of Contemporary Semantic Theory*, 2nd edn, 168-202. Oxford: Wiley-Blackwell.

- Read, T., & Cressie, N. (1988). *Goodness-of-fit statistics for discrete multivariate data*. New York: Springer.
- Regenwetter, M., Dana, J., & Davis-Stober, C. P. (2011). Transitivity of preferences. *Psychological Review*, *118*, 42–56.
- Reips, U. D. (2002). Standards for Internet-based experimenting. *Experimental Psychology*, *49* (4), 243-256. doi: 10.1027//1618-3169.49.4.243
- Rescher, N. (2007). *Conditionals*. Cambridge, MA: The MIT Press.
- Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, *95*, 318–339.
- Rouder, J. N., Lu J., Morey R. D., Sun D., & Speckman P. L. (2008). A hierarchical process dissociation model. *Journal of Experimental Psychology: General*, *137*, 370-389.
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, *8*. Available from Methods of Psychological Research Online, <http://www.mpronline.de>
- Schroyens, W. (2010). A meta-analytic review of thinking about what is true, possible, and irrelevant in reasoning from or reasoning about conditional propositions. *European Journal of Cognitive Psychology*, *22* (6), 897-921.
doi: 10.1080/09541440902928915
- Schwarz, N., Strack, F., Hilton, D. and Naderer, G. (1991). Base rates, representativeness, and the logic of conversation: The contextual relevance of “irrelevant” information. *Social Cognition*, *9* (1): 67-84
- Self, S. G., & Liang, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, *82*(398), 605-610. doi :
10.1080/01621459.1987.10478472

- Singmann, H., B. Bolker, J. Westfall, S. Højsgaard, J. Fox, M. Lawrence, et al. (2016). afex: Analysis of Factorial Experiments. R package version 0.13-145, Available via <http://cran.rproject.org/package=afex>.
- Singmann, H., & Kellen, D. (2013). MPTinR: Analysis of Multinomial Processing Tree models with R. *Behavior Research Methods*, 45, 560-575.
- 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., Singmann, H., and Klauer, K. C. (2016a). The relevance effect and conditionals. *Cognition*, 150, 26-36. [doi:10.1016/j.cognition.2015.12.017](https://doi.org/10.1016/j.cognition.2015.12.017)
- Skovgaard-Olsen, N., Singmann, H., and Klauer, K. C. (2016b). Relevance and Reason Relations. *Cognitive Science*. doi: 10.1111/cogs.12462
- 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. doi: 10.1111/cogs.12057
- Wright, E. F. and Wells, G. L. (1988). Is the Attitude-Attribution Paradigm Suitable for Investigating the Dispositional Bias? *Personality and Social Psychology Bulletin*, 14 (1), 183-190.

Appendix: Statistical Analyses

Experiment 1: MPT Analysis

The observed response frequencies in Experiment 1 were analyzed with multinomial processing tree models (MPT; Riefer & Batchelder, 1988). The Models were fitted with R package MPTinR (Singmann & Kellen, 2013). The MPT framework is typically used to characterize the mixtures of processes and cognitive states that underlie individuals' categorical responses (for reviews, see Batchelder & Riefer, 1999; Erdfelder et al., 2009). However, they can also be used to directly test hypotheses at the level of the observed response distributions through goodness of fit and model-selection statistics

(e.g., Birnbaum, 2013; Hilbig & Moshagen, 2014; Karabatsos, 2005; Klauer & Kellen, 2011, 2015; Klauer, Singmann, & Kellen, 2015). We will evaluate the models' absolute performance via the G^2 statistic (Read & Cressie, 1988) and their relative performance with the Fisher Information Approximation (FIA; Grünwald, 2007). Where traditional model-selection statistics such as Akaike and Bayesian information criteria (Burnham & Anderson, 2002) rely on the number of parameters as a proxy for model complexity, FIA penalizes models according to their functional flexibility.

In the present study, the different theories establish distinct predictions for the truth evaluations. For example, according to the material-implication account, individuals should consider the TT, \perp T, and $\perp\perp$ cells of the truth table as true (T), but deem the T \perp cell false (\perp). These predictions are deterministic in the sense that no other response is considered to be possible. The deterministic nature of axiomatic accounts represents a long-standing problem in psychology due to the need to recast these accounts in order to accommodate the stochastic nature of responses with respect to an experiment's empirical sample space (e.g., Luce, 1995, 1997; Regenwetter, Dana, & Davis-Stober, 2011). For example, the predictions of the material-implication account for the TT cell of the truth table would then be relaxed in order to allow false or NN responses with some probability. The exact manner of relaxation is not entirely clear though. For example, one could assume that individuals almost invariably respond true (e.g., 90% of the times), or alternatively that true responses constitute an absolute or a relative majority, among other possibilities. This issue has been thoroughly explored in studies focusing on whether preferences are transitive (when an individual prefers A to B, and B to C, then A is expected to be preferred to C), in which different stochastic implementations have been considered (e.g. weak, moderate, and strong stochastic transitivity, the triangle inequality; see Regenwetter et al., 2011). In the present case we adopted what we view as the most lenient stochastic implementation; to wit, that the predicted response should occur at last as often as each of the other responses (i.e., that it should enjoy at least a relative

majority). For example, in the case of the TT cell, the stochastic implementation of the material implication account then predicts that

$$P(T) \geq P(\perp), P(NN)$$

The reason behind the adoption of this specific stochastic specification is the diagnostic power associated to its failure, as any theory that fails to succeed under these minimal constraints should be seriously questioned. Note that this stochastic specification is completely agnostic regarding the nature of the deviations from the predictions: Individuals might commit “errors” due to a misreading of the sentences, a failure in their evaluation, or a motor-response error, among other possibilities (see Birnbaum, 2013).

Experiment 2: LMM Analysis

For Experiment 2, a mixed linear model was fitted to the data with fully crossed fixed effects for the predictor (IV: $P(\psi \mid \phi)$ and $P(\phi \ \& \ \psi)$), relevance condition (PO, NE, and IR), and sentence type ($P(\text{if } \phi, \text{ then } \psi)$ and $P(\phi \ \text{and}/\text{but}/\text{therefore } \psi)$) and crossed random effects for participants and scenarios. In lme4 syntax (Bates, Maechler, Bolker, & Walker, 2015), the LMM used took the following form:

$$\begin{aligned} \text{DV} \sim \text{IV} * \text{relevance} * \text{sentence_type} & \quad + (\text{IV} * \text{relevance} \mid \text{participant}) \\ & \quad + (\text{IV} * \text{relevance} \mid \text{scenario}) \end{aligned}$$

The R package afex (Singmann *et al.*, 2016) was used to obtain the statistical significance of the fixed effects while controlling for variability due to participants and scenarios. Note that for the analysis, we controlled for the family-wise error rate of follow-up tests using the Tukey method for comparing a family of 12 estimates (indicated by the subscript ‘T’).

5. GENERAL DISCUSSION

Niels Skovgaard-Olsen
University of Konstanz and
Albert-Ludwigs-Universität Freiburg

A revised version of this chapter is planned for publication as:

Skovgaard-Olsen, N. (invited book chapter). Relevance and Conditionals: A Synopsis.

In: S. Elqayam, I. Douven, J. Evans, and N. Cruz (eds.),

Festschrift for David Over, Routledge Press.

Relevance and Conditionals: A Synopsis

Throughout the dissertation we have encountered two implementations of Inferentialism. On the one hand, the Default and Penalty Hypothesis as an alternative to the Equation, defended by the Suppositional Theory of conditionals. On the other, truth-conditional Inferentialism, which provided an alternative truth table to the de Finetti truth table, defended by the Suppositional Theory of conditionals. As we have seen, while the former received empirical support in Chapter 2, the latter was not supported by the results reported in Chapter 4.

One methodological point of all the empirical studies in the dissertation is that if we are going to draw conclusions about equations and inequalities such as $P(\text{if } A, \text{ then } C) = P(C|A)$ and $P(\text{if } A, \text{ then } C) \geq P(A, C)$ based on our experimental findings, then we should ensure that we have investigated all the permutations of high and low priors and probabilistic dependency and independency first. These equations and inequalities are not bounded in their domain to cases of probability raising, and hence the evidence we cite in support of them should not be thus bound either. Concrete examples of this methodological point are given in Chapters 2 and 3, where previous studies drawing negative conclusions about the role of relevance for $P(\text{if } A, \text{ then } C)$ are cited, and levels of probabilistic coherence in the PO condition is used as evidence for the Equation in the uncertain and-to-if inference task.

To illustrate the impact of these results, consider the following statement from a recent, acclaimed introduction to research on conditionals in cognitive science (Nickerson, 2015:199):

Treating conditionals as probabilistic statements is one of the defining features of what has come to be referred to as the *new paradigm* in cognitive psychology or the “new psychology of reasoning” (Chater & Oaksford 2009, Evans, 2012; Manktelow, 2012; Oaksford & Chater, 2013; Over, 2009; Pfeifer, 2013). Gilio and Over (2012) see the conditional-probability hypothesis as fundamental to the new paradigm and express it this way: “*The conditional probability hypothesis* is that

people will tend to judge the probability of the indicative conditional of natural language, $P(\text{If } A \text{ then } C)$, as the conditional probability of C given A , $P(C|A)$ ” (p. 119). [Notation modified for uniformity]

In Chapter 2 it was found that this conditional probability hypothesis—which is stated above to be *fundamental* to the new paradigm in psychology of reasoning—only holds under the condition of positive relevance (where $P(C|A) - P(C|\neg A) > 0$). In the case of negative relevance ($P(C|A) - P(C|\neg A) < 0$), or irrelevance ($P(C|A) - P(C|\neg A) = 0$), the strong relationship between $P(\text{if } A, \text{ then } C)$ and $P(C|A)$ is disrupted, because participants tend to view natural language indicative conditionals as defective under these conditions.³⁸ Moreover, it was shown that these results generalize to evaluations of acceptability, when the participants are specifically instructed to judge the adequacy of the information provided when considering the statement as a contribution to a conversation. In contrast, the Equation was found to fit the probability and acceptability evaluations of concessive conditionals (‘Even if A , then still C ’) remarkably well across all relevance conditions. But the Equation was formulated as a thesis about indicative conditionals and not as a thesis concerning concessive conditionals, as noted in Chapter 2.

The empirical study in Chapter 3 shows that the results from Chapter 2 generalize to the uncertain and-to-if inference and that empirical support could be found for the explications of reason relations and epistemic relevance in terms of ΔP used in these studies. Indeed, the absolute values of the participants’ conformity to $P(\text{if } A, \text{ then } C) \geq P(A,C)$, which is normatively prescribed by the Suppositional Theory, showed a drop from 87% in the positive relevance condition to 54% in the irrelevance condition. And this was a drop that was not reflected in either the participants’ conformity to $P(C|A) \geq P(A,C)$ or $P(\text{Even if } A, \text{ then still } C) \geq P(A,C)$, which both stayed constant at around 78% across relevance conditions. As explained in Chapter 3, this finding presents supporters of the Suppositional Theory with a dilemma: either the

³⁸ Subsequently, some of the findings have been replicated independently by Vidal and Baratgin (in review) using different methods.

participants are less probabilistically coherent than it appeared in Cruz *et al.* (2015) or a substantial part of them do not follow the Equation across relevance levels.³⁹

Finally, the findings from the empirical study in Chapter 4 strongly indicate that there is a dissociation between the influence of relevance on assessments of acceptability and probability compared to truth evaluations, when investigating ‘and’, ‘but’, ‘therefore’, and ‘if then’ sentences. In Chapter 4 these results were interpreted as showing that there was a deeply entrenched modularization between the processes and/or representations tapped by the experimental tasks reported.

Taken together with the acceptability evaluations from Chapter 2, the results in Chapter 4 indicate that indicative conditionals behave like ‘therefore’ sentences in their probability and acceptability evaluations, and opposite to ‘but’ sentences, with respect to the relevance manipulation. Yet in relation to the truth evaluations, ‘and’, ‘therefore’, and ‘but’ sentences did not differ across relevance conditions, and no evidence for a relevance effect on the true antecedent cells of the truth table of the indicative conditional could be found. Moreover, the consistent high confidence ratings of the participants did not indicate that they were in a state of conflict when assigning truth values to sentences that conflicted with their reason relation evaluation. However, for the FF cell a moderate relevance effect was found on the truth evaluation of the indicative conditional.

These results from Chapter 4 present a puzzle (in addition to the puzzle discussed in Chapter 4 that none of the dominant truth tables seemed to fit the truth evaluations of the indicative conditional at the group level). On the one hand, it is possible to interpret the dissociation found between truth evaluations and probability/-acceptability evaluations as indicating a dissociation between semantic and pragmatic processing of content—with relevance almost exclusively affecting the latter. However, on this interpretation it is still odd that strong probabilistic relevance effects could be found in Chapters 2-3 on experimental tasks, which have been used

³⁹ In a subsequent study that is not part of this dissertation, individual variation in these results was investigated (Skovgaard-Olsen, Kellen, Hahn, Klauer, ms).

by supporters of the Suppositional Theory to provide evidence in favor of a semantic theory. On the other hand, it is possible that some other explanation for the dissociation in Chapter 4 can be found. In particular, the following circumstance might be part of the explanation: the truth evaluation and the probability assignment tasks differed on whether the cognitive evaluation of the sentences were based on a particular truth table cell that was to be treated as certain. To be sure, a strong relevance effect was also found on the 'and', 'but', and 'therefore' sentences, when the participants were provided with the TT cell that they were supposed to treat as certain. But this acceptability task featured a comparative judgment of the acceptability of these three sentences in contrast to the truth and probability tasks, where the participants only had to focus on a particular sentence at a time in their evaluations. It is possible that these differences in the tasks may need to be taken into account in the final interpretation of our results.

Since the boundary between semantics and pragmatics will feature centrally in the interpretation of Chapters 2-4, the next section outlines some of the open issues that these results raise.

Semantic and Pragmatic Factors

In deciding whether the dissociation found in Chapter 4 has implications for whether relevance is to be counted as a semantic or a pragmatic factor in the content of indicative conditionals, the following questions merit further investigation:

- (I) What interpretation of 'truth' do the participants have when they provide truth evaluations?

The importance of (I) can be illustrated as follows. Suppose that the participants make truth value assignments based on an understanding of 'truth' as 'what can be proven in principle independent of whether it has actually been proven'. In that case, the truth table for negation would no longer be truth functional inasmuch as there would

be propositions for which both a proposition and its negation would be 'False'. Moreover, for this understanding of truth, ' \vee ' and ' \supset ' would no longer be truth functional, since it can be shown that the truth functionality of ' \vee ' and ' \supset ' depends on the truth functionality of negation (McCawley, 1993: 107ff).

Indeed, many other ways of interpreting truth exist. In the philosophical literature (Künne, 2005), realistic conceptions of truth (e.g. the correspondence theory of truth, “truth is what corresponds to the facts”) are contrasted with epistemic concepts of truth (e.g. the coherence theory “truth is what belongs to a maximally coherent set of beliefs” or “truth is what all investigators would agree on at the limit of an ideal inquiry”), pragmatic theories (“truth is what works”), and with deflationary theories (e.g. “the predicate ‘true’ is merely a convenient device for disquoting sentences, or a device for forming pro-sentences, which allows us to endorse assertions that we would not be able to endorse otherwise (e.g. “the next thing Pete says is true”, “Everything the Pope says is true”)).

At present it is unknown which understanding of the notion ‘truth’ the participants bring to bear on the truth table task. It is unknown whether the different truth tables elicited by the participants reflect different notions of truth or diverging interpretations of the conditional. And it is moreover unknown whether the participants understand the notion of truth in the same way as the semantic theory they are being tested according to.

Interestingly, Oberauer *et al.* (2007) found that the same group of participants that tend to conform to the de Finetti table in a ternary truth table task tended to conform to the material implication in a binary truth table task—although the two theories stand as diametrically opposite in the literature. It is thus possible that the participants interpret the truth values differently in the two experimental paradigms.

Indeed the truth table of the material implication sounds most plausible, if one interprets the truth value as indicating consistency.⁴⁰ Accordingly, the material implication treats the conditional as true in the false antecedent cells, because the

⁴⁰ I thank Christoph Klauer (p. c.) for this observation.

falsity of the antecedent is *consistent* with the truth of the conditional. This might account for the fact that the material implication is useful in mathematical and logical contexts, where the goal is to keep inconsistency at bay. In line with this idea of the material implication as especially suited for mathematics and deductive logic, Rescher (2007: 43) points out that what makes the material implication appropriate for these contexts is the following link between implications and deducibility that it establishes for demonstratively true instances: $p \vdash q$ iff $\vdash p \supset q$.⁴¹

In contrast, the de Finetti truth table seems to be most plausible if ‘true’ is interpreted as ‘verification’ and ‘falsity’ is interpreted as ‘disconfirmation’. In this context, it is interesting to observe that in some experiments cited in favor of the de Finetti truth table (like Evans *et al.*, 2007), the instructions explicitly ask for whether a truth table cell “conforms” to a conditional rule, “contradicts” it or “is irrelevant” to it”, rather than for the truth or falsity of the conditional *simpliciter*.

- (II) What is the relationship between (a) the semantic values invoked by a given semantic theory and (b) what the participants are evaluating in a given experimental paradigm?

There is some precedence in the linguistic literature for not taking intuitive judgments of truth and falsity at face value. For example, von Stechow (2004) and Abrusán and Szendrői (2013) have argued that intuitive judgments on presupposition failures⁴² as true or false are influenced by pragmatic factors such as possibility of verification and need not represent the sentences’ actual semantic values. Instead, it is argued that it is more decisive whether the semantic values assigned would allow us to construct a systematic theory of the projection behavior of the linguistic expressions in question. Accordingly, in Winter (2016: 20) it is made an empirical adequacy condition of theories in formal semantics that they agree—not with the intuitive truth value

⁴¹ However, as Rescher (2007: 44) also points out, a *demonstrated* true material implication in mathematics and deductive logic is in fact a strict implication, $\Box(p \supset q)$.

⁴² E.g. ‘the Danish Pope is in his sixties’ carries the false presupposition that there is a Danish Pope.

judgments—but with the intuitive entailment judgments. If so, then the strong tendency to focus on the truth table task in the psychology of reasoning as decisive evidence for or against semantic theories may turn out to be misguided.

Moreover, semantics for the indicative conditional and other connate epistemic expressions exist, which pose constraints on probability distributions as semantic values (e.g. Yalcin, 2012; Moss, 2015). Yet constraints on probability distributions do not themselves impose truth conditions that can be interpreted as representing ways that the world can be.

- (III) To avoid a free-license in invoking pragmatics as an explanation of divergences from the semantic theory (such as the divergences from the Suppositional Theory reported in Chapters 2-3), mechanisms that give rise to the pragmatic phenomena need to be posited, which give rise to predictions that can be tested independently.

In commenting on the findings from Chapter 2, Over and Cruz (forthcoming) suggest that the effect might be pragmatic and not semantic, because there is some evidence that relevance also affects conjunctions and disjunctions. The implicit assumption is that if relevance is supposed to be part of the semantic content of indicative conditionals, then it should serve to distinguish the content of indicative conditionals from the semantic content of other connectives.

In Chapter 4, we did not investigate disjunctions. But the results on the probability assignments to conjunctions indicate that while the probability assignments are somewhat higher for the PO condition, there is no evidence for an analogous defect to the one reported in Chapter 2, which would make the participants assign low probabilities to ‘A & C’ in the IR and NE condition.

Moreover, based on Table 1 below, the conjecture could be made that disjunctions are most probable for negative relevance items. This is especially pronounced for the ‘either... or...’ formulation, which can be read as exclusive disjunction—but even for a reading of ‘... or ...’ based on inclusive disjunctions, the NE

formulations that present the two disjuncts as alternatives seem to be more probable than the PO formulations. At any rate, disjunctions don't seem to exhibit the NE/IR defect that was documented for indicative conditionals in Chapter 2. If so, then disjunctions have a distinct relevance profile from indicative conditionals, which is not captured when one talks indiscriminately about whether 'disjunctions require a connection between the disjuncts'.⁴³

Table 1. Stimulus Materials, Mark Scenario illustrated with Disjunctions

	PO	NE	IR
HH	(Either) Mark presses the on switch on his TV OR his TV will be turned on	(Either) Mark lacks an appointment with the repairman OR his TV will work.	(Either) Mark is wearing socks OR his TV will work.
HL	(Either) Mark looks for popcorn OR he will be having popcorn.	(Either) Mark presses the on switch on his TV OR his TV will be turned off.	(Either) Mark is wearing socks OR his TV will malfunction.
LH	(Either) the sales clerk in the local supermarket presses the on switch on Mark's TV OR his TV will be turned on.	(Either) Mark pulls the plug on his TV OR his TV will be turned on.	(Either) Mark is wearing a dress OR his TV will work.
LL	(Either) Mark pulls the plug on his TV OR his TV will be turned off.	(Either) Mark refuses to look for popcorn OR he will be having popcorn.	(Either) Mark is wearing a dress OR his TV will malfunction.

Note. PO = positive relevance; NE = negative relevance; IR = irrelevance.

However, irrespectively of how this empirical issue is resolved, (III) still suggests that if the relevance effect on the probabilistic assessments of indicative conditionals is to be declared a pragmatic effect, we need to require that a suitable mechanism is specified which will lead to new predictions.

⁴³ However, one source of opposition against the claim that disjunctions have a negative relevance profile might be that one takes their meaning to be characterized by the or-introduction rule in natural deduction, which holds that a disjunction may be introduced in a proof whenever one of the disjuncts is true. But even granting this, it could still be argued that negative relevance is a conventional implicature of disjunctions which does not affect their truth conditional content (just as one way of interpreting the results from Chapter 4 is that the reason relation reading is a conventional implicature of indicative conditionals and 'therefore' and 'but' sentences, which (almost) does not affect their truth conditional content). Finally, that there should be a relationship between negative relevance and disjunctions is suggested by the fact that there are acceptable instances of inferences from 'A or C' to 'if non-A, then C' (see Skovgaard-Olsen (2016) for a discussion under which conditions such inferences are acceptable on a reason relation reading of the conditional).

Since Grice (1989) has a maxim of relevance, which Grice never elucidated further than “Be relevant!”, it is tempting to invoke it to account for relevance effects on indicative conditionals. However, it should be noted that relevance can be assessed at different levels and that whereas Grice’s maxim concerns the contribution of complete speech acts to a conversational context, the epistemic notion of relevance used in Chapters 2-4 concerned the internal relationship between two components in a sentence. That these evaluations of relevance can come apart is nicely illustrated by the example ‘If it snows in July, the Government will fall’ introduced in Douven (2015a). The point is that although this conditional violates the expectation that the antecedent is positively relevant for the consequent, there may nevertheless be a rhetorical point in making this assertion, which makes the assertion relevant as a speech act in the conversation. More specifically, the speaker may be interpreted as making the rhetorical point that it is so obvious that the consequent will hold that the consequent will hold no matter what happens (and thus even under such absurd circumstances as it snowing in July). In Douven (2015b) an argument is moreover made that the other Gricean maxims of informative and non-misleading conversation do not put us in a better position to account for the influence of relevance on our assessments of indicative conditionals.

Alternatively, other roles that have been assigned to pragmatics could be considered. As an example, Carston (2002: ch. 1) argues at some length that the semantic content of sentences in itself only suffices to provide a schema for a proposition and that processes of pragmatic interpretation apply even before a truth conditional content has been determined (by resolving reference assignments, ambiguities etc.). However, given that relevance was found only moderately to affect the truth evaluations of indicative conditionals in the FF cell in Chapter 4, it is unlikely that relevance assessments is a factor that enters directly into determining the propositional content of conditionals.

At this stage, further experiments are needed to determine whether support can be found for other pragmatic accounts or whether relevance is part of the

probabilistic, semantic content of indicative conditionals. Some initial first steps are outlined below.

Outlining Future Studies

The purpose of this section is to outline some ideas for future studies that will help us make progress on the issue of whether to classify the relevance effect on conditionals, reported in Chapter 2, as a pragmatic or semantic phenomenon.⁴⁴

One of the characteristics of conversational implicatures is that they can be cancelled without contradiction (e.g. the conversational implicature that the event described by the second conjunct followed the event described by the first conjunct temporally can be cancelled by saying ‘Peter and Ann got married and had a child, but *not necessarily in that order*’). Accordingly, if the relevance effect on conditionals is the result of a conversational implicature, it is expected that it should be cancellable without giving rise to a contradiction. Accordingly, assertions like ‘If A then C, *but A does not have anything to do with C*’ and ‘If A, then C, *but A is not in any way connected to C*’ should sound fine.

Furthermore, it is well-known that presuppositions differ from entailments in their projection behavior under embeddings (Beaver and Geurts, 2014). In contrast to entailments, presuppositions tend to be preserved in embeddings under negation (e.g. both the sentence ‘the King of France is bald’ and ‘the King of France is *not* bald’ wrongly presuppose that there is a King of France). Accordingly, if the relevance effect on conditionals is a presupposition, then it should be preserved under negations of conditionals. In contrast, if a reason relation assessment is part of the semantic value of a conditional, then it is to be expected that it interacts with the negation operator such that the negation of a conditional is to be understood as denying that A is a reason for C (which is true, if, for instance, A is irrelevant for C).

⁴⁴ In a subsequent study that is not part of this dissertation, experiments to implement some of these suggestions have been made (Skovgaard-Olsen, Collins, Krzyżanowska, Hahn, Klauer, ms).

In Stalnaker's possible worlds semantics for conditionals, the principle of Conditional Excluded Middle ($\models (A \rightarrow C) \text{ or } (A \rightarrow \neg C)$) is valid and so the following negation principle holds: $\neg(A \rightarrow C) \models A \rightarrow \neg C$ (Arlo-Costa, 2007). Moreover, the latter negation principle has also been invoked by proponents of the Suppositional Theory of conditionals to account for compound conditionals by providing paraphrases into sentences only containing basic conditionals (for discussion see Skovgaard-Olsen, 2016). In contrast, on a reason relation reading, it should be possible to negate a missing link conditional (e.g. 'It is not the case that if I will run out of milk tomorrow, then Merkel will give a speech about Syrian refugees') without thereby being committed to a conditional that negates the consequent of a conditional ('if I will run out of milk tomorrow, then Merkel will not give a speech about Syrian refugees').

Finally, it was suggested in (II) that the semantic value of a linguistic expression might depart from its evaluation by our intuitive truth value judgments, if assigning that semantic value to the expression allows us to build a compositional semantics that correctly accounts for its projection behavior. Accordingly, an avenue of further research would be into whether the internal relevance relations in sentences play a role in composing the semantic content of complex sentences in which they appear. In particular, the embedding behavior of conditionals under relevance conditionals could be investigated by examining principles such as the so-called Import-Export principle ($A \rightarrow (B \rightarrow C) \leftrightarrow (A \ \& \ B) \rightarrow C$). The Import-Export principle has been used by proponents of the Equation to account for right-nested conditionals (for discussion, see Skovgaard-Olsen, 2016) and it is also accepted as a valid argument schema by other conditional logics in the literature (see Arlo-Costa, 2007). However, on a reason relation reading of the conditional, counterexamples to the Import-Export principle can be found.

For instance, in the Sophia scenario cited in Chapter 2, one of the positive relevance conditionals sounded as follows: 'If Tim receives an orc costume, then he will be excited about his present'. But given that we have already accepted that conditional we might be inclined to accept the following slightly strange conditional as

well 'If Tim receives an orc costume, *and Sophia regularly wears shoes*, then Tim will be excited about his present'. For in spite of the fact that that added italicized conjunct is irrelevant for the consequent, the first conjunct remains probability raising for the consequent, and thus the conjunction in the antecedent as a whole remains probability raising for the consequent. However, accepting the latter conditional would not incline us to accept the following conditional, which is entailed according to the Import-Export Principle: 'If Tim receives an orc costume, then Tim will be excited about his present, *if Sophia regularly wears shoes*'. The reason is, of course, that the added italicized sentence is not probability raising for the sentence stating that Tim will be excited about his present. Accordingly, the reason relation reading gives us a general recipe for generating counterexamples to the Import-Export Principle, which could be investigated empirically to shed light on the question of whether the internal relevance of the parts in a sentence enters as a factor in composing the content of complex sentences.

Argument for Semantic Defect

In Skovgaard-Olsen (2016) arguments were presented for counting relevance part of the semantic content of indicative conditionals. Here I would like to focus on a particular argument, which focuses on the cognitive utility of the linguistically encoded content of normal conditionals and points out that missing-link conditionals are semantically defective, because they have a literal content that prevents them from fulfilling this cognitive role. In stating the argument, it was pointed out that the defect of missing link conditionals could not be limited to violations of Gricean norms, because Gricean norms pertain to conversational contexts, and missing-link conditionals are prevented from fulfilling the cognitive role of normal conditionals even in individual reasoning. In particular, it was pointed out that the appeal to indicative conditionals as "inference tickets" that give the right to infer the consequent

from the antecedent is blocked for missing link conditionals both in conversational contexts and in individual reasoning.

Here I would like to extend this argument by pointing to further aspects of the cognitive role of conditionals in individual reasoning which are blocked for missing link conditionals, because they require conditionals to express reason relations.

As explained in Chapter 1, according to the Suppositional Theory of conditionals the word 'if' is to be understood through its role in hypothetical thought of initiating imagination and simulation of possibilities (Evans and Over, 2004; Evans, 2007). This type of mental simulation is thought to play a central role in entertaining hypotheses, forecasting future events, and supporting decision making by imagining the consequences of alternative courses of action (Evans, Handley, Neilens, and Over, 2007). All of these mental processes are without doubt central to human thought and they have frequently been presented in the philosophical literature as relying on conditional reasoning.

A further central role of conditional reasoning is in argumentation, where inferential relations can be expressed by means of conditionals, which are often compared with 'condensed arguments' (Rescher, 2007; Krzyżanowska, 2015).

Now when Inferentialism puts the focus on conditionals' role in expressing reason relations is it then committed to denying the central role of conditionals in hypothetical thought as emphasized by the Suppositional Theory? No, because missing-link conditionals are just as useless in explanatory reasoning, forecasting, and decision making as they are in argumentation. When considering alternative explanatory hypotheses, predicting the future, and computing consequences of alternative courses of action, the agent needs to make assessment of which propositions are probability raising or probability lowering for other propositions. Because hypothetical thought is unbounded in that it can transcend the here and now and consider even remote possibilities of which there are infinite alternatives, propositions that don't make a probabilistic difference to the propositions of interest need to be set aside as irrelevant. It is a sign of rationality in hypothetical thought that

probabilistic dependencies are respected even if the basis of the reflection may depart from the actual course of events.

The point should be evident. But to illustrate, suppose the color of the socks of Stalin stood in the center of explanatory reasoning aimed at resolving why Operation Barbarossa turned out to be an utter failure, that the color of the socks of Angela Merkel was used to predict the outlines of the next European treaty, or that the premier minister of a European country used the color of the socks of other European leaders to calculate the consequences of alternative courses of actions guiding his/her decision making. The thought is so absurd that it is hard to take seriously.⁴⁵

What this illustrates is that the mental processes of suppositional reasoning are only useful as long as only hypotheses that preserve probabilistic dependencies are considered. In the Default and Penalty Hypothesis the conditional probability is per default processed in the PO condition as a way of assessing the sufficiency of the reason relation. As long as only probability raising scenarios are considered, the Ramsey test is an effective mental algorithm for engaging in hypothetical thought. However, if there are no constraints on which hypotheses it is applied to, then it will not in itself help us explain past events, predict the future, or decide among alternative courses of action.

In Rescher's (2007: 75) words: "conditionals effectively summarize the result of hypothetical inferences". And in making hypothetical inferences we are, of course, constrained by probabilistic dependencies that govern all other types of thought.

However, even if such reflections suggest that irrelevance is a semantic defect of conditionals, the jury is still out on empirically determining its precise nature. In Skovgaard-Olsen (2016), it was tentatively suggested that epistemic relevance could

⁴⁵ It is, of course, possible to restore sense in such examples by creating elaborate scenarios, where factors that initially appear irrelevant turn out surprisingly to be relevant after all. However, what this shows is not that probabilistic dependency plays no role for hypothetical thought but rather that it is so important for hypothetical thought that we invest substantial cognitive effort into restoring it.

be thought of as part of the sense-dimension of meaning characterizing its cognitive role.⁴⁶ In Chapter 4, we have seen a stark dissociation between the effect of relevance on probability and truth evaluations of indicative conditionals. Since presuppositions and conventional implicatures are distinguished by whether there is a dependency on their failure on the truth conditions of the sentences in which they occur (Potts, 2015), the arrow presently points in the direction of conventional implicatures. This conjecture is also supported by the linguistic arguments in McCawley (1993: ch. 15).

References

- Abrusán, M. and Szendrői, K. (2013). Experimenting with the King of France: Topics, verifiability and definite descriptions. *Semantics & Pragmatics*, 6(10), 1-43.
- Arlo-Costa, Horacio (2007). The Logic of Conditionals, in: *The Stanford Encyclopedia of Philosophy* (spring 2016 Edition), Edward N. Zalta (ed.). URL = <http://plato.stanford.edu/archives/fall2016/entries/logic-conditionals/>.
- Beaver, D. I. and Geurts, B. (2014). Presupposition. In: *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2014/entries/presupposition/>.
- Carston, R. (2002). *Thought and Utterances: The Pragmatics of Explicit Communication*. Oxford, Blackwell Publishing.
- Cruz, N., Baratgin, J., Oaksford, M. and Over, D.E. (2015). Bayesian reasoning with ifs and ands and ors. *Front. Psychol.*, 6, 192. doi: 10.3389/fpsyg.2015.00192
- Douven, I. (2015a). *The Epistemology of Indicative Conditionals: Formal and Empirical Approaches*. Cambridge: Cambridge University Press.
- Douven, I. (2015b). How to account for the oddness of missing-link conditionals. *Synthese*, 1–14.

⁴⁶ At the same time it was pointed out that it would have to be on an expanded notion of sense (*Sinn*), whereby it didn't play a role in determining reference (*Bedeutung*), as in Frege (1892).

- Evans, J. St. B. T. (2007). *Hypothetical Thinking: Dual Processes in Reasoning and Judgment*. New York: Psychology Press.
- Evans, J. St. B. T., Handley, S. J., Neilens, H., and Over, D. E. (2007). Thinking about conditionals: A study of individual differences. *Memory & Cognition*, 35(7), 1772-84.
- Evans, J. St. B. T. & Over, D. (2004), *If*. Oxford: Oxford University Press.
- Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100, 25-50.
- Grice, P. (1989). *Studies in the Way of Words*. Cambridge, MA.: Harvard University Press.
- 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>>.
- Künne, W. (2005). *Conceptions of Truth*. Oxford: Oxford University Press.
- McCawley, J. D. (1993). *Everything that Linguistics have Always Wanted to Know about Logic—but were ashamed to ask*. Chicago: The University of Chicago Press. (Second Edition)
- Moss, S. (2015). On the Semantics and Pragmatics of Epistemic Vocabulary. *Semantics and Pragmatics*, 8 (5), 1-81.
- Nickerson, R. S. (2015). *Conditional Reasoning. The Unruly Syntactics, Semantics, Thematics, and Pragmatics of "if"*. Oxford: Oxford University Press.
- Oberauer, K., Geiger, S. M., Fischer, K., and Weidenfeld, A. (2007). Two meanings of „if“? Individual differences in the interpretation of conditionals. *The Quarterly Journal of Experimental Psychology*, 60(6), 790-819.
- Over, D. E. and Cruz, N. (forthcoming). Probabilistic accounts of conditional reasoning. To appear in: Linden J. Ball and Valerie A. Thompson (Eds.), *International Handbook of Thinking and Reasoning*. Hove, Sussex: Psychology Press.

- Potts, C. (2015). Presupposition and Implicature. In: Lappin, S. and Fox, C. (eds.), *The Handbook of Contemporary Semantic Theory* (2nd edn). Oxford: Wiley-Blackwell, 168-202.
- Rescher, N. (2007). *Conditionals*. Cambridge, MA.: The MIT Press.
- Skovgaard-Olsen, N. (2016). Motivating the Relevance Approach to Conditionals. *Mind & Language*, 31(5), 555-79.
- Skovgaard-Olsen, N., Kellen, D., Hahn, U., and Klauer, K. C. (ms). Norm Conflicts and Conditionals.
- Skovgaard-Olsen, N., Collins, P., Krzyżanowska, K., Hahn, U., Klauer, KC (ms). Cancellation, Negation, and Rejection.
- von Stechow, P. (2004). Would you believe it? The king of France is back! In: Reimer and Bezuidenhout (ed.), *Descriptions and Beyond*. Oxford: Clarendon Press, 269-96.
- Winter, Y. (2016). *Elements of Formal Semantics. An Introduction to the Mathematical Theory of Meaning in Natural Language*. Edinburgh: Edinburgh University Press.
- Yalcin, S. (2012). A Counterexample to Modus Tollens. *Journal of Philosophical Logic*, 41 (6), 1001-24.

CURRICULUM VITAE

DR. NIELS SKOVGAARD-OLSEN

niels.skovgaard.olsen@psychologie.uni-freiburg.de

n.s.olsen@gmail.com

Talstrasse 10, 79102 Freiburg,
Germany

Date of birth: 30.07.1985

Køge, Denmark

Danish citizen

Education and Positions

- 2015 Visiting Scholar at Berkeley University.
- 2014- Postdoc with Professor Dr. Wolfgang Spohn as mentor. Project: *Reason Relations, Argumentation, and Conditionals: Applying Ranking Theory to Psychology of Reasoning*.
- 2014- Ph.D. Student (Psychology), University of Freiburg.
Supervisor: Professor Dr. Karl Christoph Klauer.
- 2014- Associate, *What if* – an interdisciplinary research group on conditionals and thought experiments with researchers from the fields of philosophy, linguistics, and history.
- 2012-4 Ph.D. (philosophy), University of Konstanz. Supervisors: Professor Dr. Wolfgang Spohn (philosophy), Professor Dr. Sieghard Beller (psychology). Dissertation defended with *summa cum laude* (“highest honour”).
- 2011-12 Studies in psychology, University of Konstanz.
- 2010 MA, philosophy, University of Copenhagen.

Other Experiences

- 2016-2017 Research visits at *Munich Center for Mathematical Philosophy* and *Birkbeck University* to collaborate with Prof. Dr. Ulrike Hahn.
- 2016 Participant in International Rationality Summer Institute 2016
- 2015 Award: Preis des Landkreises Konstanz zur Förderung des wissenschaftlichen Nachwuchses an der Universität Konstanz. (Award for excellent dissertations in philosophy and history.)
- 2014 Funding for 3 year post-doc project as part of *New Frameworks of Rationality*, DFG Schwerpunktsprogramm SPP1516.
- 2012 Scholarship from *New Frameworks of Rationality*, DFG Schwerpunktsprogramm SPP1516.

Talks

- Upcoming* CogSci 2017, Annual Meeting of the Cognitive Science Society, London
ECAP9, European Congress of Analytic Philosophy, München
- Invited Talks* 2016, Roundtable with Igor Douven (Konstanz)
2015, Gießener Abendgespräche Kognition und Gehirn
2015, Berkeley, Working Group in the History and Philosophy of Logic, Mathematics and Science
2014, 2015, 2016, 2017 annual meeting of New Frameworks of Rationality
- Conferences* Reasoning Club 2017, Turin
/Workshops 2016, ICT-2016, Brown University.
2015, 2016 What-if research group, University of Konstanz.
2014, AISB-50, London University
2012, 2013, 2015, Danish annual meeting of philosophy
2013, Public lecture, University of Konstanz
- Topics* Experimental work on conditionals.
The problem of logical omniscience.
The compositionality of conditionals.
Perceptions and reasons.
How to make philosophical theories useful for scientific purposes.

Ranking theory and conditionals.
Relevance and Reason relations.
Semantic objectivity.
Norms conflicts and attributing reasoning errors.

Teaching Experiences

- 2016 Seminar on Conditionals in Philosophy and Psychology with Dr. Eric Raidl, University of Konstanz.
- 2015-2017 Sessions on Rationality and Psychology of Reasoning for MSc students, Psychology Department, University of Freiburg.
- 2015-2017 Supervision of 1) BSc. project, 2) research internship, and 3) experimental group of psychology students, Psychology Department, University of Freiburg.

Peer Review

Review Editor of *Frontiers in Psychology: Theoretical and Philosophical Psychology*.

Ad-hoc reviewer for: *Synthese, Journal of Cognitive Psychology, Episteme, Cognition, Logique et Analyse, Memory and Cognition*.

Organization

- 2016 Organization of Symposium ‚Normativity and Rationality‘ with Dr. Shira Elqayam for the International Conference of Thinking (ICT2016) Brown University.
- 2013 Organization of 1) Journal Club and 2) Reading Group for PhD-Students and Post-Docs at University of Konstanz with Dr. Johannes Schmitt and Dr. Giuliano Bacigalupo.

Papers

Impact Factor: 12.32

Skovgaard-Olsen, N., Singmann, H., and Klauer, K. C. (2016), ‚Relevance and Reason Relations‘, in: *Cognitive Science*. doi: 10.1111/cogs.12462

- Raidl, E. and Skovgaard-Olsen, N. (2016), 'Bridging Ranking Theory and the Stability Theory of Belief', in: *Journal of Philosophical Logic*. doi: 10.1007/s10992-016-9411-0
- Skovgaard-Olsen, N. (2016), 'Motivating the Relevance Approach to Conditionals', in: *Mind & Language*, 31 (5), 555-79. doi: 10.1111/mila.12120
- Skovgaard-Olsen, N., Singmann, H., and Klauer, K. C. (2016), 'The Relevance Effect and Conditionals', in: *Cognition*, 150, 26-36.
[doi:10.1016/j.cognition.2015.12.017](https://doi.org/10.1016/j.cognition.2015.12.017)
- Skovgaard-Olsen, N. (2015a), 'Ranking Theory and Conditional Reasoning', in: *Cognitive Science*. doi: 10.1111/cogs.12267
- Skovgaard-Olsen, N. (2015b), 'The problem of logical Omniscience, the preface paradox, and doxastic commitments', in: *Synthese*. doi: 10.1007/s11229-015-0979-7
- Olsen, N. S. (2014), 'Philosophical Theory-Construction and the Self-Image of Philosophy', in: *Open Journal of Philosophy*.
- Olsen, N. S. (2010), 'Reinterpreting Sellars in the Light of Brandom, McDowell, and A. D. Smith', in: *European Journal of Philosophy*, 18 (4), 510-38.
doi: 10.1111/j.1468-0378.2009.00360.x

Dissertations

- Skovgaard-Olsen, N. (2017), *Putting Inferentialism and the Suppositional Theory of Conditionals to the Test*. (Psychology Dissertation, University of Freiburg)
- Olsen, N. S. (2014), *Making Ranking Theory useful for Psychology of Reasoning*. (Philosophy Dissertation, University of Konstanz)
URL = <http://nbn-resolving.de/urn:nbn:de:bsz:352-0-262692>

Conference Proceedings

Skovgaard-Olsen, N., Kellen, D., Hahn, U., and Klauer, K. C. (in press), 'Conditionals, Individual Variation, and the Scorekeeping Task', in: M. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*. London, UK: Cognitive Science Society.

Skovgaard-Olsen, N. (2014), 'Logical Omniscience and Acknowledged vs. Consequential Commitments', in: *Proceedings of AISB50*. London, UK: Society for the Study of Artificial Intelligence and Simulation of Behavior.

Book Reviews

Skovgaard-Olsen, N. (2017), 'Sellars and His Legacy Ed. By James O'Shea (Book Review)', in: *Journal of the History of Philosophy*, 55 (2), 358-59.

In Review and under Preparation

Skovgaard-Olsen, N., Kellen, D., Krahl, H., and Klauer, K. C. (forthcoming), 'Relevance differently affects the truth, acceptability, and probability evaluations of 'And', 'But', 'Therefore', and 'If Then''. (*Thinking and Reasoning*)

Skovgaard-Olsen, N. (invited book chapter), 'Relevance and Conditionals: A Synopsis', for *Festschrift for David Over*, Routledge Press.

Spohn, W., Skovgaard-Olsen, N., and Kern-Isberner, G. (invited book chapter), 'Ranking Theory', for *Handbook of Rationality*, The MIT Press.

Skovgaard-Olsen, N. (ms), 'Compositional Reason Relation Semantics'.

Skovgaard-Olsen, N., Kellen, D., Hahn, U., and Klauer, K. C. (ms), 'Norm Conflicts and Conditionals'

Skovgaard-Olsen, N., Collins, P., Krzyżanowska, K., Hahn, U., Klauer, K. C. (ms), 'Cancellation, Negation, and Rejection'

Raidl, E. and Skovgaard-Olsen, N. (ms), 'Simulating Lewis/Stalnaker Conditionals in Ranking Theory'

EIDESSTATTLICHE ERKLÄRUNG (DECLARATION OF ORIGINALITY)

Ich erkläre hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungsbeziehungsweise Beratungsdiensten (Promotionsberater oder anderer Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

(I hereby declare that this dissertation is my own work and that all the sources that I have used or quoted have been acknowledged by means of complete references. This work has not been submitted previously for a degree at any university or other academic institution.)

Freiburg, den 25.09.2017