PRAGMATIC USE & INTERPRETATION OF CONDITIONALS

Dissertation
zur Erlangung des Doktorgrades
des Fachbereichs Humanwissenschaften
der Universität Osnabrück

vorgelegt
von

BRITTA GRUSDT

aus
München

Osnabrück, May 2023

SCIENTIFIC SUPERVISORS:
Prof. Dr. Michael Franke
Prof. Dr. Mingya Liu
NON-SCIENTIFIC SUPERVISORS:
Prof. Dr. Gordon Pipa
CO-EXAMINER:
Prof. Dr. Nicole Gotzner

DATE OF THE DISPUTATION: SEPTEMBER 28, 2023

ABSTRACT

'If'-sentences can be used to express many different beliefs and intentions, including, among others, warnings, promises, causal relations, or the speaker's uncertainty with respect to the propositions mentioned in the conditional, an expression of the form 'if . . . , (then) . . . '. Thus, speakers use conditionals in a wide range of situations and listeners easily interpret them accordingly. However, it remains an unsolved question how such diverse interpretations of the word 'if' arise. Across a number of different disciplines, people, therefore, aim to find answers to the supposedly simple question of what 'if' means. In this thesis, I investigate the pragmatic meaning of conditionals. In the first part of the thesis I present a probabilistic Bayesian model, more precisely a Rational-Speech-Act model, building on recent developments in computational pragmatics. We show that pragmatic processes formalized by the model, in combination with a representation of rich structural prior beliefs of the interlocutors, can explain common observations within the communication with conditionals. These include, among others, the usually inferred relation between the propositions mentioned in the conditional ('If A, C' suggests that A and C are related) as well as the infelicity of so-called *missing-link* conditionals which, as suggested by their name, lack this relation entirely. These are two observations that have been the focus of a current debate about whether, and if so, how, it is possible to explain them within pragmatics instead of ascribing them to the semantics of conditionals. In the second part of the thesis, I first present experimental data, collected in a behavioral experiment that we designed, on speakers' use of conditionals. Then, I compare this empirical production data with the quantitative predictions of our model, which shows that the model is able to account for major aspects observed in the data. The third part of the thesis addresses a particular phenomenon called *conditional perfection* which refers to the interpretation of a conditional 'if A, (then) C' as biconditional 'if and only if A, (then) C'. In this part, I focus on the empirical side again, presenting two behavioral experiments with which we aim to test an account proposed by von Fintel (2001) about the influence of the focus of a conversation, a so-called *Question under discussion*, on the listener's interpretation of a conditional, in particular as biconditional.

DEUTSCHE ZUSAMMENFASSUNG

"Wenn, dann"-Sätze können verwendet werden, um viele verschiedene Überzeugungen und Absichten auszudrücken, unter anderem Warnungen, Versprechen, kausale Beziehungen oder die Unsicherheit des Sprechers in Bezug auf die im Konditional, ein Ausdruck der Form "Wenn . . . , (dann) . . . ", genannten Aussagen. Die Sprecher verwenden also Konditionale in einer Vielzahl von Situationen, wobei Hörer diese problemlos entsprechend interpretieren. Wie es zu solch unterschiedlichen Interpretationen des Wortes "wenn" kommt, ist jedoch offen. In verschiedenen Disziplinen versucht man daher, Antworten auf die vermeintlich einfache Frage zu finden, was "wenn" bedeutet. In dieser Arbeit untersuche ich die pragmatische Bedeutung von Konditionalen. Im ersten Teil der Arbeit stelle ich ein probabilistisches Bayessches Modell vor, genauer gesagt ein Rational-Speech-Act-Modell, welches auf jüngsten Entwicklungen in dem Bereich der computergestützten Pragmatik aufbaut. Wir zeigen, dass pragmatische Prozesse, die durch das Modell formalisiert werden, häufige Beobachtungen in der Kommunikation mit Konditionalen erklären können, wenn das Weltwissen der Gesprächspartner, insbesondere bezüglich der strukturellen Zusammenhänge der modellierten Variablen, adäquat repräsentiert wird. Zu den Beobachtungen, die wir mit dem Modell erklären können gehören u.a. die üblicherweise gefolgerte Beziehung zwischen den im Konditional erwähnten Propositionen ("Wenn A, C" suggeriert, dass es einen Zusammenhang zwischen A und C gibt) sowie die Tatsache, dass sogenannte *missing-link*-Konditionale, denen diese Beziehung, wie ihr Name vermuten lässt, völlig fehlt, keine sinnvollen Aussagen darstellen. Diese beiden Beobachtungen stehen im Mittelpunkt einer aktuellen Debatte darüber, ob, und wenn ja, wie es möglich ist, sie innerhalb der Pragmatik zu erklären, anstatt sie der Semantik von Konditionalen zuzuschreiben. Im zweiten Teil der Arbeit stelle ich zunächst experimentelle Daten vor, die in einem von uns konzipierten Verhaltensexperiment zum Gebrauch von Konditionalen erhoben wurden. Anschließend vergleiche ich diese empirischen Daten mit den quantitativen Vorhersagen unseres Modells, was zeigt, dass das Modell in der Lage ist, die wichtigsten in den Daten beobachteten Aspekte zu erklären. Der dritte Teil der Arbeit befasst sich mit einem besonderen Phänomen, welches als *conditional perfection* bezeichnet wird und sich auf die Interpretation eines Konditionals "Wenn A, (dann) C" als Bikonditional "Wenn und nur wenn A, (dann) C" bezieht. In diesem Teil konzentriere ich mich wieder auf die empirische Seite, indem ich zwei Verhaltensexperimente vorstelle, mit denen wir eine Theorie von von Fintel (2001) über den Einfluss des Gesprächsfokus, einer so genannten *Question under Discussion*, auf die Interpretation eines Konditionals, insbesondere als Bikonditional, testen.

PUBLICATIONS

Chapters 3 to 5

Grusdt, B., Lassiter, D., & Franke, M. (2022). Probabilistic modeling of rational communication with conditionals. *Semantics and Pragmatics*, *15*. https://doi.org/10.3765/sp.15.13

Chapter 6

Grusdt, B., & Franke, M. (2021). Communicating uncertain beliefs with conditionals: Probabilistic modeling and experimental data, In *Proceedings of the Annual Meeting of the Cognitive Science Society*.

Chapter 8

Grusdt, B., Liu, M., & Franke, M. (2022). Testing the Influence of QUDs on Conditional Perfection, In *Experiments in Linguistic Meaning* 2. https://doi.org/10.3765/elm.2.5413

ACKNOWLEDGEMENTS

Needless to say, this thesis would not have been possible without my supervisors. Thank you, Michael and Mingya, for all the time that you spent with me. I am very grateful for all that I could learn from you, on a professional and also a personal level. It was a true pleasure to work with both of you. Thank you, Mingya, for having integrated me into your group in Berlin and having motivated me to present my work on a regular basis. Thank you, Michael, for all the motivation and inspiration I gained from our meetings, for the patience that you had with me, and all your highly constructive, unadorned, yet always appreciative and motivating feedback — I would not have learned as much without it.

Then, I would like to thank Daniel Lassiter, who is co-author of my main paper for this thesis. Thank you for your support and for having made this paper possible! A big thank you also goes to my first academic advisor, Frank Jäkel. Who knows whether I would have ended up doing a PhD if it was not for the great experience that I had when writing my Bachelor thesis with Frank. Further, I would like to thank Malin Spaniol and Josefine Zerbe for their great support in implementing my experiments — I wish you all the best for your own PhD research! Talking about experiments, thanks to everyone who took part in my studies for the time and energy that they spent to help me with their participation.

Supervisors and colleagues are, of course, indispensable for writing a thesis. Yet, this thesis would neither have been possible without my friends and family and their emotional support that helped me to eventually finish writing it. A few of them, who were particularly important to me during my time as a PhD student, I would like to mention explicitly, in no meaningful order.

A big thank you goes to Malin Spaniol, Tabea Kossen, Pablo Prietz, Xenia Ohmer and Vinicius Macuch who all read parts of this thesis and provided very helpful feedback that certainly helped to make the thesis better. Thank you, Georg, for being a great colleague and friend, and for having shown me Hanabi, one of the best games I know — we will hopefully keep on playing and thinking about it in Berlin! Thank you, Betty, for the countless hours that we spent together virtually motivating each other to keep on writing. I really don't want to imagine how it would have been without your company during the last couple of months. I wish you all the best for your defense and am looking forward to celebrating with you! Thank you,

Xenia, for the great (but way too short) time that we had as office-mates and more important than that, thank you for being a wonderful friend! I am very grateful that our ways crossed each other once more in Osnabrück, and they will hopefully do again elsewhere. I will miss you! Thank you, Tabea, for having insisted in going to Chile in 2022 — without you I would probably not have paused writing my thesis for a month and would have missed this wonderful trip. Thank you, Jana, for having been a wonderful flatmate and friend, my time in Osnabrück would certainly have been much less exciting without you. Thank you, Pablo, for your immense support during the last one and a half years since we met. I am very grateful for all that you did for me!

Last but not least I would like to thank my family. Thank you for supporting me in everything that I do, and in particular for giving me not only shelter and countless delicious meals during the Covid-19 lockdowns but also for your great company during these otherwise lonely times.

CONTENTS

LIST OF TABLES

Part I

INTRODUCTION & BACKGROUND

1 OVERVIEW

The fact that this entire thesis deals with the little word "if", like many theses did before, already gives an idea of the special role that it takes on. Part of what makes it so special are the diverse interpretations that are observed in the communication with conditionals. Consider the following examples:

(1) If Ann goes to the party, Bob will go there, too.

(2) If it rains, the grass will be wet.

(3) If you happen to be in Osnabrück in autumn, you will need an umbrella.

The conditional in (1) suggests that Bob will *not* go to the party in case that Ann *does not* go. One may analogously infer from the conditional in (2) that the grass will *not* be wet in case that it *does not* rain. However, this inference seems less strong since, depending on the context in which the conversation takes place, other reasons for the grass to be wet easily come to mind (e.g., someone dropping a bottle of water). The speaker of (3) seems to communicate even less, if anything at all, about the case when the proposition in the if-clause, the *antecedent*, is false. In this example, the speaker only seems to convey that it usually rains a lot in Osnabrück during autumn and that thus, the addressee might need to bring an umbrella when visiting the speaker in autumn in Osnabrück. What the speaker conveys in all three examples equally is the uncertainty about the antecedent: the speaker does not seem to know whether Ann goes to the party, whether it rains or whether the addressee will be in Osnabrück in autumn. Neither does the speaker seem to know whether the proposition in the main-clause, the *consequent*, is true. The different interpretations of these few examples already bring up the question how to define the meaning of the word "if".

Semantics and pragmatics are the two large subfields in linguistics that investigate meaning. While semanticists look at the meaning of words and phrases in isolation, pragmatists also take into account the context in which an utterance is made; that is, pragmatics studies the meaning of sentences that speakers utter when interacting with a listener. When taking into account the context, an utterance may be perceived as false, or better-to-say misleading or inappropriate, although it might be objectively true. An example, is given in (4).

(4) He cleaned the kitchen.

Assuming that the person the speaker is referring to did clean the kitchen, (4) is true. If that person also cleaned the bathroom, this utterance yet seems misleading as it suggests that *only* the kitchen was cleaned. It is this meaning, that the speaker *communicates* beyond the actually uttered words, that is studied in pragmatics.

This thesis is concerned with the pragmatics of conditionals, seeking to gain insights of how the diverse interpretations of conditionals like in (1)–(3) may arise when taking into account the interaction between interlocutors. Although conditionals have been (and are still) studied extensively across disciplines, including linguistics (e.g., Kaufmann, 2023; Lassiter, 2018a; Schulz, 2015), psychology (e.g., Evans et al., 1993; Oaksford & Chater, 2020; Wason, 1968), logic and philosophy (e.g., Adams, 1975; Bennett, 2003), they remain one of the most elusive natural language constructs to provide an account of meaning for, quite in contrast to other logical expressions like 'some' or 'or'. Conditionals attract researchers from a number of different disciplines because they touch on various phenomena: to name a few, they are used to reason about hypothetical worlds, to formulate wishes, promises or warnings, to communicate causal relations, and everyday decisions are based on conditionals. In short, conditionals form an essential part of human reasoning.

The main contribution of the thesis is a formal computational model that builds on recent advances in the field of computational pragmatics (Frank & Goodman, 2012; Goodman & Stuhlmüller, 2013). The model makes quantitative predictions (i) about the situations in which speakers are likely to utter conditionals and in which they rather choose utterances without conditional structure and (ii) about the listener's interpretation of the speaker's selected utterance and how this interpretation is influenced by the speaker's alternative utterances. Moreover, I present empirical data from three behavioral experiments, one of which was conducted to put the proposed model to a test. With the other two experiments, we investigate the phenomenon that people sometimes interpret 'if' as 'if and only if', known as *conditional perfection* (Geis & Zwicky, 1971). Note that we exclusively consider English conditionals here.[1]

The structure of the thesis is as follows. I start with an introduction of relevant previous work on conditionals in Chapter 2, in particular on their semantics and pragmatics. The second part of the thesis comprises Chapter 3-7 which concern our computational model. The necessary background is given in Chapter 3 in which I will first introduce the vanilla version of the model, and then discuss the adaptations necessary for making it suitable to model communication with conditionals. In Chapter 4 and Chapter 5 I then present the predictions of our model, assuming general, unspecific utterance contexts

1 For cross-linguistic studies on conditionals see for example, Comrie (1986), Zaefferer (1991).

in the former and some concrete contexts in the latter chapter. In Chapter 6 I present the behavioral experiment that we developed to test the proposed model and describe the collected empirical data which is subsequently modeled in Chapter 7. The third part of the thesis concerns the phenomenon conditional perfection — the interpretation of 'if' as 'if and only if'. I present two empirical studies in Chapter 8 which we ran to investigate factors that possibly elicit a conditional perfection reading. Lastly, I summarize the contributions of my work as well as its limitations and discuss its predictions in light of other empirical studies in Chapter 9 and terminate with a short conclusion in Chapter 10.

2 BACKGROUND ON CONDITIONALS

2.1 WHAT DOES "IF" MEAN?

Before delving into various aspects of conditionals, let me set the boundaries for what general types of conditionals there are and which of them this thesis is about.

The broadest distinction of conditionals considers two main types, *indicative* conditionals on the one hand and *subjunctive* conditionals on the other hand (e.g., see Bennett, 2003). The former are conditionals like those in (1)–(3) and (5), while (6) is an example for the latter.

(5) If it did not rain, someone must have watered the plants.

(6) If Osnabrück was in the south of Spain, it would rain less.

Subjunctive conditionals usually communicate the falsity of the antecedent; (6), for instance, implicates that Osnabrück is not located in the south of Spain. Indicative conditionals, on the other hand, usually (but not always, see Section 5.4.2) communicate neither the truth nor the falsity of the antecedent; (5), for instance, leaves open whether or not it rained. Subjunctive conditionals are often interchangeably referred to as *counterfactual* conditionals (e.g., see Bennett, 2003, p.12), referring to the suggested implication that the antecedent is false, even though this is not necessarily the case (e.g., see Anderson, 1951, p.36).[1]

What we can note is that indicative conditionals concern the actual world whereas subjunctive conditionals concern possible alternative worlds to the actual world. This is also reflected in the grammatical tenses used in both types respectively (e.g., see Starr, 2021); only subjunctive conditionals use the modal 'would' in the consequent (e.g., 'would' or 'would have').

Whether subjunctive and indicative conditionals have different underlying semantics and should be treated as two different kinds of conditionals is still debated. Traditionally, a clear distinction has been made between them (e.g., Bennett, 2003; Jackson, 1979), but recently some researchers have argued for a uniform treatment of both (e.g., Schulz, 2015).

This thesis focuses on indicative conditionals. Therefore, I will write "conditional" to refer to *indicative* conditionals, more precisely to in-

1 Anderson (1951) gives the following example to show that subjunctive conditionals do not need to communicate the falsity of the antecedent: "If Jones had taken arsenic, he would have shown just exactly those symptoms which he does in fact show." As Anderson notes, this conditional, assumed to be uttered by Jones' doctor, may even suggest the *truth* of the antecedent.

p	q	$p \supset q$	$p \Leftrightarrow q$	deFinetti
T	T	T	T	T
T	F	F	F	F
F	T	T	F	void
F	F	T	T	void

Table 1: Truth conditions of the material implication ($p \supset q$), the biconditional ($p \Leftrightarrow q$) and according to the *defective truth table* (De Finetti, 1995).

dicative conditionals whose antecedent and consequent are simple propositions like in (1) to (5). This particularly excludes conditionals whose antecedent and/or consequent are, for instance, conditionals themselves. A final note on the notation used throughout the thesis. As commonly done, I will abbreviate (English) natural language conditionals by a simple arrow ($\rightarrow$), whereby the proposition in the antecedent is denoted by A and the proposition in the consequent by C (i.e., $A \rightarrow C$ denotes a conditional "If A, (then) C").

2.1.1 *Semantics of conditionals*

In truth-conditional semantics, meaning is defined by specifying the conditions when a proposition is true; the truth value of a sentence or phrase is determined by an evaluation of the truth values of its composites. Formal accounts of natural language meaning are traditionally rooted in logical analysis and therefore pay attention specifically to sentential connectives. A sentence "A and C", for instance, evaluates as true if and only if both propositions, that is, A and C, are true — corresponding to the truth conditions of the logical *and* ($\wedge$). While natural language conjunctions (Blakemore & Carston, 2005), disjunctions (Simons, 2001) and negation (Horn, 1989) all feature their own respective subtleties, possibly deviating from a classical logical analysis, natural language conditionals are among the most elusive expressions to provide an account of meaning for.

In the following, I will consider several semantics proposed for conditionals, starting with the logical if.

CONDITIONALS & TRUTH CONDITIONS. Table 1 displays the truth conditions of the logical 'if', called *material implication* that is denoted by the horseshoe symbol $\supset$. When the material implication is assumed as semantics for natural language conditionals "*if . . . , then . . .*", the conditional is referred to as *material conditional*. That is, in this case "*if A, then C*" evaluates as true for all combinations of truth conditions of A and C except for the case in which A is true and C is false. This semantic, however, evaluates many conditionals as true that are intuitively considered false; examples are given below.

The general inferences that follow from the material implication but often do not match peoples interpretation of "*if ..., then ...*" are summarized as the *paradoxes of the material implication* (e.g., see Egré & Rott, 2021). Since the material conditional is true whenever the consequent is true, the truth of $A \supset C$ follows, for example, from the truth of C — independently of A. Therefore, under the material implication, the truth of "Bob's mum is happy", for instance, implies that "If her son does not pass the final exam, Bob's mum is happy", which is not very reasonable. Similarly, $A \supset C$ is true whenever the antecedent is false, also yielding nonsensical inferences when applied to natural language conditionals. Another example for a well-known paradox of the material implication is given in (7) and (8) — known as *Strengthening the antecedent*: from the truth of $A \supset C$ (e.g., 7), we can conclude the truth of $A \land B \supset C$ for any B (e.g., (8)).[2]

(7) If you do sports regularly, you lead a healthy life.

(8) # If you do sports regularly and eat a pizza a day, you lead a healthy life.[3]

Some of the paradoxes of the material implication are avoided in C. I. Lewis's (1912) *Strict conditional*-account, according to which a conditional evaluates to true *if and only if* in *all* worlds in which the antecedent is true, the consequent is true as well; put differently, the material conditional is necessarily true (see Egré & Rott, 2021).

Under the assumption that conditionals have truth conditions, which remains an open issue (see Rothschild, 2015, for a review), it is uncontroversial that a conditional $A \rightarrow C$ is false when the antecedent (A) is true and the consequent (C) is false — this is maybe the single aspect of conditionals that is not debated. The classical semantics for conditionals based on truth conditions agree on $A \rightarrow C$ being true when both, antecedent and consequent are true. However, this case raises questions, in particular for conditionals like (9) where the antecedent and the consequent are true facts but have no relation whatsoever, rendering (9) a *missing-link* conditional, which I will discuss in light of our model in Chapter 5 (Section 5.4.1).

(9) # If Paris is the capital of France, Queen Elizabeth died in 2022.

Less clear for truth-conditional semantics of conditionals are the situations when the antecedent is false. Reconsider examples (1) and (2) from the beginning, repeated here as (10) and (11).

(10) If Ann goes to the party, Bob will go there, too.

2 If $A \supset C$ is true, it follows that it is not the case that A is true and C is false. The only case where $A \land B \supset C$ is false is when the antecedent, $A \land B$, is true and the consequent, C, is false. But since $A \land B$ is only true when A is true and B is true, and knowing that it is *not* the case that C is false when A is true (due to the truth of $A \supset C$), the truth of $(A \land B) \supset C$ is implied whenever $A \supset C$ is true.

3 Examples that are infelicitous are marked with a preceding '#'.

(11) If it rains, the grass will be wet.

While (11) seems to remain true, independently of the actual whether, it seems more difficult to evaluate the truth condition of (10) when it is known that Ann does not go to the party. Under a material implication account a conditional is true as soon as its antecedent is false, accordingly in that case (10) is true. Another possibility is to define the truth conditions of a conditional only in case that the antecedent is true and declare them as 'void' when the antecedent is false, which was done by De Finetti (1995). This definition is also known as *defective truth table* (see Table 1).

Under the *possible-world* semantics from Stalnaker (1968) (10) can either be true or false, depending on other aspects of the world. Stalnaker proposed to define the truth conditions of a conditional $A \rightarrow C$ based on a possible world in which the antecedent A is true and which is only minimally different from the actual world. Therefore, when the antecedent is false in the actual world, the conditional may evaluate to true or false — depending on the truth value of the consequent C in the considered possible world in which A is true and which is minimally different from the actual world; if in that case, C is false, the conditional is false, but if C is true, the conditional is true. In case that in the actual world the antecedent is true and the consequent is false, the conditional is false, and in case that the antecedent and the consequent are both true in the actual world, the conditional is true as well.

The last semantic account of conditionals that I would like to mention is the *restrictor analysis* by Kratzer (1986), which is based on work of D. Lewis (1975) and Heim (1988) and has gained a lot of attention in linguistics. Kratzer does not treat *if* as a sentential connective, but as an operator that by itself does not add any meaning to a sentence. Instead she considers it a restrictor of the modal or frequency operator in the main clause (e.g., must, rarely, usually, etc.). For the conditional "If it is not in the kitchen it *must* be in the bathroom", 'must' is the overt modal operator that is restricted by the 'if'-operator (see Edgington, 2020). Importantly Kratzer also explicitly allowed the modal operator to be a covert operator, so that conditionals without modal operator are not excluded from the analysis.

CONDITIONALS & PROBABILITIES. One of the first accounts of conditionals that explicitly draws on probabilities goes back to Adams (1965). According to Adams, "*the assertion of a conditional statement [...] is probabilistically justified if the likelihood that both the antecedent and consequent are true is much greater than that the antecedent is true and the consequent false*" (p.175 Adams, 1965); put differently, a conditional is deemed assertable when the corresponding conditional probability exceeds a certain threshold. This proposition has become known as

Adam's thesis.[4] Similar proposals were put forward that also relate a conditional $A \to C$ to the conditional probability $P(C \mid A)$ (e.g., Jackson, 1979; Stalnaker, 1970), yet differ in important ways. Stalnaker (1970), for instance, combined his earlier possible-world account with conditional probabilities, but argued that the conditional probability $P(C \mid A)$ is the probability of the conditional $A \to C$ to be *true*:

$$P(A \to C) = P(C \mid A), \text{ where } P(A) > 0 \qquad \textit{Stalnaker's hypothesis}$$

Contrary to that, Adams (1965) denied conditionals to have truth conditions at all, arguing instead that the conditional probability is the extent to which the conditional is assertable/acceptable. This is an important difference in light of the so-called *triviality results*. D. Lewis (1976, 1986) and Hájek (1989) showed that it is impossible for a proposition $A \to C$ to equal the conditional probability $P(C \mid A)$ unless one is willing to accept rather absurd consequences; in the uncontroversial case that $P(A, C) > 0$ and $P(A, \neg C) > 0$, the probability of the conditional, $P(A \to C)$, will for instance equal the probability of the consequent, $P(C)$, when assuming Stalnaker's hypothesis ($P(A \to C) = P(C \mid A)$).[5] Since the triviality results are based on the assumption that conditionals are propositions with truth conditions, they are not problematic for Adams' account as he denies this assumption in the first place.

The first, albeit indirect, reference of the relation between conditionals and the corresponding conditional probability goes back to Ramsey (1931), which is also mentioned by Stalnaker to have motivated his possible-worlds theory. In the following quote from Ramsey, he describes his view on how conditionals are processed which has become known as *the Ramsey Test*.

> *If two people arguing "If* p*, then* q*?' and are both in doubt as to* p*, they are adding* p *hypothetically to their stock of knowledge and arguing on that basis about* q*; so that in a sense "If* p*,* q*" and "If* $p, \neg q$*" are contradictories. We can say that they are fixing their degree of belief in* q *given* p*. If* p *turns out false, these degrees of belief are rendered void. If either party believes* $\neg p$ *for certain, the question ceases to mean anything to him except as a question about what follows from certain laws or hypotheses.* (Ramsey, 1931, p.247, footnote 1)

The novelty of Ramsey's (1931) argument was the interpretation of probability as degree of belief as well as the drawn connection between the "*degree of belief in q given p*" and "*ordinary, typically uncertain,*

4 Despite Adams (1965) referring to *assertability*, *Adam's Thesis* is often cited to refer to *acceptability* (e.g., see Mellor, 1993; Skovgaard-Olsen et al., 2016). But, as noted by Douven and Verbrugge (2010, p.305), Adams seems to have conceived $P(C \mid A)$ as measuring the acceptability of a conditional $A \to C$ as well.

5 For a proof, see Egré and Rott (2021).

conditional judgments" (MacBride et al., 2020). Theories of conditionals that are based on Ramsey's idea that a conditional is processed by supposing the antecedent based on which the consequent is evaluated are summarized under the name *Suppositional Theory* (Edgington, 2020). The importance of the connection between the conditional probability and natural language conditionals is once more emphasized by Edgington (1995) who called the equation $P(\text{"If A, C"}) = P(C \mid A)$ "*the Equation*". Like Adams (1965) Edgington is a proponent of a non-truth functional account of conditionals, arguing that conditionals do not have truth conditions, and that the conditional probability is the measure of the extent to which a conditional is acceptable (Edgington, 2003, p.388).

Analogously to what has been called "the Equation" among philosophers, psychologists refer to the *conditional probability hypothesis* predicting people's probability judgments of natural language conditionals to be in line with the corresponding conditional probability (Over & Cruz, 2021). The questions about conditionals that people try to find answers for in the psychology of reasoning focus on how they are processed and how people reason with them, rather than on their truth conditions. The material implication had long been considered as the normative standard in the psychology of reasoning with the most influential account based on the material implication being *Mental Model Theory* (see Johnson-Laird, 1986, 2001; Johnson-Laird & Byrne, 1991, 2002). Mental Model theory represents the reasoning process as a simulation of mental models where each model corresponds to a set of possibilities, which of these models will be represented is further due to pragmatic and semantic modulation, taking into account the broader utterance context, respectively the semantic content of antecedent and consequent. To give an example, consider (12) and (13) taken from Johnson-Laird and Byrne (2002, p.661).

(12) If the patient has malaria, then she has a fever.

(13) If the ball rolls to the left, then the red light comes on.

The authors argue that (12) receives a conditional interpretation (according to the material implication) whereas (13) receives a biconditional interpretation (according to $\Leftrightarrow$, the logical 'if and only if'). The interpretation of (12) corresponds to the three models where (i) the patient has malaria and a fever (ii) the patient does not have malaria but a fever and (iii) the patient neither has malaria nor a fever, whereas in the interpretation of (13) only two models are represented explicitly, (i) the ball rolls left and the red light comes on and (ii) the ball rolls right and the green light comes on. As it is the case in the biconditional interpretation of (13), the set of mental models that people explicitly represent, and are thus aware of, might not correspond to the complete list of theoretically possible cases which is how the authors explain the observed divergences between peoples' interpretations of conditionals and the material implication.

The fundamental shift away from a binary truth-conditional paradigm with a focus on the material implication to a non-truth functional paradigm, focusing on conditional probabilities, that allows for non-deductive and uncertain inferences, was dubbed as "*new paradigm psychology of reasoning*" (Elqayam & Over, 2013; Over, 2009).[6]

CONDITIONALS & (CAUSAL) RELATIONS. One of the most characteristic inferences with respect to indicative conditionals is the dependency relation between the antecedent and the consequent. If we were to ask anyone to replace A and C in the sentence "*if A, C*" by simple propositions, most likely the chosen propositions would be related in some way. A naturally forthcoming reaction when being prompted with a conditional whose antecedent and consequent do not directly seem to be related, is to ask for such a relation. This seems to happen in (14),[7] taken from Douven and Romeijn (2011) and discussed in Douven (2012), that will be considered in light of our model in Chapter 5 (Section 5.3).

(14) If it rains tomorrow, we cannot have sundowners at the Westcliff. (Douven & Romeijn, 2011)

In the original example, the conditional is supposedly uttered by a speaker who agreed with the listener to have sundowners at a hotel called the 'Westcliff'. Even though the addressee might not understand why exactly the event of rain would hinder them from having sundowners at the said hotel, it seems reasonable and natural to infer from (14) that the speaker knows that there is some relation.

The conditional in (15) is a similar example: one may be puzzled by this utterance, as it is difficult to imagine what the winning of a soccer team has to do with the vacation of the speaker's babysitter.

(15) ? If soccer team A wins, our babysitter will go on a long vacation.

(16) # If her favorite color is yellow, it will snow tomorrow.

Yet, it seems natural to immediately think about a possible relation; the babysitter might, for instance, have bet on team A and may thus win a lot of money that would allow for a long vacation. In other examples, like (16), a possible relation between antecedent and consequent is so far-fetched that the conditional simply seems to be nonsense; this kind of conditionals is called *missing-link* conditional, emphasizing the importance of the communicated relation once more.

6 Note that recently, Johnson-Laird et al. (2015) proposed a revised version of mental model theory that is not based on the material implication.
7 Like infelicity is marked by a preceding '# ', we use a preceding '?' for examples that are neither clearly infelicitous nor clearly felicitous.

Some people go so far as to argue that the dependency relation is so characteristic for conditionals that it is an inherent part of their meaning (e.g., Douven, 2008; Krzyżanowska et al., 2014; Skovgaard-Olsen et al., 2016; van Rooij & Schulz, 2019) — this kind of semantics for conditionals has been called *Inferentialism*, summarized by Douven et al. (2022) as follows:

> *A conditional "If* A*,* B*" is" true if there is a compelling argument from* A *plus contextually determined background premises to* B*, with* A *being pivotal to that argument (i.e., with* A *removed, the argument would cease to be compelling), false if there is a compelling argument from* A *plus contextually determined background premises to the negation of* B*, and indeterminate otherwise.* (Douven et al., 2022, p.7f)

One of the first hypotheses proposed in this vein is the so-called *Evidential Support Thesis* (EST) from Douven (2008) which predicts a simple conditional $A \to C$ to be acceptable if and only if $P(C \mid A) > \theta$ and $P(C \mid A) > P(C)$, inducing a positive relevance on the antecedent A with respect to the consequent C. Building on EST, van Rooij and Schulz (2019) consider as assertability condition for conditionals a measure known as *relative difference* (Sheps, 1958), which is defined in terms of *contingency* (Equation [1]).

$$\triangle P_A^C := P(C \mid A) - P(C \mid \neg A) \qquad [1]$$

Positive contingency values ($\triangle P_A^C$) mean that both, the presumable cause (A) and effect (C), tend to co-occur, whereas negative contingency values mean that when the cause is present, the effect tends to be absent, and contingency values of 0 indicate that there is no relation between both observations. In causal terms, $\triangle P_A^C > 0$ denotes a *generative* and $\triangle P_A^C < 0$ a *preventive* cause. For a conditional $A \to C$ to be assertable, van Rooij and Schulz (2019) argue that $\triangle^* P_A^C$ (Equation [2]) must take on high values.

$$\triangle^* P_A^C := \frac{P(C \mid A) - P(C \mid \neg A)}{1 - P(C \mid \neg A)} = \frac{\triangle P_A^C}{1 - P(C \mid \neg A)} \qquad [2]$$

They show that besides accounting for the dependency between antecedent and consequent, this single measure can account for intuitions that Douven's (2008) proposed assertability condition cannot account for, including that $P(C \mid A)$ should count for more than $P(C \mid \neg A)$ does.[8] In Chapter 4 (Section 4.4), we will consider van

8 That $P(C \mid A)$ should count for more than $P(C \mid \neg A)$, means that when the latter is fixed, a change of the former (adding/subtracting a value δ) should lead to a greater change of the considered measure $\triangle^* P_A^C$ as compared to when the former is fixed and the latter changes by the same δ. In other words, $P(C \mid A)$ should have more influence on $\triangle^* P_A^C$ than $P(C \mid \neg A)$.

Rooij and Schulz's (2019) proposal from the point of view of our pragmatic model for the communication with conditionals. More precisely, we will show that, in what we will call default contexts, listeners interpreting a conditional $A \rightarrow C$ infer that $\triangle^* P_A^C$ is indeed large. However, we will also see that large values of $\triangle^* P_A^C$ are not a sufficient condition for our speaker to utter $A \rightarrow C$.

Focusing on the relevance relation between antecedent and consequent, Skovgaard-Olsen et al. (2016) observed a diminished influence of the conditional probability in participants' acceptance ratings of conditionals when the antecedent and the consequent are negatively, or not at all related. To account for this empirical observation, the authors propose what they called the *Default and Penalty Hypothesis*, a heuristic according to which in the default case — when A is positively relevant for C — participants draw on the conditional probability $P(C \mid A)$ to evaluate the acceptability or truth of the conditional $A \rightarrow C$. When the positive relevance assumption is violated, the penalty-part is activated which manifests itself, so they argue, in participants' lower sensitivity with respect to $P(C \mid A)$.

Another recent proposal in this vein is the *Hypothetical Inferential Theory* (short HIT), a psychological dual-processing hypothesis put forward by Douven et al. (2018). The theory combines the intuitions from Inferentialism (Krzyżanowska et al., 2014), and the suppositional conditional (Douven et al., 2018, p.54). HIT assumes that in the default case, the mental representation of a conditional is an inferential link from the antecedent to the consequent which exists as long as it is strong enough to be subjectively supported, determined by pragmatic cues, inferences to the best explanation, etc. (intuitive, effortless process). When the cues for the existence or strength of a link between antecedent and consequent are not sufficient so that it is not subjectively supported, HIT predicts a truth value gap (Douven et al., 2018, p.54) accounting for the defective truth table (effortful process).

An argument that is often cited as evidence for Inferentialism — that is, for the inferred dependency relation between antecedent and consequent being part of the meaning of conditionals — is the infelicity of missing-link conditionals like (9) or (16) with no reasonable conceivable relation between antecedent and consequent (e.g., Krzyżanowska & Douven, 2018; Vidal & Baratgin, 2017). However, there is another subset of conditionals without a (dependency) relation between antecedent and consequent that poses a major challenge for Inferentialists since they are — unlike missing-link conditionals — felicitous. Two examples that belong to this subset of conditionals are so-called *biscuit conditionals* and *concessive* conditionals; see (17) and (18) for an example of each kind.[9]

9 Recently, van Rooij and Schulz (2020, 2021) proposed a generalization of their Inferentialist assertability/acceptability conditions (van Rooij & Schulz, 2019) that is also

(17) If you're hungry, there are biscuits on the sideboard. (Austin, 1956)

(18) Even if it rains, they will go hiking.

They communicate the truth of the consequent *independently* of the antecedent. While with (17) the speaker does not only communicate that there are biscuits on the sideboard, this conditional may further be interpreted as an offer for the addressee to take some, and (18) seems to put emphasis on the independence of 'raining' and 'them going hiking' besides communicating a strong belief in the consequent ('them going hiking').

Whether the commonly inferred relation between antecedent and consequent is part of the semantic meaning of conditionals or rather due to pragmatic enrichment is strongly debated. We will come to arguments for the latter option — which is also what we argue for — later in the next section, in which we turn to the pragmatics of conditionals, starting with a short introduction of the pragmatic concepts relevant for the work presented in this thesis.

2.1.2 *Pragmatics of conditionals*

While semanticists consider the meaning of words and sentences in isolation, mostly in terms of their truth conditions, pragmatists investigate what the speaker *communicates* or, to say it in the words of Grice (1975), *implicates* with the spoken words. From the perspective of the listener, pragmatics investigates how the listener's interpretation of the speaker's utterance is affected by the speaker's choice of words and other contextual assumptions, in short the so-called *utterance context*. This is a broad expression that may include anything from the prior knowledge of the speaker relevant in the considered situation, the speaker's intentions, the relation of speaker and listener, the setting in which the utterance is made (e.g. in a formal vs. an informal context) or the set of relevant alternative utterances that the speaker could have said but didn't.

In the following, I will first shortly introduce in Section 2.1.2.1 the pragmatic concepts that will be relevant for the discussion of conditionals from a pragmatic point of view. In Section 2.1.2.2 I proceed with some background on relevant issues concerning the pragmatics of conditionals, in particular the commonly inferred relation between antecedent and consequent and conditional perfection-readings. Then, I will turn to examples that showcase the importance of taking into account world knowledge as well as the interlocutors' epistemic states — two integral parts of our model for the communication with con-

able to account for Biscuit conditionals, more precisely for the inferred truth of the consequent.

ditionals — in order to infer the meaning that the speaker communicates with a conditional.

2.1.2.1 *Gricean Pragmatics*

In order to explain the *implicated* meaning of an utterance that goes beyond its *literal* meaning, Grice (1975) defined a theory of *implicatures*, differentiating between so-called *conventional* and *conversational* implicatures. A conventional implicature is an utterance whose communicated meaning is due to the conventional meaning of the chosen words; an example (adapted from Grice and White, 1961, p.127) is given in (19). While the truth conditions for (19) and (20) are the same, only the connective 'but' in (19) communicates an expected contrast between *being rich* and *being honest*. If the communicated meaning of an utterance, for instance the contrast communicated by 'but', is *detachable* from the utterance (here 'but') in the sense that there is another expression with the same truth conditions that does *not* carry the same implicated meaning (here 'and'), the *implication* (e.g., the contrast in case of 'but') is considered to arise from a *conventional* implicature (see Potts, 2005).[10]

(19) She is rich but honest.

(20) She is rich and honest.

Conversational implicatures, on the other hand, are explained by Grice to result from conversational rules that interlocutors normally adhere to, which he summarized as the *Cooperative principle*:

> COOPERATIVE PRINCIPLE: *Make your conversational contribution such as is required, at the stage at which it occurs, by accepted purpose or direction of the talk exchange in which you are engaged.* (Grice, 1991, p.26)

Grice formulated a set of maxims, nowadays called *Gricean Maxims*, which require the speaker to be as informative as necessary but not more informative than necessary (Maxim of *Quantity*), to only say what one believes to be true and does not lack adequate evidence (Maxim of *Quality*), to only say what is relevant (Maxim of *Relation*) and to say that on point (Maxim of *Manner*). Each of these maxims further comprises several submaxims; 'Avoid ambiguity' and 'Be brief' are, for instance, two submaxims of the Maxim of Manner. Horn (1984) later reduced Grice's (1975) maxims and submaxims to what he called Q- and R-implicatures where the former comprises implicatures that are due to the alternative utterances that the speaker did not choose to say (e.g., 'some' $\rightsquigarrow$ 'some, but not all') while the latter is

10 In examples, I will use $\rightsquigarrow$ to denote implications; so $x \rightsquigarrow y$ means 'x implicates y'.

motivated by non-linguistic factors as well, including for instance indirect speech-acts (see Horn, 2000, p.309f.). Similarly, Levinson (2000) summarized the Gricean Maxims and submaxims into three kinds of implicatures, Q-,I- and M-Implicatures. Q-implicaturess were defined as inferences that result from a reference to an alternative utterance that the speaker did *not* choose, without any non-linguistic knowledge being involved whereas I-implicatures were defined as positive strengthenings toward stereotypical situations that result without reference to other alternative utterances (e.g., 'Harry and Sue bought a piano' ⇝ 'They bought one piano together, not one each' (Levinson, 2000, p.117)). The third kind of implicature, M-Implicatures, result from using marked expressions which implicate special, non-typical situations (e.g., 'The blue cuboid block is supported by the red cube' ⇝ 'The blue block is not, strictly, a cube' (Atlas & Levinson, 1981, p.31)).

Consider (21) and (22) for two examples of Gricean conversational implicatures (or Q-implicatures according to Levinson and Horn). (21) is the prime example for so-called *scalar implicatures* where the salient alternative utterances that the speaker might have chosen can be ordered on a scale: ⟨none (0), some (1,. . . , n-1), all (n)⟩.

(21) She ate some cookies.
⇝ She did *not* eat *all* cookies.

(22) A: Where does C live? B: Somewhere in the south of France. (Grice, 1991, p.32)
⇝ B *does not know* where *exactly* C lives (given that this information is relevant in the utterance context).

Given that the speaker respects the Cooperative Principle, in particular the Maxim of Quantity, and thus does not withheld information relevant in the utterance context, it can be concluded from (21) that the speaker is not in a position to say something more informative like "*She ate all cookies*", leading to the interpretation of '*some*' as '*some but not all*'. The inference in (22) is similar to the implicature in (21), but without a clearly defined scale: B is expected to be as informative as possible (and required), and thus the unspecific response *somewhere in the south of France* implicates that the speaker does not know C's exact place of residence. (21) and (22) are both *generalized conversational implicatures* since they occur independently of a concrete utterance context.[11]

Similarly to conventional implicatures being detachable (see explanation above), there are characteristics that are associated with conversational implicatures that can be tested for to find out whether the implicated meaning of an utterance can be explained by a conversa-

11 The counterpart of generalized conversational implicatures, defined by Grice (1975), are *particularized* conversational implicatures which do not normally arise, but are bound to specific features of the utterance context (e.g.,see Horn, 2006, p.4).

tional implicature. One of these characteristics is *Cancellability*. Upon saying (21), it is, for instance, no problem to continue saying "...in fact, she ate *all* cookies", whereby the conversational implicature from 'some' to 'some but not all' is canceled. Contrary to that, the (conventional) implication that 'being rich' and 'being honest' are contrasting, communicated by (19), does not seem to be cancellable — "She is rich but honest, however, I do not want to imply that there is a contrast between being rich and being honest" sounds odd.

In a nutshell, Grice's (1991) ideas are based on considering communication as something rational, assuming that speakers choose the words that they utter for a reason — a *good* reason.

2.1.2.2 *Gricean Pragmatics and conditionals*

In line with Jackson (1979) and Stalnaker (1976), Grice (1989) argued that the truth values of the natural language conditional are those of the material implication and that the apparent differences in the interpretation of natural language conditionals and the material implication can be explained by pragmatic considerations (Grice, 1989). This difference was dubbed by Grice as *Indirectness Condition* representing the 'non-truth functional grounds for accepting $A \supset C$' (Grice, 1991, Ch.4, p.58). These include the relation between antecedent and consequent which others ascribe to the semantics of conditionals as we have seen in Section 2.1.1. We turn to pragmatic explanations of this relation next.

inferred relation between A & C. To explain the commonly inferred relation between A and C, which is not accounted for by the material implication $A \supset C$ as underlying semantic of natural language conditionals, Grice (1989) considers the possibility that it is due to a conventional implicature. Yet, he refutes this possibility since the inferred relation is non-detachable: reformulations of the conditional ("If A, C") in terms of logical equivalent expressions to the material implication (e.g., $\neg A \vee C$, $\neg(A \wedge \neg C)$) still carry the implication of a relation between A and C (e.g., "If Smith is in London, he attends the meeting" vs. "Either Smith is not in London or he attends the meeting" Grice, 1991, p.67).

So, Grice argues that the inferred relation results from a conversational implicature; that is, a speaker who utters $A \rightarrow C$ is *conversationally* committed to the material implication $A \supset C$ *and* the Indirectness condition.[12] More precisely, Grice contends that the relation between antecedent and consequent is implicated by the speaker's utterance of the conditional because of the speaker's commitment to the conversational maxims, in particular the Maxim of Quantity: the conditional

12 For completeness, note that Grice explored another possibility for the origin of the inferred relation, considering the supposed role or function in the language of a conditional particle (see Grice, 1991, Ch.4).

$A \rightarrow C$ is logically weaker, and thus less informative than other utterances as, for instance, the direct claim of the consequent (C). The latter should yet be considered relevant to the conversation, given that the conditional $A \rightarrow C$ is relevant.

As mentioned above, it is strongly debated whether the commonly inferred link between antecedent and consequent is pragmatically implicated or part of the semantics of conditionals. Proponents of the semantic approach (e.g., see Douven, 2017; Krzyżanowska & Douven, 2018) have criticized those who argue for a pragmatic account for not being very explicit in the description of the exact the assumed pragmatic processes. In Chapter 3 we present a computational model that makes the pragmatic processes explicit and which we argue is able to explain the commonly inferred link without the need to ascribe it to the semantic, the core meaning of the conditional.

Here, I would like to shortly introduce another pragmatic explanation of the inferred link that has recently been spelled out by van Rooij and Schulz (2022). As described above in Section 2.1.1, van Rooij and Schulz proposed the so-called relative difference ($\triangle^* P^C_A$, see Equation [2]) as acceptability/assertability condition for conditionals. More precisely, they argued that $\triangle^* P^C_A$ should be reasonably large (≈ 1) for the conditional $A \rightarrow C$ to be acceptable/assertable. In a recent paper (van Rooij & Schulz, 2022), they now derived — in line with our results in Chapter 4 (Section 4.4) — that $\triangle^* P^C_A \approx 1$ results from the assumption that $P(C \mid A) \approx 1$ for $A \rightarrow C$ to be assertable and the conversational implicature that the speaker is not in a position to claim the consequent directly, which would be more informative than the selected conditional. From these assumptions, $P(C \mid A) \approx 1$ and $P(C) << 1$, it follows that $P(C \mid A) - P(C) >> 0$, and thus $P(C \mid A) >> P(C)$. This further implies (i) $P(C \mid A) >> P(C \mid \neg A)$ because $P(C \mid A) >> P(C)$ holds if and only if (i) holds.[13] Taken together, this implies that their proposed measure of relevance, $\triangle^* P^C_A$, takes on a large value close to 1. Thus, like we independently derive in Chapter 4, this shows that it is not necessary to explicitly require large values for $\triangle^* P^C_A$ for the conditional $A \rightarrow C$ to be assertable: the implicature that the speaker does not know the truth value of C when

13 (1) $P(C) = P(C \mid A) \cdot P(A) + P(C \mid \neg A) \cdot P(\neg A)$
1. Assume (a_1) $P(C \mid A) >> P(C \mid \neg A)$.
$\overset{(1)+(a_1)}{\Rightarrow} P(C) << P(C \mid A) \cdot P(A) + \mathbf{P(C \mid A)} \cdot P(\neg A)$
$\Rightarrow P(C) << P(C \mid A) \cdot (P(A) + P(\neg A))$
$\Rightarrow P(C) << P(C \mid A)$
2. Assume (a_2) $P(C \mid A) >> P(C)$.
$P(C \mid \neg A) \overset{(1)}{=} (P(C) - P(C|A) \cdot P(A)) / P(\neg A)$
$\overset{(a_2)}{\Rightarrow} P(C \mid \neg A) << (P(C) - \mathbf{P(C)} \cdot P(A)) / P(\neg A)$
$\Rightarrow P(C \mid \neg A) << P(C)$
$\overset{(a_2)}{\Rightarrow} P(C \mid A) >> P(C \mid \neg A)$

asserting $A \rightarrow C$, together with the assumption that $P(C \mid A) \approx 1$ accounts for $\triangle^* P_A^C \approx 1$.

CONDITIONAL PERFECTION. A phenomenon in the context of conditionals that has particularly been discussed on pragmatic grounds, more precisely as a conversational implicature (e.g., Atlas & Levinson, 1981; Horn, 2000; van der Auwera, 1997) is the biconditional interpretation of "*if*" as "*if and only if*" (iff) referred to as *conditional perfection* (CP) (Geis & Zwicky, 1971) — which Chapter 8 of this thesis is concerned with.

For some conditionals, a 'perfected' interpretation seems to be endorsed by default, without the specification of a concrete utterance context; an eminent example are promises and threats communicated with conditionals like (23) below. Here, the biconditional interpretation is forthcoming naturally: the speaker seems to communicate to scratch the addressee's back if and only if the addressee scratches the speaker's back.

(23) If you scratch my back, I'll scratch yours.

(24) If you drive much too fast in town, you'll lose your driver license.

Other conditionals, like (24), do not seem to receive CP-reading by default; one may, for instance, not drive much too fast in town , but lose one's driver license for drunk driving.

It is not only unclear what exactly triggers conditional perfection readings, the difficulties already start with the definition of the phenomenon itself, which is surprisingly vague. Most papers about CP consider $\neg A \rightarrow \neg C$ as additional inference, fewer focus on 'only if A, C' (e.g., see Herburger, 2016) or $C \rightarrow A$ to lead to the biconditional interpretation (see Van Canegem-Ardijns & Van Belle, 2008, p.350). Note that logically, it does not matter with which of these inferences the conditional $A \rightarrow C$ is enriched; all result in the biconditional $A \iff C$.

Van Canegem-Ardijns and Van Belle (2008) considered various different inferences all related to CP in more detail and found that they are connected to different kinds of conditionals. The most general inference that they considered, $\neg A \rightarrow \neg C$, was found to be related with a number of different speech acts, whereas 'only if A, C' and 'only if $\neg A, \neg C$' were related to specific kinds of conditionals only. The former was found to be related to conditional promises or permissions and the latter to conditional threats or recommendations.

In the linguistic literature, CP-readings of conditionals have often been explained by pragmatic effects (e.g., Atlas & Levinson, 1981; Horn, 2000; van der Auwera, 1997), differing in the exact mechanisms at play that are assumed to generate the implicature.

Atlas and Levinson (1981), for instance, ascribe the biconditional interpretation of a conditional to an I-implicature, that is, as an inference to a stereotypical situation. According to van der Auwera (1997), CP arises from a scalar Q-implicature, assuming the conjunction of several conditionals ($(A \rightarrow C) \wedge (B \rightarrow C) \wedge \ldots)$ as alternative, more informative utterance than the conditional $A \rightarrow C$, which has been criticized on several grounds. If the more informative alternative conjunction of conditionals consists of more than one conjunct other than the uttered conditional (e.g., $(A \rightarrow C) \wedge (B \rightarrow C) \wedge (D \rightarrow C)$), one of the alternative conditionals can still be true, which means that A is *not* a necessary condition for C (also see Franke, 2009, p.238). Further, it has been argued that for a scalar Q-implicature, the considered alternative utterances must be equally brief and equally lexicalized (Levinson, 1987), or the alternatives on top of the scale must be at least as lexicalized as the utterance at the bottom of the scale (Horn, 2000). Otherwise it cannot be excluded that the speaker chose the less complex utterance, not because she was not in a position to utter the more specific, and thus more informative utterance, but simply made this choice for economic reasons. The alternative utterances in van der Auwera's (1997) account, however, have different complexities; $A \rightarrow C \wedge B \rightarrow C$ is more complex than $A \rightarrow C$.

Horn (2000) prefers a different alternative utterance, namely the mere claim of the consequent (C) and argues that CP-readings are due to an R-implicature, according to his distinction between Q- and R-implicatures (Horn, 1984) where R-implicatures are speaker-based and result from the speaker being required to make the contribution a necessary one, corresponding to the Maxims of Quantity, Relation and Manner in terms of Grice (see Horn, 2000, p.310).

Other pragmatic aspects have been considered to influence the occurrence of CP. Participants' interpretation of a conditional as biconditional was, for instance, shown to be influenced by the availability of alternative causes or disabling conditions for the consequent, making a biconditional interpretation more likely when fewer alternative causes are conceivable (Cummins et al., 1991; Markovits, 1986).

The account proposed by von Fintel (2001) that we experimentally test in Chapter 8 is also related to the set of possible alternative causes for the consequent, but it focuses on the question-under-discussion (QUD) that specifies the purpose of the conversation: von Fintel predicts the interpretation of a conditional $A \rightarrow C$ to depend on whether the conversation focuses either on the conditions that make the consequent true or on the consequences that follow from the truth of the antecedent, determined by the QUD. The former case may, for instance, result from the QUD being something like "*How to achieve C? In which cases C?*", whereas for the latter, the QUD may be "*What happens if / follows from A?*". Depending on the QUD, and to that effect, depending on the focus of the conversation, the speaker's set of the relevant alternative utterances will differ and the conditional is predicted to either receive a biconditional (when QUD="*How to achieve C?*") or a simple conditional reading (when QUD="*What if A?*").

CONTEXT DEPENDENCY & WORLD KNOWLEDGE. In much of the early research on conditionals, in particular in the psychology of conditional reasoning, conditionals were mostly treated as abstract rules. However, in everyday conversations, we have already seen that different conditionals imply fundamentally different things, which suggests that their interpretation is highly context-dependent; they may be used to make an offer like in so-called biscuit conditionals (e.g., "If you're hungry, there is pizza left."), they can communicate promises or threats (e.g., "If you don't give me the money, I will kill you", "If you win, I'll bake a cake for you") or they can be used to describe consequences of possible actions (e.g., "If you throw an even number, you will win (the game).") to name a few possibilities.

When taking into account background knowledge about the propositions expressed in the conditional and the relation among them, different kinds of conditionals may, however, appear more similar than different in the sense that they can be treated in a uniform way. Promises and threats, for instance, both communicate the condition that would bring about the consequent, for threats this is something to be avoided since it is common knowledge that people generally dislike it (e.g., getting killed), while for promises it is something positive that people generally like (e.g., cake, especially if someone else bakes it). Another example are biscuit conditionals: it is uncontroversial to assume as given world knowledge that the antecedent (e.g., someone being hungry) has no influence on the truth of the consequent (e.g., that there is pizza left). This explains the common inference that the speaker has a strong belief in the consequent — assuming that $A \rightarrow C$ is only assertable when $P(C \mid A) \geq \theta$ for a reasonable large threshold θ, as we would assume for any non-special conditional. If A and C are independent, $P(C) \geq \theta$ follows from $P(C \mid A) \geq \theta$ since in that case $P(C \mid A) = P(C)$. Yet note that there is more to biscuit conditionals than that: the assumption of the background knowledge that antecedent and consequent are independent alone does not address the question *why* the speaker chose to utter the biscuit conditional instead of claiming the consequent straight away.

A famous example from Gibbard (1981), that I would like to introduce in a bit more detail, nicely showcases the importance of contextual knowledge on the interpretation of conditionals. Gibbard described a situation with two people playing poker, Pete and Mr. Stone, and two trustworthy henchmen, Jack and Zack. Jack sees both players' cards, Zack only sees Mr. Stone's cards, but shares this information with the other player, Pete. In this situation, Jack (seeing both players' cards) truthfully says "*If Pete called, he lost*" and at the same

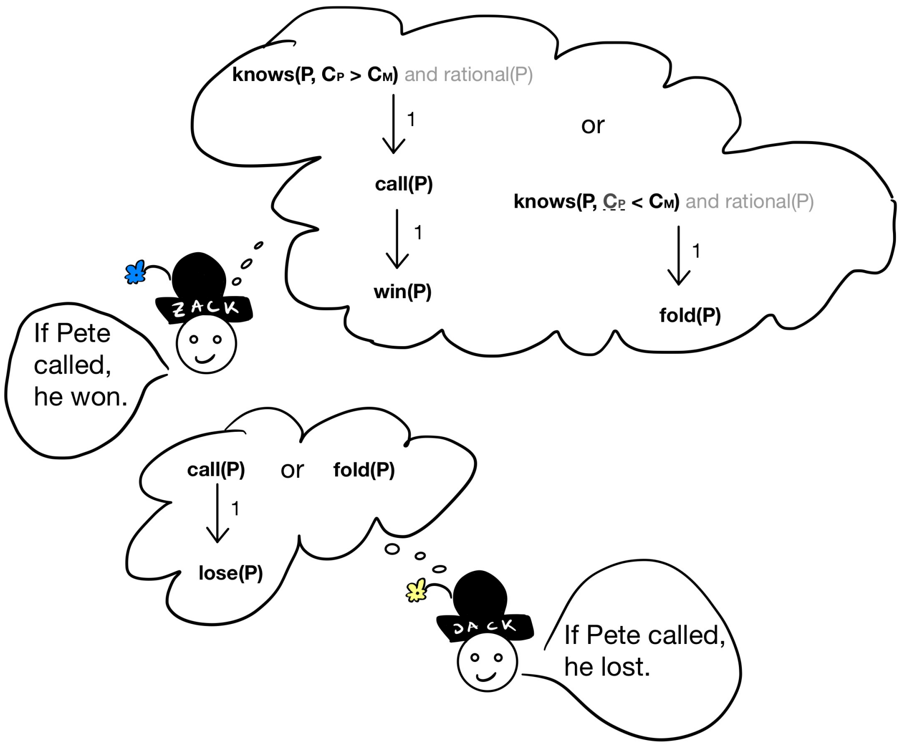


Figure 1: Relevant background knowledge in the Poker example from Gibbard (1981); knows(x,y) means that x knows y, C_P and C_M denote Pete's, respectively Mr. Stone's cards and $C_P > (<) C_M$ means that Pete's cards are better (worse) than Mr. Stones' cards.

time, Jack (seeing Mr. Stone's cards and telling Pete) truthfully says "If Pete called, he won", which are contradictory but intuitively both assertable by the respective henchman.

Gibbard used this example to argue against conditionals being propositions having truth conditions: assuming that conditionals do have truth conditions, the conditionals asserted by Zack and Jack should reasonably both evaluate as true. This, however, contradicts the *principle of Conditional Non-Contradiction* (CNC) which says that conditionals with the same antecedent and contradictory consequents (e.g., $A \rightarrow C, A \rightarrow \neg C$) cannot both be true, unless the antecedent is inconsistent. Contrary to Gibbard, Krzyżanowska et al. (2014) argued that it is possible to define a truth-conditional semantics for conditionals that evaluates both conditionals in the example as true while still respecting CNC. To achieve this, Krzyżanowska et al. defined a set of conditions that all involve the speaker's background knowledge.

Independently of the question about truth conditions, considering the relevant background knowledge, shown in Figure 1, helps to see that both speakers made a perfectly reasonable choice in saying the respective conditional although they are presumably contradictory. To be able to make sense of both conditionals, it is indispensable to consider the *different* utterance contexts of the two speakers. While neither Jack nor Zack know whether Pete called, their utterances are based on two fundamentally different facts: Zack knows that Pete knows whether or not he will win since thanks to Zack, Pete knows Mr. Stone's cards, and thus, given that Pete is rational — which is silently assumed, therefore it is grayed out in Figure 1 — he will only call in case that his cards are better than Mr. Stones' and so, in that case, he will win. On the other hand, Jack knows for sure who will win because he sees both players' cards, but no information is shared with any of the two players. Thus, from the point of view of Jack, Pete cannot know whether he will win in case he calls since he does not have any information about Mr. Stones' cards. So, Jack's utterance is based on complete knowledge about Pete and Mr. Stones' cards whereas Zack's utterance is based on his knowledge about (i) Mr. Stone's cards (ii) Pete's epistemic state with respect to Mr. Stone's cards *and* (iii) the assumption that Pete is rational. If Pete wasn't rational, he might call even though he knew that he would lose — which is clearly absurd — and therefore Zack would not be in a position to utter the asserted conditional.

The contextual circumstances of both Zack's and Jack's situation legitimate the respectively uttered conditional and the principle of Conditional Non-Contradiction remains valid given that it is interpreted such that two conditionals $A \to C$ and $A \to \neg C$ cannot both be true (or assertable) given *identical* utterance contexts. Similarly, we can say that "the grass is wet" in one situation and that "the grass is dry" in another situation without any problems. The peculiarity in Gibbard's (1981) example is that the *objective* situation is indeed the same, both conditionals refer to one and the same situation, assuming that we do not know anything about the epistemic states of the henchmen. Though, this is what is crucial here, the additional knowledge that the two speakers respectively have. Yet, even if we, as the addressee of Zack and Jacks' respective utterance, do not know anything about their epistemic states, we would not end up believing $A \to C$ and $A \to \neg C$ at the same time: either we withdraw the assumption that the henchmen are liable or in case that this is unassailable, we should infer that they must have different pieces of information on which their claims are based. To check whether this holds, we might ask each of them whether Pete calling is really the only condition for the mentioned conclusion, that is, that Pete wins (Zack) or looses (Jack). While Jack would affirm this, Zack should not: he should respond by saying something like "... well, if he is *not* rational he might also call and lose".

This brings us to the issue considered in the next section, namely the question how one's beliefs change upon receiving information in form of conditionals.

2.1.3 *Belief update with conditionals*

As illustrated by the Poker-Example discussed in the previous section, the interpretation of utterances comes along with an update of the interpreter's beliefs. That is, the meaning of an utterance is strongly related to how listeners update their beliefs in response to a speaker's utterance.

Someone who is, for example, told by a cooperative and trustworthy speaker that 'it is snowing outside' ($S = s$) is assumed to update her prior belief $Pr(S = s) \in (0, 1]$ to a posterior value of 1. While belief update with factual knowledge like in this simple example can be straightforwardly formalized with the well known *principle of conditionalization*, using conditional probabilities (Lin, 2022), it is less clear how belief update works with conditionals. To see why, let us shortly explain conditionalization. Conditional probabilities describe how one's belief, represented by a probability distribution $Pr(\cdot)$, in a certain event changes when another event is observed. That is, when my belief in a certain hypothesis is described by $Pr(H_1)$ prior to my observation of the truth or falsity of an event E, the conditional probability $Pr(H_1 \mid E = 1) = \frac{Pr(H_1, E=1)}{Pr(E=1)}$ describes my updated belief, P^*, in H_1, $P^*(H_1) = Pr(H_1 \mid E = 1)$. Here, the information which is conditioned on is — and necessarily so — a proposition that is either true or false. For conditionals, however, no consensus has even been reached about whether or not they have truth conditions at all, and thus, there is no straight forward way like basic conditionalization to describe the update of one's beliefs given conditional information.

A more general question than how to condition one's belief on a conditional, is the question how to condition one's beliefs on *uncertain* information which can be described by a generalization of basic conditionalization called *Jeffrey conditionalization* (Jeffrey, 1983). It is applicable in case that there is *uncertain* evidence; one might, for instance, not know for sure whether an event occurs or not, but have evidence for it to occur with a certain probability (e.g., $P(H_1) = 0.7$). This method is defined by *Jeffrey's rule*, shown in Equation [3]; the updated belief in H_1 is defined as the weighted sum of one's old belief in H_1 given that E_1 is true, respectively false, weighted by one's respectively updated belief in E_1.

$$P^*(H_1) = P(H_1 \mid E = 1) \cdot P^*(E_1 = 1) + P(H_1 \mid E = 0) \cdot P^*(E = 0) \quad [3]$$

For a perfectly reliable information source, that is when $P^*(E_1 = 1) = 1$ or $P^*(E_1 = 0) = 1$, [3] is identical to basic conditionalization with $P^*(H_1) = Pr(H_1 \mid E) \cdot Pr(E)$. Another possibility to tackle belief update with uncertain information is to look for a probability distribution constraint to the received uncertain information that is minimally different from the distribution representing the addressee's prior beliefs (see Grove & Halpern, 1997, p.208). However, this method, as well as Jeffrey conditionalization have been shown to yield undesired results (Trpin, 2020) and results that are incompatible with common intuitions when applied to conditionals (e.g., see van Fraassen (1981), but see Douven and Romeijn (2011), Grove and Halpern (1997) for counterarguments).

We will consider belief update with conditionals in light of our model in Chapter 5 (Section 5.1–5.3), in which we discuss three exemplary conditionals ((25) and (26) below and (14) from above), embedded in specific utterance contexts that were first discussed by Douven (2012).[14] The challenge that these examples bring along is that each of the three utterance contexts has different effects on the addressee's posterior belief in the antecedent: after learning the conditional, it may either remain unchanged, increase or decrease.

(25) If Sue passed the exam, her father will take her on a skiing vacation. (Douven, 2012)

(26) If Kevin passed the driving test, his parents will throw a garden party. (Douven, 2012)

The observation concerning (25) is that a listener who just saw Sue buying skiing clothes and then gets to know (25) would increase her belief in Sue having passed the exam whereas an addressee of (26) who saw Kevins' parents spade their garden, would decrease her belief in Kevin having passed the driving test. The mostly qualitative model proposed by Douven (2012) draws on the explanatory value of the antecedent, with different update mechanisms depending on whether there is a change in the explanatory status of the antecedent or not. As will be seen in Section 5.1 and 5.2, representing the (silently assumed) background knowledge that one usually wears skiing clothes when skiing and that spading the garden is incompatible with throwing a garden party, which can reasonably be taken for granted, helps to treat (25) and (26) — as well as (14) mentioned above — in a uniform way even though the intuitive inferences are orthogonal to each other. Pragmatic reasoning further allows to account for the intuitive interpretations of these examples. Approaches similar to ours have been advanced recently to tackle these benchmark examples from Douven (2012), yet without considering the role of pragmatic reasoning (Eva et al., 2020; Günther, 2018; Vandenburgh, 2021).

14 To be precise, we will consider the effect of the speaker's utterance of a conditional on the listener's beliefs assuming that the listener takes over the speaker's beliefs completely. That is, we will *not* consider how the listener's beliefs represented by a probability distribution P_L are integrated with the speaker's beliefs represented by another probability distribution, P_S.

2.2 EXPERIMENTAL DATA

This section shall provide some background about the empirical findings on the use and interpretation of conditionals. I will start with experiments of how people reason with conditionals, and continue with experiments that aim to test semantic and pragmatic accounts of the meaning of conditionals.

2.2.1 *Conditional Reasoning*

CONDITIONAL INFERENCE TASKS. A large part of the experiments that have been conducted in the context of conditionals are credited to psychologists investigating how people reason with conditionals. The early experiments in particular investigated whether participants adhere to logically correct inferences. Participants were, for instance, confronted with a conditional like (27) and one of the premises given in (27a)-(27c) to test whether they endorse the four classical conditional inferences, spelled out in Table 2.

(27) If we continue to blow CO_2 into the atmosphere, the climate on Earth will change dramatically. [$A \rightarrow C$]

a. We continue to blow CO_2 into the atmosphere. [A]
b. The climate on Earth will not change dramatically. [$\neg C$]
c. We do not continue to blow CO_2 into the atmosphere. [$\neg A$]
d. The climate on Earth will change dramatically. [C]

Accepting (27a) and (27b) given the conditional in (27) corresponds to the logically valid inferences *Modus Ponens* (MP), respectively *Modus Tollens* (MT). Contrary to that, accepting (27c) and (27d) corresponds to the inferences *Denying the antecedent* (DA) and *Affirming the consequent* (AC) which are logically *in*valid.

Throughout studies, MP is almost universally endorsed by all participants. Usually the the second highest endorsement rates are observed for Modus Tollens (MT), which is, like MP, logically valid. The two invalid inferences, AC and DA, typically yield lower, albeit still moderate, endorsement rates. Evans et al. (1993, Chapter 2) collected the results from several experiments (n=14) and summarized the inference rates observed across studies to lie between 0.89 and 1 (MP), 0.41 and 0.81 (MT), 0.23 and 0.75 (AC) and 0.21 and 0.73 (DA).

WASON SELECTION TASK. One of the best known and largely applied tasks in the literature on conditional reasoning is the so-called *Wason-Selection Task* (Wason, 1968) in which participants are given a rule in form of a conditional like (28) and are then asked to verify it by turning around as few cards as possible among a set of four shown cards.

premise 1	premise 2	conclusion	inference name
$A \to C$.	A.	C.	Modus Ponens (MP)
$A \to C$.	$\neg C$.	$\neg A$.	Modus Tollens (MT)
$A \to C$.	C.	A.	Affirming the Consequent (AC)
$A \to C$.	$\neg A$.	$\neg C$.	Denying the Antecedent (DA)

Table 2: Classical conditional inferences. MP and MT are logically valid, whereas DA and AC are logically *in*valid inferences.

(28) If there is a vowel on one side of the card,
then there is an even number on the other side (Wason, 1966).
$[A \to C]$

The four cards show a single side respectively with, in this example, a vowel (A), a consonant ($\neg A$), an even (C) or an odd ($\neg C$) number, corresponding to the premises of the four conditional inferences (MP, DA, AC, MT). According to the material implication, it is necessary to turn around the card with a vowel (A) and the card with an odd number ($\neg C$), corresponding to the premises of the valid inferences MP and MT. The upshot of the studies from Wason (1966, 1968) is that nearly all participants correctly turn over the A-card, but only a minority selects the $\neg C$-card. Many studies investigating peoples' reasoning in the Wason-selection task and variants of it followed, making it "*the most intensively researched single problem in the history of the psychology of reasoning*" in the upcoming quarter of a century (Evans et al., 1993, Chapter 4, p.99). These studies showed that the rate of participants who give logically correct responses depends on the content of the conditional rule to be verified; when the conditional involved familiar content like in (29) instead of abstract rules, the endorsement rates increased substantially (e.g., Griggs & Cox, 1982; Johnson-Laird et al., 1972; Wason & Shapiro, 1971).

(29) If a person is drinking beer, then the person must be over 19 years of age.[15] (Griggs & Cox, 1982)

Several studies showed that for a better performance, that is, a higher endorsement rate of MT, reflected in the choice of the $\neg q$-card, and lower endorsement rates of DA and AC, participants, however, had to be familiar with the rule; familiar content alone was not sufficient (e.g., Griggs & Cox, 1982; Manktelow & Evans, 1979). What is clearly emphasized by these experiments, is a strong contextual component in how people interpret and reason with conditionals.

RELATION BETWEEN $A \to C$ AND $P(C \mid A)$. l in the psychology of reasoning, the shift from the material conditional and mental models towards suppositional theories of conditionals that focus on the

15 At the time of the experiment the law in Florida — the experiment took place at the University of Florida — allowed drinking beer at the age of 19.

conditional probability came along with a shift from experiments focusing on the Wason Selection Task towards experiments investigating the relation between participants' beliefs in a conditional and the corresponding conditional probability. The resulting large body of experimental research on the subject has accumulated evidence for a positive relation between peoples' judgment of the truth of a conditional $A \rightarrow C$ and their subjective conditional probability $P(C \mid A)$ (e.g., Evans et al., 2003; Oberauer & Wilhelm, 2003; Over et al., 2007; Singmann et al., 2014).

Evans et al. (2003), for instance, provided their participants frequency information about the four cases $A\&C$, $A\&\neg C$, $\neg A\&C$, $\neg A\&\neg C$ and asked them about the probabilities of the conditional $A \rightarrow C$ and its so-called contrapositive, $\neg C \rightarrow \neg A$, where A and C referred to a color or a shape printed on cards. Based on the frequency information, the authors calculated the probability of the material implication ($P(A \supset C) = P(\neg A \lor C)$) and the conditional probabilities $P(C \mid A)$, respectively $P(\neg A \mid \neg C)$, and compared these and the joint probabilities $P(A \land C)$, respectively $P(\neg A \land \neg C)$, with participants' probability ratings for the conditional, respectively the contrapositive. Their results provide evidence for an interpretation according to the conditional probability and also according to the joint probability hypothesis, but not the material implication:[16] they found a high correlation between participants' ratings of the conditional and its contrapositive and the respective conditional probabilities as well as the respective joint probabilities while contrary to the predictions of the material implication, participants' ratings decreased when the frequency of $\neg A$ cases increased, and the ratings for the conditional and the contrapositive differed, which are, however, identical under the material implication. A regression analysis revealed strong individual differences though: some participants based their ratings on the joint probability while others based them on the conditional probability.

2.2.2 *Experiments on the meaning of conditionals*

In the following, I will first consider some experiments that showcase the context-dependent interpretations of conditionals. Then, I will turn to experiments that investigate whether conditionals are interpreted in line with the predictions of several theories about their meaning, including 'the Equation' and Inferentialism. Lastly, I will consider experiments that aim to find out whether the inferred dependency relation between antecedent and consequent can be explained by pragmatic enrichment or whether it is rather part of the semantics of conditionals.

16 The interpretation of a conditional $A \rightarrow C$ as the conjunction $A \land C$, i.e. $A \rightarrow C$ is evaluated as true if and only if A and C are both true, has particularly been shown to be endorsed by young children (see Barrouillet et al., 2000; Evans et al., 1993, p.42).

CONTEXT DEPENDENCY. An early experiment that investigated the context-dependency of peoples' interpretation of conditionals was done by Fillenbaum (1975). He let participants paraphrase different kinds of conditionals (e.g., threats, promises, etc.) and observed systematic differences in participants' responses, which suggests a different understanding depending on the respectively considered kind of conditional. Conditional threats were, for instance, often paraphrased as disjunctions which was only very rarely the case for conditional promises. Consider the utterance "You give me the money or I kill you" as paraphrase for the conditional "If you do not give me the money, I'll kill you' in comparison to the utterance "You lose or I will bake a cake for you" as paraphrase for the conditional "If you win, I'll bake a cake for you'; while the former is felicitous (although of course extremely questionable content-wise), the latter is clearly odd.

Dieussaert et al. (2002) empirically investigated how the number of alternative reasons for the consequent (other than the antecedent), speaker control (in terms of whether or not the speaker can influence the truth of the consequent) and pragmatic type (promise, threat, temporal, etc.) influence participants' interpretation of conditionals. All three factors were shown to have an effect on how participants interpreted the conditional (in terms of estimated likelihoods $L(A \mid C), L(C \mid A)$ for $A \rightarrow C$), yet the degree how influential they were varied; pragmatic type was found to be the most important of the considered factors.

For another example of context-dependency, consider (30) and (31) from Fugard et al. (2011). If peoples' interpretation of conditionals was merely guided by the conditional probability, (30) and (31) should be interpreted equally as the corresponding conditional probabilities are both 1. Fugard et al.'s (2011) results, however, showed that most participants indicated a degree of belief of zero for (30) whereas (31) was mostly assigned a probability of 1.

(30) If the card shows a 2, then it shows a 2 or a 4. $[A \rightarrow A \vee C]$

(31) If the card shows a 2, then it shows an even number. $[A \rightarrow B]$

An explanation for this result that the authors give, refers to the Gricean Maxim of Quantity: the consequent of (30) does not add any information whereas the consequent of (31) does and thus, the former is not assertable due to a violation of the Maxim of Quantity.

THE EQUATION. Despite the vast number of studies that investigated how people reason with conditionals, Douven and Verbrugge (2010) were the first who empirically investigated the relation between people's *acceptability* of a conditional $A \rightarrow C$ — instead of its *truth* — and the corresponding conditional probability. They tested slightly weakened forms of Adam's Thesis (AT), where the acceptability of an indicative conditional $A \rightarrow C$ is predicted to be only *approximately* equal to the conditional probability $P(C \mid A)$ or where acceptability and conditional probability correlate to a high or at least a moderate degree.

They put this general thesis to the test by considering it in light of three different types of *inferential* conditionals: conditionals $A \to C$ where the consequent, C, follows *deductively, abductively* or *inductively* from the antecedent, A, taking into account background premises that are salient in the utterance context. In deductive inferences, the consequent necessarily follows from the antecedent (e.g., "If Chelsea wins the Champions League in 2011, then that will be a first in the club's history" given background knowledge that Chelsea had never won the Champions League by that time), in abductive inferences, the consequent is the best explanation of the antecedent (e.g., "If Tom and Hank are jogging together, then they are friends again.", knowing that they have had an argument) and in inductive inferences, the consequent follows with a certain statistical probability (e.g., "If John is a second-year psychology student, he has passed his statistics exam", knowing that hardly any second-year psychology student failed). Whether they found evidence for AT or a weakened version of it was dependent on the type of conditional: a moderate correlation was found between the acceptability of inductive conditionals and the conditional probability $P(C \mid A)$, while for abductive conditionals the correlation was high. Yet, only deductive conditionals yield evidence for AT (Acc($A \to C$) $\approx P(C \mid A)$).

In a second experiment, Douven and Verbrugge (2010) asked participants to evaluate how probable it is that the conditional is true instead of asking for the acceptability of the given conditional statements. With this change, the results were in line with former experimental work showing participants' judgments of the truth of the conditionals to closely match the estimated conditional probabilities.

INFERENTIALISM It is uncontroversial that conditionals commonly convey a (dependency) relation between the antecedent and the consequent. But why do we infer a relation between antecedent and consequent of a conditional? Why do we tend to infer from (1), repeated here as (32), that *if* Bob goes to the party, it is *because* of Ann?

(32) If Ann goes to the party, Bob will go there, too.

According to Inferentialism, this is because the dependency relation is part of the core meaning of conditionals. Recently, a group of researchers has devoted their work on the empirical investigation of Inferentialism (e.g., Douven et al., 2018; Krzyżanowska et al., 2021; Krzyżanowska, Wenmackers, & Douven, 2013; Krzyżanowska, Wenmackers, Douven, & Verbrugge, 2013; Skovgaard-Olsen et al., 2016; Vidal & Baratgin, 2017). With the *Default and Penalty Hypothesis*, Skovgaard-Olsen et al. (2016), for instance, predict a diminished influence of the conditional probability in participants' acceptance ratings when the antecedent and the consequent are negatively, or not at all related, which cannot be explained with suppositional theories alone, in which the assertability/acceptability of conditionals is only based on the conditional probability. To test their hypothesis, the authors systematically varied relevance in their experiment (besides the prior probabilities of antecedent and consequent) as a within-participant manipulation so that the antecedent was either irrelevant for the consequent or had positive or negative impact on it. Their results showed that participants were indeed sensitive to the relevance (or irrelevance) between antecedent and consequent; high estimates for the conditional probability predicted higher acceptability ratings of the conditional in the positive relation condition than in the negative and irrelevance condition — yet, with substantial individual differences among participants.

Another experiment for testing Inferentialism was done by Douven et al. (2018) who showed their participants fourteen color patches of gradient colors with the left- and rightmost patches being clearly of a certain color (e.g., blue on the left and green on the right) while the inner patches evenly changed from the color on the leftmost patch to the color of the rightmost patch. They were then asked to judge conditionals like "If patch number 8 is green, so is patch number 11" as 'true', 'false' or 'neither true nor false', whereby the distance between the patch numbers from the antecedent and the consequent was manipulated as well as the direction of the inference, that is, whether the patch number in the consequent was smaller or larger than the patch number in the antecedent. As predicted by HIT (their proposed semantic combining Inferentialism and the suppositional conditional), the number of 'true'-responses was found to depend on both, direction and distance, which the authors referred to as *inferential strength effect*. The larger the distance between the two patches, mentioned respectively in the antecedent and the consequent, the less the conditional was judged as 'true', and conditionals with congruent direction (e.g., "If patch number x is green, so is patch number >x"; from blue to green) yield more true responses than conditionals with incongruent direction (e.g., "If patch number x is green, so is patch number <x"; from blue to green). Further, Douven et al. got almost no indeterminate answers (conditional rated as neither true nor false) which are commonly prevalent, in particular when the antecedent is false. This result was in accordance with the authors' predictions, because of, so they argue, the relevance between antecedent and consequent that was assured in their experiment.

In a follow-up paper Douven et al. (2020) analyzed these results again with the aim to explicitly compare them with respect to Inferentialism and other semantics of conditionals (Mental Model theory, suppositional theory) among which Inferentialism was found to explain their data best. In order to exclude the possibility that this finding was due to the fact that the conditionals in their experiment with the soritical series of color patches were rather abstract, they run another experiment with more realistic material. This was realized by using abductive conditionals whose consequent is considered (in a given context) to be an explanation for the antecedent (e.g., "If Judy and Pam are jogging together, then they have patched up their friendship") using varying degrees of the explanation quality, that is, how well the consequent explains the antecedent. Again, the results support Inferentialism — however, they do *not* discredit the conditional probability as reliable predictor for participants' truth ratings of the conditional, yet it was found to be a weaker predictor than explanation quality.

On the other hand, there is also empirical evidence speaking against Inferentialism (e.g., Cruz et al., 2016; Oberauer et al., 2007) which we will discuss next.

INFERRED LINK BETWEEN A AND C. In the previous paragraph, we considered some experiments that tested whether the dependency relation between antecedent and consequent that is commonly associated with conditionals is part of their core meaning, thus inherently associated with the word 'if'. Let us now turn to an experiment from Cruz et al. (2016) whose results speak in favor of a pragmatic explanation, that is, against Inferentialism.

They tested participants' readiness to draw various inferences, in particular, the inference from $A \wedge C$ to $A \rightarrow C$ (called *centering* or *conjunctive sufficiency*) which is valid under the conditional probability hypothesis and generally invalid given Inferentialism.[17] Although Cruz et al. observed lower ratings for the conditional conclusion $A \rightarrow C$ given the premise $A \wedge C$ when there was no inferential link between A and C (i.e., for missing-link conditionals), they showed that this result could be explained by the lack of a shared topic in A and C: when there was *no* inferential link between antecedent and consequent, but both propositions concerned the same topic, participants' ratings did *not* differ significantly from participants' ratings in the connected condition, which contradicts the predictions that Inferentialism makes. Therefore, the authors conclude, that instead of the missing link, it might be the lack of discourse coherence that causes the oddity of missing-link conditionals.

Moreover, participants' ratings for centering were more similar to the valid inferences that Cruz et al. (2016) tested for than for the invalid inferences, which also speaks against Inferentialism.

In response to the results from Cruz et al., Krzyżanowska et al. (2017) run another experiment to further test the influence that the

17 Centering is not an uncontroversial inference; evidence for peoples' readiness to draw or accept this inference was further found by Cruz et al. (2015), but see Skovgaard-Olsen et al. (2017) for evidence against it.

relation between antecedent and consequent on the one hand and discourse coherence on the other hand have on the assertability of conditionals. They asked participants for the sensibleness (would it make sense to say ...?) or assertability (would it be natural to assert ...?) of conditionals ($A \rightarrow C$) and of assertions of A and C as subsequent claims. Participants' ratings for the two types of assertions diverged only in the condition where there was no relation between A and C (i.e., $\Delta p = P(C \mid A) - P(C \mid \neg A) = 0$), but A and C were both about the same topic (e.g., "If Sophie likes the Alps, then mountaineering can be dangerous."). In this case, only conditionals received very low ratings which made the authors conclude that for a conditional to be assertable the antecedent must be relevant for the consequent — and accordingly, they argue that the oddity of missing-link conditionals is not sufficiently explained by the lack of a common topic in antecedent and consequent.

Note the different questions that Krzyżanowska et al. and Cruz et al. respectively asked for in their experiments: while Cruz et al. asked their participants *how confident they could reasonably be in the conclusions* (i.e., in $A \rightarrow C$), Krzyżanowska et al. asked for the assertability (either directly or by asking whether it 'would make sense to say ...') of the conditionals. However, it seems a reasonable possibility that one is confident in a certain conclusion without finding the respective assertion a natural thing to say; there might for instance be better alternatives or the speaker may simply not have a good reason for making this assertion, which might have influenced participants' responses.

2.3 SHORT SUMMARY

This chapter meant to give a broad overview of current and past research on conditionals investigating some of the many questions regarding conditionals that remain a subject of debate. As the experiments and theories considered in this background section showed, the utterance context in which a speaker chooses a conditional is important for the listener's interpretation thereof. However, while many theories of the meaning of conditionals require some kind of pragmatic enrichment to account for the diverse and context-dependent interpretations, no formal account has so far been proposed in this regard. This is the main goal pursued in this thesis, to develop a formal model of the pragmatics of conditionals. One of the open and recently strongly debated questions that we will also address in light of our model concerns the inference of the dependency relation between antecedent and consequent, and how it arises.

Part II

MODELING CONDITIONALS IN THE RATIONAL-SPEECH-ACT FRAMEWORK

In this part of the thesis I present a Rational-Speech-Act model — a probabilistic, Bayesian model — for investigating the use and interpretation of conditionals. The content of Chapters 3–5 was published in slightly different form (Grusdt, Lassiter, et al., 2022) in the Open Access journal *Semantics & Pragmatics* (vol. 15, 2022, article 13, see https://semprag.org/index.php/sp/article/view/sp.15.13/3055). Note that the paragraph on concessive conditionals in Section 5.4.3 was not part of the publication.

In Chapter 6, I present a novel behavioral experiment that we designed in order to put the proposed model to the test and in Chapter 7, I analyze the model with respect to the collected empirical data. A preliminary version of the analyses in Chapters 6–7 were published in the *Proceedings of the 43^th^ Annual Meeting of the Cognitive Science Society* (Grusdt & Franke, 2021).

3 RSA: A FORMAL MODEL OF PRAGMATIC REASONING

The Rational-Speech-Act (RSA) model is a prominent instance of a formalization of Gricean pragmatic reasoning using tools from probability calculus, decision and game theory (see Franke & Jäger, 2016; Goodman & Frank, 2016, for an overview). RSA models are probabilistic, data-driven and based on the assumption that linguistic behavior is goal-oriented and in this regard (approximately) optimal. A noteworthy benefit of a probabilistic and data-driven approach is that the predictions by the models can often be directly compared to quantitative aspects of experimental data, thereby allowing the statistical comparison of theoretically relevant models based on empirical data (e.g. Degen et al., 2020; Franke & Bergen, 2020; Qing & Franke, 2015). Moreover, as probabilistic modeling is prominent in other areas in the cognitive sciences, it becomes relatively easy to integrate insights and modeling components from these other areas as well, such as belief formation (Goodman & Stuhlmüller, 2013; Herbstritt & Franke, 2019), sequential adaptation (Schuster & Degen, 2019), or learning biases (Brochhagen et al., 2018). Since it was first introduced by Frank and Goodman (2012) in the context of a language reference game, the number of RSA models and modeled phenomena has been growing; scalar implicatures (Goodman & Stuhlmüller, 2013), hyperboles (Kao, Wu, et al., 2014) as well as more complex phenomena such as the interpretation of vague or polite language (Lassiter & Goodman, 2017; Yoon et al., 2016), projective content (Qing et al., 2016), metaphors (Kao, Bergen, et al., 2014) or social meaning (Burnett, 2019) have been investigated by means of RSA.[1]

There are several reasons why RSA also seems promising for modeling the interpretation of conditionals. First of all, we expect it to be influenced by the availability of non-conditional utterances that the speaker might have chosen instead. In this way, RSA formalizes a Gricean account of listeners' interpretations and the speakers' utterance choices. Since RSA is able to flexibly integrate contextual knowledge, we can model quite different utterance contexts, including situations in which richly structured world knowledge is relevant. This will be crucial in our account of the effects of causal knowledge on the interpretation of conditionals.

1 A hands-on introduction to RSA modeling with examples in the probabilistic programming language WebPPL(Goodman & Stuhlmüller, 2014), in which also the model presented here is implemented, is provided in the web-book *Probabilistic Language Understanding* (Scontras et al., 2018).

In the following, Section 3.1 first gives a general introduction to the vanilla version of the RSA model. Subsequent sections then elaborate on the specific adaptations necessary in order to capture reasoning about the use of conditionals against the background of richly structured causal world knowledge. First of all, we pin down an appropriate set of state representations and alternative utterances, together with an appropriate semantic denotation function in Section 3.2.1. Based on these definitions, we use Section 3.2.2 to demonstrate the main underlying ideas of the model by means of a concrete example with only three world states before we lay out the assumed priors on the generalized set of states in Section 3.2.3 and Section 3.2.4.

3.1 THE VANILLA RATIONAL-SPEECH-ACT MODEL

At the heart of the vanilla RSA model lies the formalization of a cooperative Gricean speaker who, when trying to communicate a state s, probabilistically selects an utterance u by preferably choosing utterances that are not only true, but also maximize the amount of relevant information conveyed to a literal listener. The pragmatic listener is simply modeled as a rational interpreter who combines, using Bayes' rule, their prior beliefs with the speaker's protocol of choosing utterances.

In order to capture the pragmatic speaker's behavior, in particular, in order to ground out a notion of truth and informativity, the RSA model considers a literal listener first, whose interpretation behavior is defined as a conditional probability distribution $P_{\text{lit}}(s \mid u)$ obtained by updating any prior beliefs $P_{\text{prior}}(s)$ about likely world states s with the set of states $[\![u]\!]$ where utterance u is considered true or permissible.[2]

$$P_{\text{lit}}(s \mid u) \propto \delta_{s \in [\![u]\!]} \cdot P_{\text{prior}}(s) \qquad [4]$$

2 The formula in Equation [4], and in Equations [6] and [7] for other model components, gives the probability up to proportionality ($\propto$), leaving the normalizing constant of the probability distribution implicit. If $F(x) \geq 0$ is a non-normalized score for any $x \in X$ with X a finite set, the notation $P(x) \propto F(x)$ is shorthand for $P(x) = \frac{F(x)}{\sum_{x'} F(x')}$. Moreover, $\delta_{s \in [\![u]\!]}$ is the Kronecker delta function, which returns 0 or 1 depending on whether its Boolean argument, here the denotation function of utterance u applied to state s, is true (1) or false (0).

The pragmatic speaker is then defined in terms of a notion of *utterance utility* $U(u; s)$, which reflects how informative u is for communicating s.[3]

$$U(u; s) = \log P_{\text{lit}}(s \mid u) \quad [5]$$

The informativity of an utterance u as description of a state s is defined as the log-likelihood (negative surprisal) of the literal listener's beliefs for state s after hearing utterance u. The probability that a speaker in state s will choose utterance u is then defined as a soft-max operation on utility scores:

$$P_S(u \mid s) \propto \exp(\alpha \cdot U(u; s)) \quad [6]$$

α is a model parameter governing how closely the speaker approximates utility maximization: the higher α is, the more likely the speaker is to choose the "rational", utility-maximizing utterance. At the extremes, a hyperrational speaker (with $\alpha \to \infty$) would only choose utterances that maximize utility, and a randomizing speaker (with $\alpha = 0$) would choose randomly among true utterances without regard for their utility.

Finally, the pragmatic listener's interpretation is captured by a conditional probability distribution, $P_{PL}(s \mid u)$, which represents the listener's *a posteriori* beliefs (after having heard utterance *u*) about the probability of state s, taking the priors over states and a Gricean speaker's utterance-choice behavior into account.

$$P_{PL}(s \mid u) \propto P_S(u \mid s) \cdot P_{\text{prior}}(s) \quad [7]$$

3.2 AN RSA MODEL FOR COMMUNICATION WITH CONDITIONALS

3.2.1 *World states, utterances & assertability*

WORLD STATES. Conditionals like $A \to C$ are often associated with the speaker's uncertainty about whether A and/or C are true. Therefore, to model pragmatic reasoning about conditionals, we include potentially uncertain speakers into our modeling. Concretely, we look at the partition of possible worlds into the four types of worlds which agree on the truth values of A and C: $W = \{w_\emptyset, w_A, w_C, w_{AC}\}$. The concrete set of states used are probability distributions over worlds $w \in W$.

3 A common modification of the basic definition of utterance utility in terms of informativity are utterance costs that are often used to account for differences among utterances, for instance regarding salience, social compatibility, or their complexity with respect to utterance length, pronunciation etc. (e.g., see Gates et al., 2018; Qing et al., 2016). Even though we use utterances of different complexities, we prefer not to rely on particular utterance costs here which could nevertheless easily be integrated into the model.

There are at least two prominent possibilities of how to interpret probability distributions: as precise *objective* chances or as *subjective* beliefs. Consequently, world states in our model can be interpreted in different ways, too. For one, we can either think of $s \in S$ as the true beliefs of a maximally competent speaker about the objective chance of each type of world. In this case, the conversational goal, implicitly defined in the vanilla RSA model, is to communicate the true (objective, but intrinsically stochastic) world state known to the speaker. For another, we can conceptualize world states as a representation of an uncertain speaker's subjective beliefs $s \in S$ about the true (non-probabilistic) state of the world $w \in W$. Under this interpretation, the vanilla RSA model implicitly treats the conversational goal as that of communicating the speaker's belief state (see Aloni, 2007; Franke, 2011). The interpretation of world states, as either objective or subjective, may matter for the interpretation of the assertability conditions presented below. If not stated otherwise, we refer to the subjective interpretation of model states.

UTTERANCE ALTERNATIVES. Predictions of Gricean pragmatic reasoning strongly depend on the assumed set of alternative utterances. There has been much discussion of alternative sets for scalar items like *some*, *warm* and *or* (e.g. Katzir, 2007; Matsumoto, 1995), but much less for pragmatic reasoning about conditionals (for some discussion see van der Auwera, 1997; von Fintel, 2001). The selection of alternative utterances to consider in the following is largely governed by the desire to present a balanced set of alternative utterances which is sufficient to describe the most salient differences in the set of world states. Utterances are compositionally built from literals, possibly negated. They may be combined with *and* to form a conjunction, with *if* to form a conditional or with the word *likely*. Table 3 below lays out the alternative utterances that our model considers, together with the rule used to compute the update effects of each — its "assertability condition", to be introduced next.

ASSERTABILITY CONDITIONS. Given a world state $s \in S$ and an utterance $u \in U$, we would like to define the semantics $[\![u]\!] \subseteq S$ to serve as the anchoring of pragmatic reasoning in literal interpretation, as defined in Equation [4]. As especially the semantic meaning of conditionals is a highly controversial issue, we would like to stay as uncommitted and encompassing as possible. This is possible, to a certain extent, if we focus not on the nature of the denotation function $[\![u]\!]$ but rather at the functional role it plays in the architecture of the pragmatic reasoning model. In particular, since the utility function in Equation [5] and the speaker rule in Equation [6] entail that whenever $s \notin [\![u]\!]$, the speaker will not choose u when in state s, the main effect of $[\![u]\!]$ is to give *assertability conditions* and — as a side-effect — give information about how informative each utterance is.[4] Consequently, our model lays out a general method of computing update effects of utterances with conditionals at the level of a literal interpreter with the ulterior goal of defining reasonable speaker behavior, while avoiding as much as possible concrete commitment to a specific semantic interpretation.

We treat utterances of literals like "A" as conveying that the target state s makes the probability that A is true (i.e., $A = a$) high enough for conversational purposes; the corresponding probability, e.g., $P^{(s)}(A = a)$ for literal "A", must exceed a certain threshold for the respective utterance to be assertable.[5] This threshold is represented in the model by the free parameter θ, which we set to 0.9 in all simulations reported below. In a model of objective chance, as we assume on an objective interpretation of the probability distributions building up the set of world states, determinate truth corresponds to probability 1. That is, a speaker will treat A as true just in case A is determinately true in s — when $P^{(s)}(A = a) = 1$. The assumption that speakers sometimes assert things that are not certain, but very likely true, yet justifies a threshold below 1 as assertability condition of utterances. We also make use of this assumption on a subjective interpretation of world states in our model; a factual sentence A is thus assertable as long as the speaker's subjective belief in A is sufficiently large ($P^{(s)}(A = a) \geq \theta$, respectively $P^{(s)}(A = \neg a) \geq \theta$ when $u = \neg A$).

Similarly, "*likely A*" directly conveys that the subjective probability of A in s is greater than 0.5 as we assume that "*likely A*" is assertable in s if and only if $P^{(s)}(A = a) > 0.5$. This aspect of our account is reminiscent of expressivist accounts of probability language (Moss, 2015; Swanson, 2016; Yalcin, 2012). On the objective interpretation, this means that *likely* expresses high objective chance, consistent with the empirical findings of Ülkümen et al. (2016) and Lassiter (2018b). Note, however, that these authors show that *likely* can also express subjective uncertainty.

4 The informativity of the modeled utterances generally follows the order shown in Table 3. As we were pointed at by an anonymous reviewer, depending on the chosen prior over states, it is, however, possible for a conditional (e.g., $A \rightarrow C$) to be *literally* more informative than a literal (e.g., C) since the assertability of a literal does not *per se* entail the assertability of a conditional, e.g. $P^{(s)}(c) \geq \theta \not\Rightarrow P^{(s)}(c \mid a) \geq \theta$ (but see Footnote 6 and the text it refers to). For example: $P(A, C) = \langle w_{AC} = 0.4, w_A = 0.09, w_C = 0.5, w_\varnothing = 0.01 \rangle$. Here $P(c \mid a) \approx 0.82 < \theta = 0.9$ but $P(c) = 0.9 = \theta$. This is an interesting observation, in particular in the context of conditionals whose consequent is inferred to be independent of the antecedent as for instance in concessive conditionals (e.g., "*Even if ...*") which is certainly worth looking at in future work, but beyond the scope of what we cover here.

5 We write $P^{(s)}$ to refer to probabilities *within* states to distinguish them from probabilities *across* states. That is, $P^{(s)}(X)$ denotes the probability assigned to any event X that may be inferred from world state s, e.g., $P^{(s)}(A = a)$ is the probability of A to be true in world state s.

In parallel fashion, we render the assertability conditions of an indicative conditional $A \rightarrow C$ as $P^{(s)}(C = c \mid A = a) \geq \theta$, corresponding either to a high objective chance of C to be true when A is true or to a strong belief in this conditional probability (on part of the speaker), assuming an objective and subjective interpretation respectively. Since we do not want or need to commit to a specific semantic theory of conditionals here, we settle for motivating this condition as a plausible minimal bound on assertability that should be acceptable from a wide range of theoretical perspectives. For theories that are able to support the equation $P(A \rightarrow C) = P(C \mid A)$ while avoiding triviality results (e.g., Hájek 1989; D. Lewis 1976), the derivation is strictly parallel to the factual case above. This includes non-propositional theories (Bennett, 2003; Edgington, 1995), trivalent theories (De Finetti, 1995; Lassiter, 2020; Milne, 1997) and various others (Kaufmann 2004; Khoo 2016; Stalnaker and Jeffrey 1994; van Fraassen 1976, among others). Our assertability condition is also particularly natural for theories that render the truth-condition of $A \rightarrow C$ as $P(C \mid A) = 1$ (e.g., Moss, 2015).

The status of our assertability condition for conditionals is somewhat murkier from the perspective of other prominent theories such as Stalnaker (1968) and Kratzer (2008) as well as strict conditional theories. Because of the complexity of the way that they assign truth-values to epistemically possible worlds where the antecedent A is false, these theories can make $P(A \rightarrow C)$ high even while $P(C \mid A)$ is low. As a result, our assertability condition is stronger than these accounts would predict. However, we note that there is by now an enormous body of empirical evidence supporting the equation between the probability of a conditional and the corresponding conditional probability (Douven and Verbrugge 2010; Evans and Over 2004; Hadjichristidis and Stevenson 2001, among many others). This evidence problematizes a key prediction made by the latter group of theories: that a conditional can be judged highly probable simply because of the likely falsehood of its antecedent. Instead, situations where the antecedent is false are generally judged irrelevant to the probability of a conditional, in a probabilistic analogue of the classic paradoxes of the material conditional (Edgington, 1995). We do not doubt that the theories under consideration have theoretical resources available that may allow them to avoid this problem — for example, by using pragmatic reasoning to explain why false-antecedent cases are not considered relevant in assertion (Grice, 1989; D. Lewis, 1976). But doing so would be tantamount to adopting our assertability condition or something quite close to it. As a result, we believe that our results should be relevant to theorists with a wide variety of semantic commitments, including those for whom probabilistic reasoning has not previously played a major theoretical role.

utterance type	assertability in state s	example: utterance u	assertability u in state s
conjunction	$P^{(s)}(\phi, \psi) \geq \theta$	$A \wedge \neg C$	$P^{(s)}(A = a, C = \neg c) \geq \theta$
literal	$P^{(s)}(\phi) \geq \theta$	A	$P^{(s)}(A = a) \geq \theta$
conditional	$P^{(s)}(\psi \mid \phi) \geq \theta$	$A \rightarrow \neg C$	$P^{(s)}(C = \neg c \mid A = a) \geq \theta$
likely + literal	$P^{(s)}(\phi) > 0.5$	$likely\ \neg C$	$P^{(s)}(C = \neg c) > 0.5$

Table 3: Types of utterances with corresponding assertability conditions and an example, ordered from most informative utterance on top to least informative at the bottom. For conditionals and conjunctions, $\phi \neq \psi$.

In sum, under a wide range of plausible semantic theories of conditionals and interpretations of our RSA model, we are led to the set of assertability conditions summarized in Table 3.

INFORMATIVITY OF UTTERANCES. How much information an utterance provides is naturally linked with the defined assertability conditions; the larger the number of states that can truthfully be described with an utterance u, the less informative u will be since fewer referents are excluded as possible target states described by the speaker. Table 3 lists the four utterance types according to their informativity: conjunctions are more informative than literal assertions since whenever a conjunction is assertable, the literal assertions of each conjunct will be assertable as well (e.g., $P^{(s)}(a, c) \geqslant \theta \Rightarrow P^{(s)}(a), P^{(s)}(c) \geqslant \theta$). That is, the assertability of a conjunction entails the assertability of the corresponding literal assertions, rendering the latter less informative. A truthful assertion of a literal (e.g. C, $\neg A$, etc.), in turn, entails the assertability of the corresponding expression with '*likely*' when the assertability threshold θ is larger than 0.5 which can reasonably be assumed (e.g., $P^{(s)}(c) \geqslant \theta \Rightarrow P^{(s)}(c) > 0.5$ for $\theta > 0.5$). Turning to conditionals, they will be less informative than conjunctions since the assertability of a conjunction entails the assertability of the corresponding conditional (e.g., $P^{(s)}(a, c) \geqslant \theta, P^{(s)}(c \mid a) = \frac{P^{(s)}(a,c)}{P^{(s)}(a)}$ and $P^{(s)}(a) \leqslant 1 \Rightarrow P^{(s)}(c \mid a) \geqslant P^{(s)}(a, c) \geqslant \theta$). A literal may, however, at least theoretically be *less* informative than a conditional. The literal 'C' is, for instance, not necessarily more informative than the conditional $A \rightarrow C$: when $P^{(s)}(c) \geqslant \theta$, that is, '$C$' is assertable, it does not entail that $P^{(s)}(c \mid a) \geqslant \theta$ and so the assertability of the conditional $A \rightarrow C$ does not follow from the assertability of 'C' (unless A and C are stochastically independent where $P(c \mid a) = P(c)$). However, we will see that for a subset of states, $P^{(s)}(C) \geqslant \theta$ will entail $P^{(s)}(c \mid a) \geqslant \theta$.[6]

6 When $P^{(s)}(c) \geqslant \theta$, i.e. when '$C$' is assertable, $P^{(s)}(c \mid a) \geqslant \frac{\theta - P^{(s)}(c \mid \neg a) \cdot P(\neg a)}{P(a)}$ which will be greater than or equal to θ, when $P^{(s)}(c \mid \neg a) \leqslant \theta$ holds. For example, for the

3.2.2 *Toy example*

Let us consider a small toy example in order to illustrate model predictions, and to motivate further generalizations to be introduced hereafter. Concretely, we will consider a case with just three world states, all equally likely and just constructed here for the sake of illustration. The topic of conversation in this example is whether Alex and Chris are likely to go to the party. The three states we look at differ in the probability they assign to all the four logical possibilities of Alex and/or Chris going to the party.

(s_1) Alex and Chris both really like to go out. Both are seen at most parties, but whether either comes is unrelated to whether the other might come.

(s_2) Alex and Chris go slightly more often than not, but usually not without each other (e.g., they might be a couple, best friends, etc.).

(s_3) Alex and Chris each go out more often than not, but have no connection with each other.

These three scenarios are translated into the probability distributions $P(A, C)$ shown at the top of Table 4 where variable A denotes whether Alex comes to the party and C denotes Chris coming to the party.

With this context in mind, imagine the following utterance of Bigi directed at Wobo:

(33) Bigi: *Chris will come to the party.* ["C."]

Given that Bigi only says what she believes to be true and s_1, s_2 and s_3 describe all possible scenarios, Wobo should infer that Bigi describes the situation modeled in s_1 since this utterance (C) is neither assertable in s_2 nor in s_3.

Now, imagine Bigi to utter the following conditional instead:

(34) Bigi: If Alex comes to the party, Chris comes too. ["If A, C."]

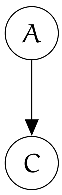
As noted in Table 4, $A \rightarrow C$ is assertable in states s_1 and s_2. That is, under a literal interpretation of Bigi's utterance, s_1 and s_2 are considered equally likely ($P_{\text{lit}}(s_1 \mid u = A \rightarrow C) = P_{\text{lit}}(s_2 \mid u = A \rightarrow C) = 0.5$). However, under a pragmatic interpretation of the utterance $A \rightarrow C$, s_2 is judged as more likely than s_1 ($^{11}/_{16}$ vs. $^{5}/_{16}$). This is because Bigi could have chosen a more informative utterance to communicate s_1.

---

relation $r = A \overset{++}{\rightsquigarrow} C$, introduced in Section 3.2.4, where $P^{(s)}(c \mid \neg a)$ is sampled from beta(1,10) the probability that $P^{(s)}(c \mid \neg a) \leq \theta$ equals 1 when $\theta = 0.9$, and when $\theta = 0.5$, the probability is still 0.999.

		s_1	c	¬c	s_2	c	¬c	s_3	c	¬c
a)	$P(A, C)$	a	0.81	0.09	a	0.6	0.05	a	0.36	0.24
		¬a	0.09	0.01	¬a	0.05	0.3	¬a	0.24	0.16
b)	$u = \text{likely } C$		1			1			**1**	
	$u = A \rightarrow C$		1			**1**			0	
	$u = C$		**1**			0			0	
	$u = A \wedge C$		0			0			0	
c)	$P_{\text{lit}}(s_i \mid u = \text{likely } C)$		$1/3$			$1/3$			$1/3$	
	$P_{\text{lit}}(s_i \mid u = A \rightarrow C)$		$1/2$			$1/2$			0	
	$P_{\text{lit}}(s_i \mid u = C)$		1			0			0	
	$\sum_{u'} P_{\text{lit}}(s_i \mid u')$		$11/6$			$5/6$			$1/3$	
d)	$P_S(u = \text{likely } C \mid s_i)$		$\frac{1}{3}/\frac{11}{6} = 2/11$			$\frac{1}{3}/\frac{5}{6} = 2/5$			$\frac{1}{3}/\frac{1}{3} = 1$	
	$P_S(u = A \rightarrow C \mid s_i)$		$\frac{1}{2}/\frac{11}{6} = 3/11$			$\frac{1}{2}/\frac{5}{6} = 3/5$			0	
	$P_S(u = C \mid s_i)$		$1/\frac{11}{6} = 6/11$			0			0	
e)	$P_{PL}(s_i \mid u = \text{likely } C)$		$\frac{2}{11}/(\frac{2}{11}+\frac{2}{5}+1) = 10/87$			$\frac{2}{5}/(\frac{2}{11}+\frac{2}{5}+1) = 22/87$			$1/(\frac{2}{11}+\frac{2}{5}+1) = 55/87$	
	$P_{PL}(s_i \mid u = A \rightarrow C)$		$\frac{3}{11}/(\frac{3}{11}+\frac{3}{5}) = 5/16$			$\frac{3}{5}/(\frac{3}{11}+\frac{3}{5}) = 11/16$			0	
	$P_{PL}(s_i \mid u = C)$		$\frac{6}{11}/\frac{6}{11} = 1$			0			0	

Table 4: Model predictions for the scenario given in Section 3.2.2 with parameters $\alpha = 1, \theta = 0.9$. a) States s_1, s_2, s_3. b) assertability of utterances given states. The most informative, assertable utterances for each state are highlighted in bold. c) Literal interpretation. d) Speaker production likelihoods. e) Pragmatic interpretation.



$$P(A = a) = 0.65$$
$$P(C = c \mid A = a) = 12/13$$
$$P(C = c \mid A = \neg a) = 1/7$$

Figure 2: Bayes net representing s_2 from Table 4, that consists of a graphical representation (left) and a set of associated (conditional) probabilities (right) that define a joint probability distribution over A and C, $P^{(s_2)}(A, C)$.

3.2.3 *Inferring latent causal relations*

The example in Table 4 shows how the model introduced so far yields the inference that $A \rightarrow C$ is most likely associated with world state s_2, by rather straight-forward Gricean reasoning. The example also demonstrates how, on top of inferring a state s, Wobo might draw inferences about the likely causal relation between propositions A and C which may have led to Bigi's beliefs as captured in s_1, s_2 or s_3. By motivation of the example, the coming of Alex and Chris was assumed to be *independent* in states s_1 and s_3, while the very reason for writing a table like in s_2 was because we assumed that there was a stochastic relationship between Alex's and Chris' coming to the party. Suppose, for simplicity, that there are two equally likely possible states $r \in \{\text{independent}, \text{dependent}\}$, meaning that A and

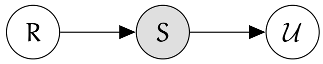


Figure 3: Through world knowledge, beliefs about the causal relation $r \in R$ inform probabilities $s \in S$ which are, in turn, considered as known by the speaker (either s represents the speaker's subjective beliefs or the speaker is considered maximally competent and knows the precise objective chances represented in s). Probabilities in s then directly influence the speaker's utterance choice $u \in \mathcal{U}$.

C are either independent or dependent. It is intuitive to think that $P(S = s_2 \mid r = \text{independent})$ is much smaller than $P(S = s_1 \mid r = \text{independent})$ or $P(S = s_3 \mid r = \text{independent})$, and also that $P(S = s_2 \mid r = \text{dependent})$ is much higher than $P(S = s_1 \mid r = \text{dependent})$ or $P(S = s_3 \mid r = \text{dependent})$. In this way, by Bayesian inference, we can obtain an indirect inference of a likely causal/stochastic relation between A and C just from probabilistic pragmatic reasoning *and* natural assumptions about the differential likelihood between different causal/stochastic relations r and states s. Note the importance of the pragmatic reasoning for drawing an inference about the likely relation in our small toy example: under a literal interpretation of the conditional $A \to C$, Wobo would not show a preference between s_1, where A and C are likely dependent and s_2, where they are likely independent.

We can think of this model as a sequence of inferences: beliefs about r stochastically inform s via world knowledge or intuitions about dependence/independence. The s variable, in turn, stochastically informs the speaker's utterance choice u due to pragmatic constraints on what counts as a good utterance. Schematically: $P(r) \Rightarrow P(s \mid r) \Rightarrow P(u \mid s)$ which is depicted in Figure 3. We can then derive $P(r \mid u)$ via Bayes' rule. In the following we specify and motivate a concrete prior structure, in particular for the $P(s \mid r)$ part, so as to be able to derive general predictions from this model for what we may consider a *default context*, where no specific world knowledge is assumed to be available regarding antecedent and consequent.

3.2.4 *Prior over world states in default context*

A state s where variables A and C are assumed to be independent ($r = A \perp\!\!\!\perp C$) is represented by probabilistically independent distributions where $P^{(s)}(A, C) = P^{(s)}(A) \cdot P^{(s)}(C)$. Since in the default context, we do not make any specific assumptions, A and C are both assigned a uniform prior probability over the interval $(0, 1)$:

$$P^{(s)}(A = a),\ P^{(s)}(C = c) \sim \text{Uniform}(0, 1) \qquad [8]$$

causal relation (R)	instance causal relation (r)	interpretation
$A \rightsquigarrow C$	$A \overset{++}{\rightsquigarrow} C$	Truth of A increases probability for truth of C
	$A \overset{-+}{\rightsquigarrow} C$	Falsity of A increases probability for truth of C
$C \rightsquigarrow A$	$C \overset{++}{\rightsquigarrow} A$	Truth of C increases probability for truth of A
	$C \overset{-+}{\rightsquigarrow} A$	Falsity of C increases probability for truth of A

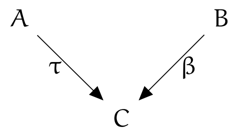
Table 5: Notation for dependent causal relations (types and instances). The instance of the causal relation provides information about the associated joint probability tables, spelled out in column 'interpretation'.

Together with the assumption of independence, this yields the following probability distribution over partitions of possible worlds, representing a single world state s when $r = A \perp\!\!\!\perp C$:

$$
\begin{aligned}
P(w_{AC}) &= P^{(s)}(A = a) \cdot P^{(s)}(C = c)\\
P(w_{A}) &= P^{(s)}(A = a) - P(w_{AC})\\
P(w_{C}) &= P^{(s)}(C = c) - P(w_{AC})\\
P(w_{\emptyset}) &= 1 - (P(w_{AC}) + P(w_{A}) + P(w_{C}))
\end{aligned}
$$

To derive the probability distributions over partitions of possible worlds that represent states where variables A and C are assumed to be dependent, we distinguish between two possible *types* of causal relation: either A has causal power to provoke C ($R = A \rightsquigarrow C$) or vice versa, that is, C has causal power to provoke A ($R = C \rightsquigarrow A$).[7] While the causal relation, R, merely provides information about the causal direction, the concrete *instances* of a causal relation, denoted as r, are distinguished based on how exactly the causal relation affects conditional probabilities, as shown in Table 5. The type of causal relation $R = A \rightsquigarrow C$, for instance, tells us that the outcome of variable A has direct influence on the outcome of variable C and the instance of the causal relation, r, further tells us *how* the outcome of A influences the outcome of C. For example, we write $r = A \overset{++}{\rightsquigarrow} C$ to pick out the class of probability distributions where the truth of A ($A = a$) increases the probability that C is true ($C = c$), whereas $r = A \overset{-+}{\rightsquigarrow} C$ picks out probability distributions where the *falsity* of A increases the probability that C is true. Putting this differently, when $r = A \overset{++}{\rightsquigarrow} C$,

7 The choice to model only a single relevant cause in the default context is corroborated by theoretical as well as empirical findings: in a small, theoretical case study from Icard and Goodman (2015) the loss of information resulting from the neglect of alternative causes was, on average, so small that taking an alternative cause into account is only justified if it has very low cost and many empirical studies have shown that people have a tendency to neglect alternative causes (see e.g., Fernbach & Darlow, 2010; Fernbach et al., 2011; Fernbach & Rehder, 2013; Hagmayer & Waldmann, 2007; Krynski & Tenenbaum, 2007).



$$P(C = c \mid A = a) = 1 - (1 - \tau) \cdot (1 - \beta) = \tau + \beta - \tau \cdot \beta$$
$$P(C = c \mid A = \neg a) = \beta$$

Figure 4: Graphical representation of a leaky noisy-or model (left) with a single explicitly modeled cause A and the corresponding conditional probabilities of C when A is true or false (right). Variable B summarizes all potential other causes of C; τ and β denote the causal power of A, respectively B, to induce the truth of C.

w_{AC} will be considerably more likely than w_A and when $r = A \overset{-+}{\rightsquigarrow} C$, $P(w_C) \gg P(w_\varnothing)$.

Formally, the probability distributions of the dependent states can be described as leaky noisy-or model with binary variables allowing for positive as well as negative causes. Noisy-or models (Pearl, 1988) describe the relationship between an effect variable and its cause variables where each cause is capable of producing the effect independently of all other causes; corresponding to a logical OR-function where the effect is only true if at least one of its causes is true. Leaky noisy-or models (Díez, 1993) comprise 'background noise' represented by an additional cause variable that is always present and summarizes all potential causes of the effect that are not explicitly modeled. Figure 4 shows a graphical representation of the leaky noisy-or model underlying the dependent states in our model with the corresponding conditional probabilities of the effect C to be true when A is true or false respectively, assuming non-deterministic relations where variables A and B have causal power τ, respectively β, to provoke C. Since we want to stick to the simplest set of induced probability distributions, but still need to cover all possible stochastic relations between the two explicitly modeled variables A and C to have a balanced set of states that preserves the informativity of utterances, we use a generalization of the classical noisy-or model which further allows negative causes: not only the truth, but also the falsity, of a cause can have an influence on the truth of the effect (e.g., see Hyttinen et al., 2011).[8] Therefore, C is likely true when A is false and a negative cause ($r = A \overset{-+}{\rightsquigarrow} C$) or when A is true and a positive cause ($r = A \overset{++}{\rightsquigarrow} C$); in both cases the truth of C may independently be due to background noise B.

To instantiate the set of probability distributions for the respective dependent causal relations, we specify the prior distributions over

8 If we only used positive causes, states where $P(w_{AC})$ and $P(w_\varnothing)$ tend to be low would be very rare and thus, conditionals like $A \to \neg C$ would only be assertable in very few states which would, in turn, render this utterance very informative. As a result, this particular conditional would become more informative than for instance literals, which does not seem reasonable. Using states with positive and negative causes results in a balanced set of states such that any conjunction is more informative than any literal and any literal is, in turn, more informative than any conditional.

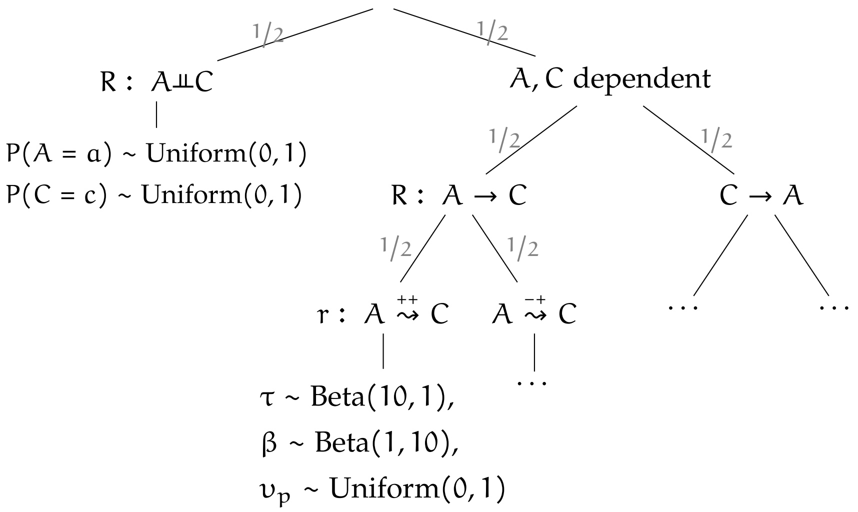


Figure 5: Graphical representation of the procedure for sampling a state s from the prior in the default context.

the respective causal power (τ), noise (β) and prior probability of the parent variable (υ_p) as shown in Figure 5.[9] The values of the hyperparameters for the beta distributions are chosen such that the mean of the distribution of the causal power τ exceeds the assertability threshold, set to 0.9 in all simulations below; for the causal power of the background noise, the parameters are simply reversed such that the prior distribution $P(\beta)$ is skewed towards 0. As we consider a default context here, where no further information is available, the prior probability of the parent (υ_p) is sampled from a uniform distribution over the interval $(0, 1)$. The joint probability distributions over A, C are then build based on the sampled values of τ, β and υ_p. Depending on the relation R and the cause variable being a positive or a negative cause, β, υ_p and υ_c correspond to different (conditional) probabilities, as listed in Table 6 where υ_c denotes the conditional probability of the effect (C) to be true when the cause variable is true: $\upsilon_c = \tau + \beta - \tau \cdot \beta$ (see Figure 4).

The joint probability distributions over partitions of possible worlds are then computed as follows when $R = A \rightsquigarrow C$ (similarly for $R = C \rightsquigarrow A$):

$$P(w_{AC}) = P(c \mid a) \cdot P(a), \qquad P(w_A) = P(\neg c \mid a) \cdot (P(a)$$
$$P(w_C) = P(c \mid \neg a) \cdot P(\neg a), \qquad P(w_\varnothing) = P(\neg c \mid \neg a) \cdot P(\neg a)$$

In total, we sample 10,000 probability distributions that comprise the set of world states used by our model. More concretely, we first

9 Note that even though the notion of *causal* relations is eventually not important as the modeling hinges on certain probabilistic dependencies between events (depending on the relation), it is important for the choice of our prior distributions which is motivated by relations that are causal by nature.

instance causal relation (r)	υ_p	υ_c	β
$A \overset{++}{\rightsquigarrow} C$	$P^{(s)}(A = a)$	$P^{(s)}(C = c \mid A = a)$	$P^{(s)}(C = c \mid A = \neg a)$
$A \overset{-+}{\rightsquigarrow} C$	$P^{(s)}(A = \neg a)$	$P^{(s)}(C = c \mid A = \neg a)$	$P^{(s)}(C = c \mid A = a)$
$C \overset{++}{\rightsquigarrow} A$	$P^{(s)}(C = c)$	$P^{(s)}(A = a \mid C = c)$	$P^{(s)}(A = a \mid C = \neg c)$
$C \overset{-+}{\rightsquigarrow} A$	$P^{(s)}(C = \neg c)$	$P^{(s)}(A = a \mid C = \neg c)$	$P^{(s)}(A = a \mid C = c)$

Table 6: Probabilities ($\upsilon_p, \upsilon_c, \beta$) that define the joint probability distribution of a state s, $P^{(s)}(A, C)$, for each instance of a dependent causal relation. υ_p is the prior probability of the cause, υ_c is the conditional probability of the effect to be true when the cause is true and β is the power of the unmodeled variables to provoke the effect which corresponds to the conditional probability of the effect to be true when the explicitly modeled cause is false.

sample a causal relation r from its prior distribution given in Equation [9]. Our choice to put a prior on the causal relation, which determines the shape of the associated sampled probability distributions, is primarily based on work from Griffiths and Tenenbaum (2005), Tenenbaum and Griffiths (2003); they were the first who put priors on the causal structure itself to predict human causal judgments instead of focusing on learning the causal strength of a, possibly non-existent, causal link (e.g., see Cheng, 1997).

$$P(r) = \begin{cases} 1/2 & \text{if } r = A \perp\!\!\!\perp C \\ 1/8 & \text{if } r \in \{A \overset{++}{\rightsquigarrow} C, A \overset{-+}{\rightsquigarrow} C\} \\ 1/8 & \text{if } r \in \{C \overset{++}{\rightsquigarrow} A, C \overset{-+}{\rightsquigarrow} A\} \end{cases} \qquad [9]$$

Based on the causal relation r, we then sample probability distributions according to the procedure described above. Since we do not assume a preference for dependent or independent relations, this results in an approximately equal number of states where A and C have a stochastic relationship and where they are probabilistically independent. A visualization of the sampled states is given in Figure 6 which shows histograms of the probabilities for each of the four possible worlds across all sampled states for three selected causal relations.[10]

3.2.5 *Communicating causal information implicitly via conditionals*

So far, we have specified how the joint probability distributions are derived under the assumption of particular causal relations. That is, strictly speaking, our world states have two components, a probability

10 The source code and all modeling results are publicly available: https://osf.io/6bshq/?view_only=1703a3417a1343a5a66b78ac8ce206c2.

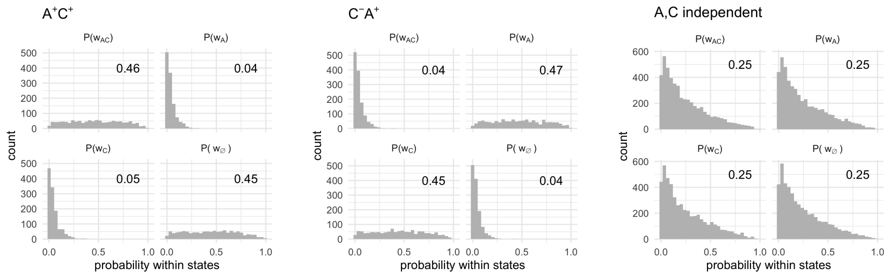


Figure 6: Histograms of the probabilities of the four possible worlds, w_{AC}, w_A, w_C and $w_{\emptyset}$ of all sampled probability tables $P^{(s)}(A, C)$ with $r = A \overset{++}{\rightsquigarrow} C$ (left), $r = C \overset{-+}{\rightsquigarrow} A$ (middle) and $r = A \perp\!\!\!\perp C$ (right). Numbers in the upper right corners are the expected values for the respective worlds.

distribution s and a causal relation r. Independently of the causal relation from which s originates, the choice probabilities of a speaker who aims to communicate her beliefs about the world can then be written as:

$$P_S(u \mid r, s) = P_S(u \mid s) \propto \exp(\alpha \cdot (\log P_{lit}(s \mid u))) \qquad [10]$$
$$\text{where } P_{lit}(s \mid u) = \sum_{r'} P_{lit}(r', s \mid u)$$

The speaker's goal of communication, when using conditional sentences, that we assume in this paper, is first and foremost to convey their beliefs about the antecedent and the consequent. In other words, we start by exploring a probabilistic model of communication with conditionals from the most austere assumption, namely that not only the assertability of a conditional does *not* hinge on any putative causal relation necessarily, but that also the purpose of communication itself is *not* to directly communicate information about the causal relation.[11]

Before we discuss the results of our simulations, we would like to add a final note on the interpretation of the world states in our model. Instead of considering them as simple pairs consisting of a probability distribution and a causal relation, we can also think of them as *causal Bayesian networks* (Pearl, 1988, 2009), henceforth abbreviated as 'Bayes nets'. They have, similar to the closely related formalism of Structural Causal Models, a rich tradition of supporting semantic theories of conditionals already, (e.g. Briggs, 2012; Hiddleston, 2005; Kaufmann, 2013; Lassiter, 2017; Lucas & Kemp, 2015; Pearl, 2009, 2013; Rips, 2010; Santorio, 2019). Bayes nets represent sets of variables (in our case binary variables) and the dependencies among them; they consist of a

11 There are circumstances where the goal of the communication reasonably includes the causal relation which we leave for future work here. What seems to be clearly wrong, is to restrict the speaker's goal only to the communication of the causal relation; this is rather attributed to explicit causal language instead of conditionals.

directed, acyclic graph that defines how the variables relate to each other, and a set of conditional probabilities of each variable given all possible instantiations of all its *direct* parent nodes, which simplify to unconditional probabilities for variables without parent nodes. This set of (conditional) probabilities is sufficient to define the joint probability distribution over all variables represented in the graph; see Figure 2 for a Bayes net that represents the joint probability distribution s_2 from Table 4.

In sum, the model we explore here can be seen as capturing the implicit communication of causal information by (i) treating world states as causal Bayes nets, (ii) identifying the purpose of utterance (the *question under discussion* (Roberts, 2012) or the relevance projection (Kao, Wu, et al., 2014)) to be the precise communication of the probability table s (the speaker's beliefs about joint truth of A and C) and (iii) a pragmatic process of utterance generation favoring true and informative utterances.

4

MODELING CONDITIONALS IN DEFAULT CONTEXTS

In this chapter, we explore model predictions in default contexts. We can think of these as the model's predictions generalized over a wide range of more specific contexts, or, relatedly, as the model's predictions for the interpretation of utterances of conditionals in unbiased, out-of-the-blue contexts. We will particularly look at the listener's inferences from an utterance of a conditional about the speaker's uncertainty about A and C, about the speaker's beliefs about any systematic relation between A and C, and about the strength of a conditional perfection reading. Doing so, we show how informativity-driven pragmatic choice of utterances leads to *de facto* assertability conditions as postulated by inferentialist accounts without having to stipulate these directly.[1]

SETTING THE SCENE. Let us reconsider example (34), repeated here as (35), but now uttered in an out-of-the-blue context.

(35) Bigi: If Alex comes to the party, Chris comes too. ["If A, C."]
Wobo: Who are these guys? What party are you talking about?

Even without any strong prior convictions, Wobo is likely to draw pragmatic inferences from Bigi's utterance. Intuitively, Wobo would infer that Bigi is uncertain about whether Alex comes to the party, similarly for Chris, and that Chris' coming to the party is not entirely (causally) unrelated to Alex's coming somehow. Moreover, the implicit dependency between antecedent and consequent may be interpreted to be so strong that the conditional is understood as biconditional; it does not seem unnatural for Wobo to infer that, according to Bigi, Chris comes to the party *only if* Alex comes.

To analyse the listener's *a posteriori* beliefs and the speaker's utterance choices we make use of the following definitions concerning the modeled states. The set of states in which the speaker is *uncertain about event* X contains all and only states s such that the probability of X in s, $P^{(s)}(X)$, is neither too low or too high: (analogous for uncertainty about X given Y) is:

$$\text{Uncertain}(X) = \{s \mid 1 - \theta \leq P^{(s)}(X) \leq \theta\} \qquad [11]$$

1 All results reported below were obtained with a threshold for the literal meaning, θ, set to 0.9, and the rationality parameter α set to 3. Qualitatively identical results have been obtained for a number of different parameter values, namely a grid with $\theta \in [0.9, 0.95, 0.975]$, $\alpha \in [1, 3, 5, 10]$.

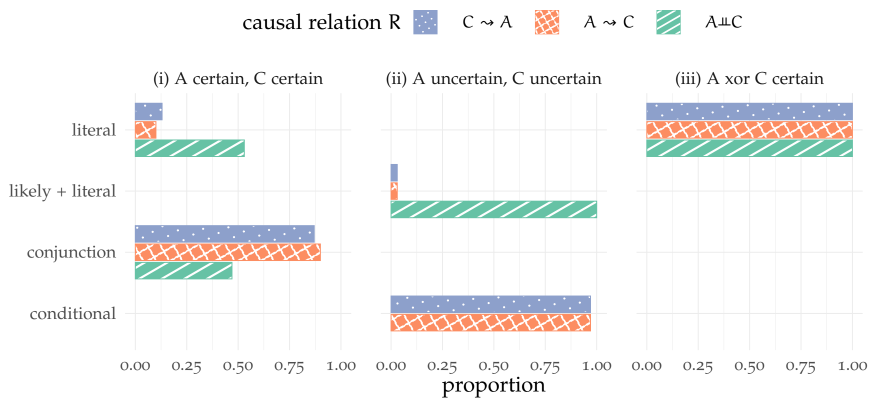


Figure 7: Relative frequency of how often each utterance type is the speaker's *best* choice for a set S of 10,000 states sampled from the prior (default context), given that the speaker is (i) certain or (ii) uncertain about A and about C, i.e., $\forall s \in S : s \in \text{Certain}(A) \wedge s \in \text{Certain}(C)$, respectively $\forall s \in S : s \in \text{Uncertain}(A) \wedge s \in \text{Uncertain}(C)$, or (iii) the speaker is uncertain about the truth of one proposition but certain about the truth of the other, e.g., $\forall s \in S : s \in \text{Uncertain}(A) \wedge s \in \text{Certain}(C)$.

A similar construction captures the speaker's *certainty about whether* X *is true* (analogous for X given Y):

$$\text{Certain}(X) = \{s \mid P^{(s)}(X) > \theta\} \cup \{s \mid P^{(s)}(X) < 1 - \theta\} \qquad [12]$$

4.1 HYPERRATIONAL UTTERANCE CHOICES IN DEFAULT CONTEXTS

According to the model presented here, the listener's inferences about the speaker's epistemic state can be put down, at least in part, to a Q-implicature (Atlas & Levinson, 1981; Horn, 1984): from the fact that the speaker decided *not* to utter a more specific and thus more informative utterance than $A \to C$, the pragmatic listener should deem it unlikely that the speaker refers to a state s in which a more informative utterance would also apply. To see whether such an alternatives-based explanation is endorsed by the present modeling setup, we look first at a speaker who always chooses the utterance with the highest utility (corresponding to a hyperrational speaker where rationality parameter $\alpha \to \infty$). Figure 7 shows model predictions from the speaker's point of view for different probabilistic beliefs with respect to the propositions A and C given the causal relation between the antecedent and the consequent. It clearly shows that a (hyperrational) speaker chooses utterances in dependence of her belief state: when the speaker is uncertain about both propositions, the best utterance is either a conditional or "*likely* Φ" which are the two least informative utterance types.[2] On the other hand, the speaker's best utterance is either a conjunction or a literal when she is certain about both propositions, that is, when she is certain that A is true (or false) and certain that C is true (or false), but she is not necessarily certain that a conjunctive event occurs (e.g. $A = a \wedge C = \neg c$). This is also the reason, why conjunctions do not seem to be preferred over literals when the speaker is certain about A and certain about C, and A and C are independent (left panel of Figure 7). In this case, when the speaker's best utterance is a literal, there simply is no conjunction that truthfully describes the given state (e.g., $P^{(s)}(a, c) = 0.82, P^{(s)}(a, \neg c) = 0.08, P^{(s)}(\neg a, c) = 0.08, P^{(s)}(\neg a, \neg c) = 0.02$). When the speaker is only uncertain about one proposition but not the other, her best utterance is a literal (right panel of Figure 7).

Figure 7 further reveals that the speaker's beliefs about the causal relation among A and C is also a highly influential factor for the speaker's utterance choice when the speaker is uncertain about both A and C (middle panel). For these states the speaker's best utterance will always be "*likely* Φ" when A and C are independent, whereas when A and C are dependent, the speaker's best utterance will almost certainly be a conditional.

4.2 INFERENCES ABOUT CAUSAL DEPENDENCY

Since pragmatic interpretation is here modeled as backwards-inference based on the speaker's utterance choice protocol, we can already anticipate from the above results of hyperrational speakers that pragmatic interpreters may draw rather specific inferences about the causal relation between A and C from an utterance of $A \to C$. In what follows, we look at inferences about the causal relationship in more detail, assuming "normal speakers" ($\alpha = 3$).

Figure 8 shows the causal inferences that listeners draw about A and C when the speaker utters the conditional $A \to C$. We observe that *a posteriori* the listener (literal and pragmatic) assigns very low probability to states where $r = A \overset{-+}{\rightsquigarrow} C$ or $r = C \overset{-+}{\rightsquigarrow} A$, as these are very unlikely to give rise to a probability table in which the conditional $A \to C$ is assertable. Interestingly, the listener is not committed to a single underlying causal relation, but instead merely infers that there is a positive dependency between the antecedent and the consequent: A tends to be true/false when C is true/false and vice versa. The pragmatic listener assigns almost the entire probability mass to the corresponding causal relations ($A \overset{++}{\rightsquigarrow} C$, $C \overset{++}{\rightsquigarrow} A$), the literal listener approximately 75%.

2 For simplicity, we discuss all results assuming an epistemic interpretation of the probabilities in our world states, but an interpretation based on objective chance is equally applicable.

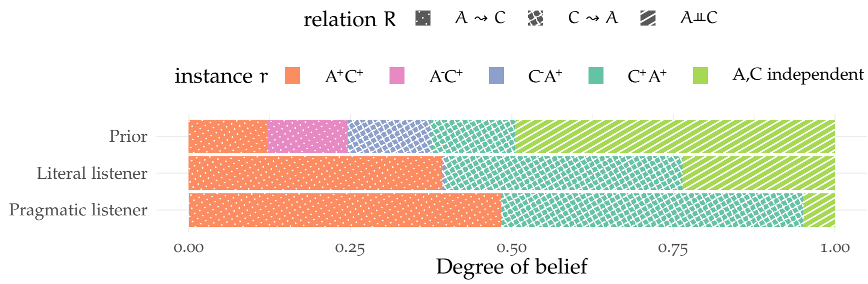


Figure 8: Degree of belief in each causal relation (A^+C^+ is shorthand for $A \overset{++}{\rightsquigarrow} C$, analogous for other relations) at all three levels of interpretation; prior to uptake of the conditional $A \rightarrow C$ and *a posteriori* given a literal or pragmatic listener.

This result suggests that also under a pragmatic interpretation, the listener needs further knowledge to disambiguate the underlying causal structure since, in this most general context, it is not possible to infer that the antecedent is a cause of the consequent or that, as under a diagnostic reading of the conditional $A \rightarrow C$, the consequent is a cause of the antecedent.

Another interesting result concerns the states where A and C are independent: even though the literal listener largely diminishes her beliefs in the independence of A and C as compared to her beliefs prior to the speaker's utterance of the conditional $A \rightarrow C$, the pragmatic listener assigns a still smaller probability to states where antecedent and consequent are causally independent. This is an important and interesting result that bears emphasis. Since the pragmatic listener combines a rich representation of the speaker's beliefs about truth of propositions and their causal relation with Gricean pragmatic reasoning, the pragmatic listener concludes more about the (speaker's beliefs about the) causal structure of the world than is entailed by the semantics of a conditional. This additional causal-pragmatic inference essentially rides piggyback on standard Gricean Quantity reasoning.

To see this, let us consider the perspective of the hyperrational speaker again (Figure 7). Those states where A and C are independent are exclusively states for which an informative Gricean speaker will either prefer to utter a bare proposition (conjunction or literal) which is more informative than a conditional or "*likely* Φ". As a result, the pragmatic listener infers a causal relation from entirely unbiased assertability conditions for conditionals and standard Gricean Quantity reasoning. Causal inference comes up as a pragmatic inference without having to stipulate an additional pragmatic constraint concerning a (causal) relation between propositions, let alone hard-coding such a requirement in the semantic meaning.

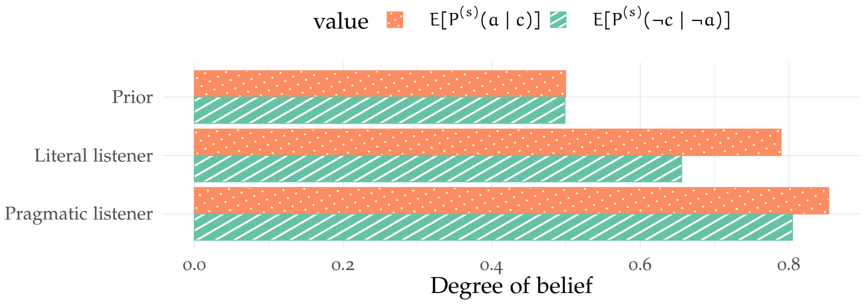


Figure 9: Degree of belief in the two CP-related conditional probabilities $P^{(s)}(\neg c \mid \neg a)$ and $P^{(s)}(a \mid c)$ at all three levels of interpretation; prior to the uptake of the conditional $A \to C$ and *a posteriori*, given a literal or pragmatic listener.

4.3 THE STRENGTH OF CONDITIONAL PERFECTION READINGS

In the example given in the beginning of this section, it does not seem unexpected or astounding when the utterance "*If Alex comes to the party, Chris comes too*" is interpreted as 'Chris will come to the party *if and only if* Alex comes to the party'. It might even seem to be a quite acceptable, if not natural, inference although we did not specify any further context and although, from a logical point of view, this inference is *not* valid.

The phenomenon that conditionals are sometimes interpreted as biconditionals, known as *conditional perfection* (CP), has caught much attention in the literature, especially since Geis and Zwicky (1971). CP remains a topic of ongoing debate; no consensus has for instance been found with respect to its prevalence or how and under which circumstances a CP reading is triggered (e.g. see Moldovan, 2013; Newstead, 1997; Oberauer & Wilhelm, 2003; von Fintel, 2001).

No general standard measure has been established that quantifies the degree to which a conditional receives a CP reading. Here we will refer to two inferences that are prominently considered in the literature on conditional reasoning, '*Denying the antecedent*' (DA) and '*Affirming the consequent*' (AC), shown below. The endorsement of DA or AC inferences suggests that participants interpret the conditional as biconditional since only then these are logically valid inferences (e.g., Evans et al., 2007; Newstead, 1997).

DA: $A \to C.\ \neg A.\ \therefore\ \neg C.$

AC: $A \to C.\ C.\ \therefore\ A.$

To learn how strongly $A \to C$ is interpreted as biconditional, we will therefore look at the listener's expected beliefs (about the speaker's beliefs) in $P^{(s)}(\neg c \mid \neg a)$ and $P^{(s)}(a \mid c)$. Only when "*if*" is interpreted as biconditional (i.e.,"*if and only if*") the two considered quantities should be large.

As can be seen in Figure 9, without any further contextual assumptions our model predicts a quite strong CP-reading for the interpretation of conditionals in the default context. The speaker's utterance of $A \rightarrow C$ elicits an increase in the listener's beliefs (about the speaker's beliefs) in the conditional probabilities $P^{(s)}(\neg c \mid \neg a)$ and $P^{(s)}(a \mid c)$ as compared to her prior beliefs. This is true for the literal and the pragmatic interpretation, yet the pattern is more pronounced in the latter. This result is to a great extent due to the representation of dependent world states as noisy-or models (see Section 3.2.4). In this way, this interpretation is explained as something akin to an I-implicature (Atlas & Levinson, 1981; Levinson, 2000), as suggested by Horn (2000); CP-readings are supported in large part by what is arguably a cognitively economic, perhaps stereotypical representation format of causal dependency between events. However, we also see that pragmatic reasoning about alternatives further strengthens the CP-reading quantitatively, similar to the accounts of van der Auwera (1997) and von Fintel (2001) which share with our model the feature that CP is a result of a listener reasoning about the speaker's production protocol — given von Fintel's (2001) account, including, crucially, the possibility that the speaker could simply have asserted the consequent.

4.4 DERIVING INFERENTIALIST ASSERTABILITY CONDITIONS

So far, we showed that the current modeling setup predicts that speakers will use conditional sentences predominantly in cases where there is a (causal/inferential) relationship between the antecedent and the consequent and that listeners, therefore, infer such a relationship from an utterance of a conditional.[3] These predictions particularly challenge the idea advanced by advocates of Inferentialism that a (causal/inferential) relation between antecedent and consequent is part, in whatever form, of the core semantics of conditionals. We here argue that an austere assertability condition for conditionals in combination with rich representations of contextual (causal) world knowledge and pragmatic reasoning is sufficient to derive the kind of assertability conditions postulated by Inferentialists' accounts. In particular, we will consider the assertability condition formulated in Equation [13] which

3 Our model predicts the dependency relation to be a defeasible inference, but it is not predicted to arise in any context; some contexts may not provoke this inference (e.g., see the discussion on missing-link and biscuit conditionals in Section 5.4). Therefore, it is also compatible with Lassiter's (2022) account of when the dependency between antecedent and consequent will arise (and when it will not) which is based on discourse coherence.

was proposed by van Rooij and Schulz (2019) improving on like-minded work of Douven, 2008.[4]

$$A \rightarrow C \text{ is acceptable/assertable only if } \Delta^{*}P = \frac{P(c \mid a) - P(c \mid \neg a)}{1 - P(c \mid \neg a)} \geq \theta. \tag{13}$$

The general idea behind assertability criteria of this kind is that an utterance of a conditional $A \rightarrow C$ is acceptable only if $P(c \mid a)$ is high (like we assume as well here) and, in addition, $P(c \mid \neg a)$ is low, that is, C being true is only likely when A is true.

We do not argue that Equation [13] necessarily holds, empirical research is needed to find out whether the acceptance/assertability of conditionals is related to such criteria.[5] Instead, we aim to investigate how the model that we propose here relates to accounts that propose this kind of assertability conditions for conditionals, possibly bringing Inferentialist's ideas and pragmatic reasoning closer together.

We find that our pragmatic model for the use of conditionals derives the criterion from Equation [13], in the sense that whenever the model predicts a (hyperrational) speaker to use a conditional in some state s, the value $\Delta^{*}P$ calculated from the probability table entailed by s is indeed very high. To see this, Figure 10 shows the distribution of $\Delta^{*}P$ measures obtained for three sets of sampled states: (i) samples from the prior (default contexts), (ii) samples from the prior conditioned on $A \rightarrow C$ being assertable (literal speaker), and (iii) the subset of states sampled in the literal speaker condition, in which a hyperrational speaker would utter $A \rightarrow C$, that is, states where $A \rightarrow C$ is the preferred choice of a pragmatic speaker. Figure 10 reveals, reassuringly, that $\Delta^{*}P$ is not always high for any state sampled from the prior. It also shows that just from our austere assertability condition alone, namely $P^{(s)}(c \mid a) \geq \theta$, the average associated $\Delta^{*}P$ increases, even if there are still quite a number of cases where what we may call a "literal speaker" might use $A \rightarrow C$ while the measure $\Delta^{*}P$ is quite low. This clearly shows that, despite biases for "simple situations" introduced by *noisy-or* parameterization of state priors, the assertability condition that $P^{(s)}(c \mid a) \geq \theta$ alone does not guarantee that $\Delta^{*}P$ is high. But for the pragmatic speaker, we see that the $\Delta^{*}P$ measure is exclusively very high, thus lending support to the idea that a straight-

4 As previously mentioned in Section 2.1.2.2, note that recently van Rooij and Schulz (2022) showed that this condition can be derived based on the conversational implicature that the speaker is not in a position to utter the consequent straight away, assuming, like we do here, that for a conditional to be assertable, the corresponding conditional probability must be reasonably large.

5 While some of the empirical studies that have been conducted to date suggest that the relationship between antecedent and consequent has an influence on the assertability of conditionals (e.g.,see Douven & Verbrugge, 2012; Skovgaard-Olsen, 2016), others found no such effect (e.g., see Oberauer et al., 2007; Singmann et al., 2014).

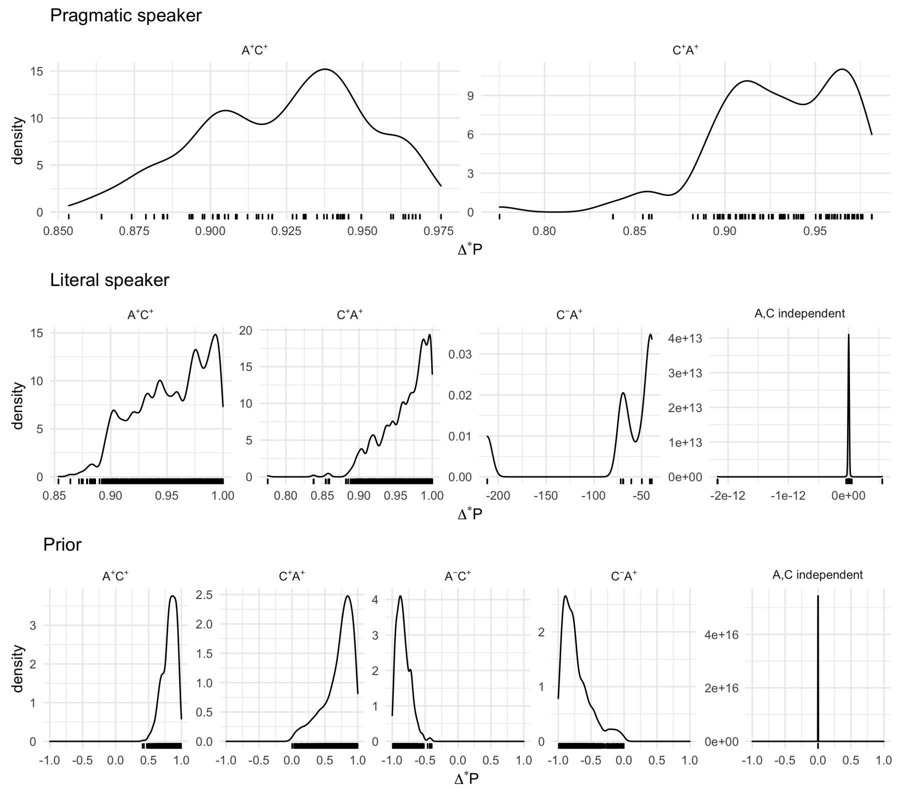


Figure 10: Distribution of $\Delta^* P$ values for 10,000 states randomly sampled from (i) the prior (bottom), (ii) from the prior given $A \to C$ is assertable (middle), and (iii) for states from (ii) where additionally $A \to C$ is the pragmatic speaker's best choice (top). Causal relations are abbreviated, e.g., A^+C^+ is shorthand for $A \overset{++}{\rightsquigarrow} C$.

forward model of Gricean pragmatic use of conditionals explains an otherwise stipulative assertability condition for conditionals.

We saw that when $A \to C$ is asserted by our pragmatic speaker, $\Delta^* P$ is large, let us also briefly consider this from the other way around: does our pragmatic speaker always prefer to utter the conditional $A \to C$ as soon as $\Delta^* P$ is large? Put differently, is a large value of $\Delta^* P$ sufficient for our pragmatic speaker to assert $A \to C$? To answer this question, we will look at the pragmatic speaker's utterance choice when $\Delta^* P$ is large. If it was a sufficient condition, the pragmatic speaker should always choose $A \to C$ in these cases. Figure 11 shows boxplots in the range $0 \leq \Delta^* P \leq 1$ of the $\Delta^* P$-values of states where $A \to C$ is assertable, but *not* the most likely utterance for a hyperrational speaker.[6] The utterance type of the hyperrational speaker's best choice for the respective states — a literal, a conjunction or a conditional other than $A \to C$ — is shown on the x-axis; color codes represent the causal relation of the states. Clearly,

6 For states where $R = C \rightsquigarrow A$, $\Delta^* P$ has a minimum value of -212 when the hyperrational speaker's best utterance is a literal, and a minimum value of -69.3 when it is a conjunction. When $R = A \perp\!\!\!\perp C$, $\Delta^* P^{(s)}$ clusters closely around 0.

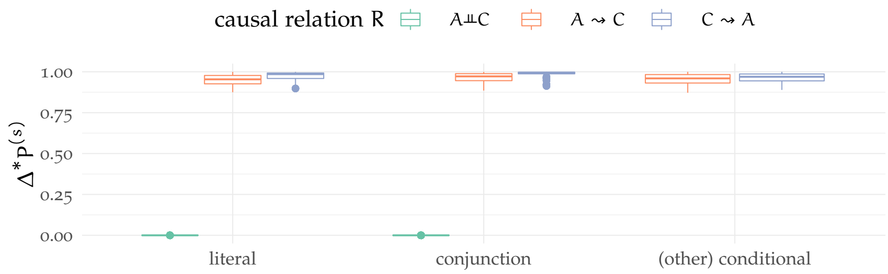


Figure 11: $\Delta^*P^{(s)}$ values, zoomed into the range where $0 \leq \Delta^*P^{(s)} \leq 1$ for s from a set of 10,000 states sampled from the prior given $A \rightarrow C$ is assertable (literal speaker condition in Figure 10) where $A \rightarrow C$ is *not* the best choice of a hyperrational speaker ($\alpha = \infty$).

there are states where $\Delta^*P^{(s)}$ is large and the speaker chooses a different utterance than $A \rightarrow C$. Particularly interesting for us are those states where the speaker does not choose a different conditional but an utterance that is more informative than a conditional, that is, a conjunction or a literal. These are situations where the predictions of our model diverge from the predictions of accounts arguing that a conditional is a assertable/acceptable when Δ^*P is large: although $A \rightarrow C$ is (literally) assertable, the speaker may have good reasons to choose a different utterance (to be maximally informative) and so, the conditional $A \rightarrow C$ might still be rejected. This explains why a criterion like the one defined in Equation [13] might be a necessary, but not a sufficient condition for the assertability of a conditional.

4.5 SUMMARY

In this chapter, we considered the predictions of our RSA-model for the communication with conditionals in what we call a default context. In this default context, we represent dependent variables as noisy-or models with large causal power of the cause variable to provoke the effect which may yet be provoked (with little probability) by alternative causes that are not modeled further. Otherwise, there is no specific world knowledge assumed to be available.

We showed that this setup together with a basic literal semantics for conditionals, based on the relevant conditional probability, makes a hyperrational pragmatic speaker (who always selects the best, i.e., most informative utterance) choose different utterances contingent on the relation between both variables and the speaker's confidence in, or better to say uncertainty of, their truth/falsity. Namely in a way that the pragmatic listener tends to interpret $A \rightarrow C$ according to a CP-reading ('if' interpreted as 'iff') and is able to infer from the speaker's utterance of a conditional that most likely antecedent and consequent are dependent, yet leaving open which of both propositions corresponds to the cause and which to the effect.

That is, the predictions of our model corroborate the hypothesis that the dependency relation does not need to be incorporated into the semantics of conditionals, but can be derived through pragmatic reasoning, given an appropriate representation of the modeled variables, in particular their underlying causal structure. We further showed that the pragmatic listener's interpretation of the conditional $A \rightarrow C$ satisfies a condition based on a measure called relative difference ($\Delta^* P$, see Equation [13]), that had been proposed as assertability condition for conditionals (van Rooij & Schulz, 2019) to ensure the presence of a dependency relation between antecedent and consequent. Moreover, our model explains why $\Delta^* P \geqslant \theta$ is not a sufficient condition for a speaker to assert $A \rightarrow C$: in many situations where $\Delta^* P \geqslant \theta$ is fulfilled, the speaker has the possibility to choose a more informative utterance than $A \rightarrow C$.

5 MODELING CONDITIONALS IN CONCRETE CONTEXTS

So far we showed that an RSA model combined with unbiased default priors is able to explain general pragmatic inferences associated with utterances of conditionals. In the following we will turn towards predictions for utterances of conditionals in rather specific and explicitly given contexts of use. In particular, we investigate how the present setup helps explain a puzzle put forward concisely by Douven (2012). Douven contrasts three cases of conditionals uttered in very concrete contexts, which were shortly introduced in Section 2.1.3 and are, for convenience, fully cited cited below. In these contexts, learning a conditional $A \rightarrow C$ either leads to an increase (the Skiing Example given in (36), see Section 5.1) or a decrease (the Garden Party Example given in (37), see Section 5.2) in the listener's degree of belief in the truth of A or the listener's degree of belief in the antecedent A does not change at all (the Sundowners Example given in (38), see Section 5.3).

We argue here that an explicit representation of world knowledge that is contextually-grounded and does not only comprise plausible prior beliefs about the probability of the represented variables, but also caters for their causal structure, is sufficient to explain how pragmatic listeners adjust their beliefs about the antecedent in one way or another after receiving information in form of a conditional. The key to the explanation we propose here is that, in each case, the listener learns a piece of causal information, that is, the listener learns that propositions are causally related where this was previously not expected or deemed rather unlikely.[1]

5.1 THE SKIING EXAMPLE

The Skiing Example (Douven, 2012) is a case where, intuitively, the listener's degree of belief in the antecedent increases.

> THE SKIING EXAMPLE. Harry sees his friend Sue buying a skiing outfit. This surprises him a bit, because he did not

1 A perhaps more realistic picture would be to model a listener as completely *unaware* of the causal relation in question. An agent unaware of a contingency holds no (explicit) belief about that contingency; it is just not represented in that agent's stock of explicitly entertained alternatives. Since adding agent's unawareness to a model of pragmatic reasoning is possible but technically rather involved (e.g. Franke, 2014; Franke & de Jager, 2011; Heifetz et al., 2006), we here make the simplifying assumption that the listener is aware of the possible causal connection but just deems it very unlikely to begin with.

know of any plans of hers to go on a skiing trip. He knows that she recently had an exam and thinks it unlikely that she passed. Then he meets Tom, another friend of Sue's, who is just on his way to Sue to hear whether she passed the exam, and who tells him:

(36) If Sue passed the exam, her father will take her on a skiing vacation.
$\rightsquigarrow$ listener belief in antecedent *increases*

In this example, there are three relevant propositions: E (Sue passed the exam), S (Sue goes skiing) and C (Sue buys skiing clothes). The listener Harry has observed C, so his beliefs in C are high, possibly 1. The speaker Tom utters the conditional $E \rightarrow S$. We want to explain how this can lead to an increase in Harry's probabilistic beliefs about E.

Our explanation hinges on assuming that Harry has certain plausible beliefs about the propositions involved and their causal relationship. It proceeds in three steps:

1. From pragmatic reasoning, the listener infers from the utterance of $E \rightarrow S$ that the speaker likely believes in a causal relation $E \overset{++}{\rightsquigarrow} S$, i.e., that passing the exam increases the chance of Sue going on a skiing trip.

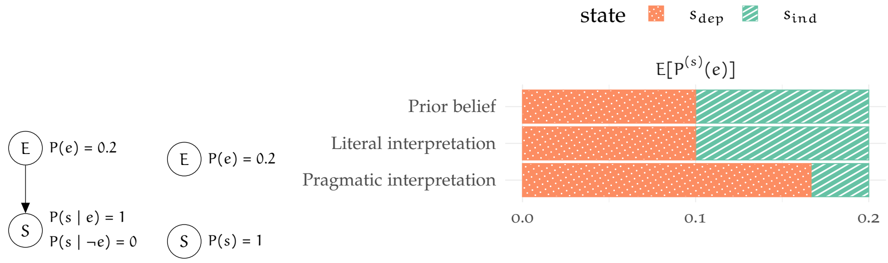
2. The listener takes the speaker to be an authority on the matter and, at least to a certain extent, also increases degrees of beliefs in the causal relation $E \overset{++}{\rightsquigarrow} S$.

3. Since the listener also has a high degree of belief in C, and given that it is plausible to assume that in general a relation of the kind $S \overset{++}{\rightsquigarrow} C$ holds, the listener ends up with a higher degree of belief in E after processing the utterance than before.

To illustrate this reasoning schema, we offer spell out one concrete context model for the Skiing Example as in Figure 12.[2] There are two Bayes nets that are in line with the speaker's utterance $E \rightarrow S$, one in which passing the exam stands in a direct causal relation to going on a skiing trip (Figure 12a), and one in which it does not (Figure 12b).[3]

2 A more realistic choice than using a single independent state with $P^{(s_{ind})}(s) > \theta$, would be to include several independent states, e.g. with $P^{(s)}(s) \in [0, 0.5, 1]$ which would, however, require a larger set of utterances such that for every state there is at least one utterance assertable. For the sake of a simpler discussion, here and in the following examples, we give just one set of concrete numbers of the relevant (conditional) probabilities. More realistically, we should assume that the numbers represented here are (something like) the listener's expected values, given uncertainty that weighs an infinity of possible values.

3 Even though it seems less probable, there is the possibility that the causal relation between E and S is reversed (i.e., $S \rightarrow E$ instead of $E \rightarrow S$); Sue may for instance study extra hard *because* her father invites her to go on a skiing trip. Note that in



(a) s_{dep}: E, S dependent (b) s_{ind}: E, S independent (c) Listener's degree of belief in the antecedent

Figure 12: Bayes nets and results for the pragmatic reasoning part in the Skiing Example with $\mathcal{U} = \{S, \text{likely } S, E \to S\}$, $\alpha = 1$, E: pass exam, S: go skiing. Both states, s_{dep} and s_{ind}, are assigned equal prior probability.

	s	¬s
e	$P(e) \cdot P(s \mid e) = 0.2$	$P(e) \cdot P(\neg s \mid e) = 0$
¬e	$P(\neg e) \cdot P(s \mid \neg e) = 0$	$P(\neg e) \cdot P(\neg s \mid \neg e) = 0.8$

(a) $P^{(s_{dep})}(E, S)$

	s	¬s
e	$P(e) \cdot P(s) = 0.182$	$P(e) \cdot P(\neg s) = 0.018$
¬e	$P(\neg e) \cdot P(s) = 0.728$	$P(\neg e) \cdot P(\neg s) = 0.072$

(b) $P^{(s_{ind})}(E, S)$

Table 7: Entailed joint probability distributions of variables E, S of the two Bayes net s_{ind}, s_{dep} shown in Figure 12; E: 'pass exam', S: 'go on skiing trip'.

Their entailed joint probability distributions are spelled out in Table 7. The situation model in Figure 12 gives the listener's (Harry's) beliefs about what Tom might believe about the relation between E and S. Given the context-story, there is no indication that Harry believes that Tom believes that the dependent Bayes net is any more or less likely than the independent one. Therefore, we assign equal prior probability to both Bayes nets, even though this is not crucial for the case at hand.

With the set of alternative utterances $\mathcal{U} = \{S, \text{likely } S, E \to S\}$, $\alpha = 1$ and the probabilities and causal relations as shown in Figure 12, we get the following probability distribution for the pragmatic listener when $u = E \to S$: $P_{PL}(s = s_{dep} \mid u = E \to S) = 5/6$ and $P_{PL}(s = s_{ind} \mid u = E \to S) = 1/6$; see Table 8 for the respective values of other model components. The dependent Bayes net becomes more likely under a pragmatic interpretation only: by taking into account the fact that the speaker could have chosen a more informative utterance (e.g. "Sue goes on a skiing trip" (S)) to refer to the independent Bayes net (see Table 8), the listener learns that, most likely, the speaker believes in a connection between Sue going on a skiing trip and her passing the exam. Contrary to that, under a literal interpretation both Bayes nets

this case, the listener's observation of Sue buying skiing clothes would, however, not increase the listener's degree of belief in Sue passing the exam.

	Literal meaning $\delta_{s\in[\![u]\!]}$			Speaker $P_S(u \mid s)$		
u	$E \to S$	S	likely S	$E \to S$	S	likely S
s_{dep}	1	0	0	1	0	0
s_{ind}	1	1	1	$1/5$	$2/5$	$2/5$

	Literal listener $P_{LL}(s \mid u)$			Pragmatic listener $P_{PL}(s \mid u)$		
u	$E \to S$	S	likely S	$E \to S$	S	likely S
s_{dep}	$1/2$	0	0	$5/6$	0	0
s_{ind}	$1/2$	1	1	$1/6$	1	1

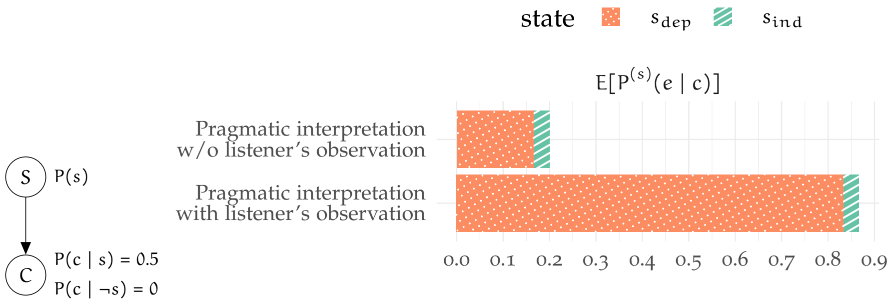
Table 8: Distributions for the Skiing Example with two Bayes nets s_{dep}, s_{ind}, utterances $\mathcal{U} = \{E \to S, S, \text{likely } S\}$ and $\alpha = 1$.

remain equally likely as the chosen utterance, $E \to S$, is literally true in both states.

The predicted interpretation of Tom's utterance, $E \to S$, with respect to the probability of the antecedent is shown in Figure 12c: the listener's degree of belief about the speaker's beliefs in the antecedent remains at an expected value of 0.2, only the degree of belief related to the causal relation between E and S is influenced by the speaker's utterance of the conditional. Crucially, since we so far only modeled the listener's pragmatic inferences about the speaker's beliefs, the listener's observation ($C = c$) is not considered yet.

How should a listener change their beliefs in the light of a (pragmatically derived) belief about the speaker's beliefs? That depends on the more general assumptions the listener makes about the speaker: is she trustworthy in general, likely well-informed on the subject matter at hand? Since nothing in the scenario described in Example (36) gives us reason to expect otherwise, we may follow the usual assumptions in Gricean belief-based reasoning that the speaker is cooperative and knowledgeable (e.g. Geurts, 2010). Even if the precise effect of integrating beliefs of a cooperative and knowledgeable agent into one's own beliefs are elusive, the effect in a scenario like the one at hand is most likely that the listener increases their own beliefs in the relation $E \overset{++}{\rightsquigarrow} S$. To keep matters simple for a fully fleshed out numerical example, we just assume that the listener adopts exactly the same probabilistic beliefs as the inferred speaker beliefs.

This takes us to the last step of the explanation, where we look at the effect of the listener's private knowledge that C is the case. Put differently, we look at the listener's inference of the speaker's beliefs about Sue passing the exam under the assumption that the speaker makes the same observation as the listener. Similar to Douven, 2012, we draw on world knowledge about the relation between C and S,



(a) s_{wk}: S, C dependent

(b) Expected value of the listener's degree of belief in the antecedent

Figure 13: Bayes net for the assumed world knowledge in the Skiing Example (left) and results for the the listener's inference about the antecedent (right); E: pass exam, S: go skiing, C: buy skiing clothes.

where the former is highly unlikely when one does not go on a skiing trip.[4] This is reflected in the model by setting $P(c \mid \neg s)$ to 0 and $P(c \mid s)$ to 0.5 (see Figure 13a) as the context story does not provide any information concerning Sue's shopping behavior.

Based on this (world) knowledge about S and C (Bayes net s_{wk}), we can compute the listener's updated belief in the probability that Sue goes on a skiing trip given the listener's observation of Sue buying skiing clothes:

$$P^{s_{wk}}(s \mid c) = \frac{P^{s_{wk}}(c \mid s) \cdot P^{s_{wk}}(s)}{P^{s_{wk}}(c)} = \frac{0.5 \cdot P^{s_{wk}}(s)}{0.5 \cdot P^{s_{wk}}(s) + 0 \cdot P^{s_{wk}}(\neg s)} = 1 \tag{14}$$

Assuming the updated probability of Sue going on a skiing trip, $P^{s_{wk}}(s \mid c)$ given in Equation [14], the updated probability of Sue passing the exam ($E = e$) for the two modeled states s_{dep} and s_{ind} is given in Equations [15] and [16]:

$$\begin{aligned} P^{(s_{dep})}(e \mid c) = {} & P^{(s_{dep})}(s, e \mid c) + P^{(s_{dep})}(\neg s, e \mid c) \\ = {} & P^{(s_{dep})}(e \mid s) \cdot P^{(s_{wk})}(s \mid c) + \\ & P^{(s_{dep})}(e \mid \neg s) \cdot P^{(s_{wk})}(\neg s \mid c) \\ = {} & 1 \cdot 1 + 0 \cdot 0 = 1 \end{aligned} \tag{15}$$

4 We argue that the relation between C and S is essential for the desired interpretation of the conditional. Imagine that Harry, the listener, had no idea what skiing is. In this case, both states should be modeled without the link between S and C making Harry's observation irrelevant with respect to his belief about Sue passing the exam (E). Therefore, his belief in the antecedent would remain unchanged. See Günther (2018) who also makes use of causal Bayes nets but argues that the intuitive interpretation of the Skiing example is independent of whether or not the relation between variables S and C is modeled.

$$\begin{aligned} P^{(s_{ind})}(e \mid c) &= P^{(s_{ind})}(s, e \mid c) + P^{(s_{ind})}(\neg s, e \mid c) \\ &= P^{(s_{ind})}(e) \cdot P^{(s_{wk})}(s \mid c) + P^{(s_{ind})}(e) \cdot P^{(s_{wk})}(\neg s \mid c) \\ &= 0.2 \cdot 1 + 0.2 \cdot 0 = 0.2 \end{aligned} \tag{16}$$

These values emphasize the importance of taking into account the causal relations among variables: only for s_{dep} the degree of belief in $E = e$ is influenced by the listener's observation of c. Given s_{ind}, the probability of Sue going on a skiing trip remains the same as without the listener's observation. The listener's updated belief in the antecedent, given the listener's independent observation of Sue buying skiing clothes, is then equal to the expected value of $P^{(s)}(e \mid c)$ $(s \sim P_{PL}(s \mid u = E \to S), s \in [s_{dep}, s_{ind}])$ which is approximately 0.87, as spelled out in Equation [17]. Remember that $P_{PL(s|u=E\to S)}$ describes the listener's beliefs after the uptake of the conditional, that is, the listener is approximately 83% confident that the speaker's beliefs correspond to s_{dep} and approximately 17% confident that they correspond to s_{ind}. Figure 13b shows the predictions for the listener's posterior beliefs in the antecedent, with and without consideration of the listener's observation.

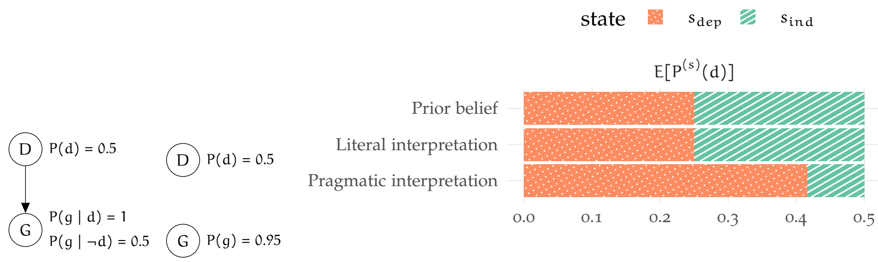
$$\begin{aligned} E[P^{(s)}(e \mid c)] = &\sum_{s_i \in \{s_{dep}, s_{ind}\}} P^{(s_i)}(e|c) \cdot P_{PL}(s_i \mid u = E \to S) = \\ &P^{(s_{dep})}(e \mid c) \cdot P_{PL}(s_{dep} \mid u = E \to S) + \\ &P^{(s_{ind})}(e \mid c) \cdot P_{PL}(s_{ind} \mid u = E \to S) = \\ &1 \cdot 5/6 + 0.2 \cdot 1/6 \approx 0.87 \end{aligned} \tag{17}$$

Conceptually, the expected value may be interpreted as integration of the listener's own observation with the information received from the speaker about the speaker's beliefs about the world that the listener takes over.

5.2 THE GARDEN PARTY EXAMPLE

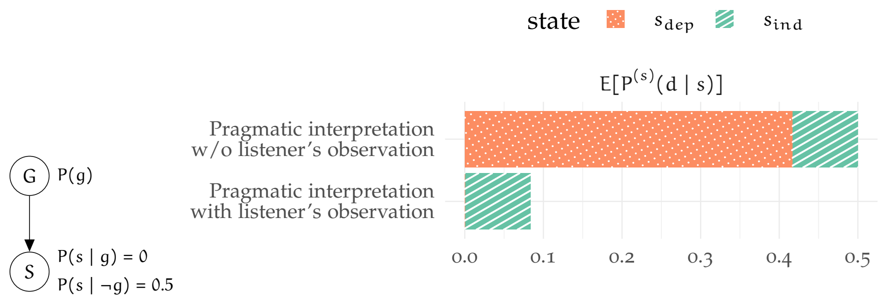
Despite its structural similarity to the Skiing Example, for completeness, let us also briefly consider the Garden Party Example (Douven, 2012) where, intuitively, the listener's degree of belief in the antecedent decreases.

> THE GARDEN PARTY EXAMPLE. Betty knows that Kevin, the son of her neighbors, was to take his driving test yesterday. She has no idea whether or not Kevin is a good driver; she deems it about as likely as not that Kevin passed the



(a) s_{dep}: D, G dependent (b) s_{ind}: D, G independent (c) Expected value of the listener's degree of belief in the antecedent

Figure 14: Bayes nets and results for Garden Party Example with $\mathcal{U} = \{G, D \rightarrow G, \text{likely } G\}$ and $\alpha = 3$, D: pass driving exam, G: throw garden party. Both states, s_{dep} and s_{ind}, are assigned equal prior probability.



(a) s_{wk}: S, G dependent (b) Expected value of the listener's degree of belief in the antecedent

Figure 15: Bayes net for the assumed world knowledge in the Garden Party Example (left) and results for the the listener's inference about the antecedent (right); D: pass driving exam, G: throw garden party, S: spade garden.

test. Betty notices that her neighbors have started to spade their garden. Then her mother, who is friends with Kevin's parents, calls her and tells her the following:

(37) If Kevin passed the driving test, his parents will throw a garden party.
$\rightsquigarrow$ listener belief in antecedent *decreases*

The difference between the Bayes nets shown in Figure 14 to the Bayes nets used in the Skiing example only lies in the intuitive instantiation of the prior probabilities, structurally they are analogous. The results for the interpretation of the conditional in the Garden Party Example (Figure 14c) also falls in with the results in the Skiing Example: only under a pragmatic interpretation, the listener increases her belief in the Bayes net where both variables are connected (s_{dep}), but the listener's belief in the probability of the antecedent remains as prior to the speaker's utterance.

Again, to consider the listener's own beliefs, we assume that the listener simply takes over the speaker's beliefs communicated by the utterance of the conditional. The world knowledge that we draw on in this example, shown in Figure 15a, concerns the fact that throwing a garden party is incompatible with spading the garden ($S = s$). Therefore, $P(s \mid g)$ is set to 0 and, due to the lack of more concrete information, $P(s \mid \neg g)$ is set to 0.5. Contrary to the Skiing Example, the listener's observation in this example is not evidence, but counterevidence for the antecedent. Betty observes her neighbors spading the garden, thus, the probability for a garden party decreases, $P^{(s_{wk})}(g \mid s) = 0$ (see Figure 15a). While in s_{ind}, the degree of belief in D is not influenced by the truth or falsity of G, $P^{(s_{dep})}(d \mid \neg g) = 0$ (as opposed to $P^{(s_{dep})}(d \mid g) = 2/3$). The combination of world knowledge about the connection between S and G, the speaker's utterance ($D \rightarrow G$), communicating a likely connection between D and G, and the listener's observation related to G therefore make Betty decrease her belief in the antecedent (Kevin passing the driving test), shown in Figure 15b.

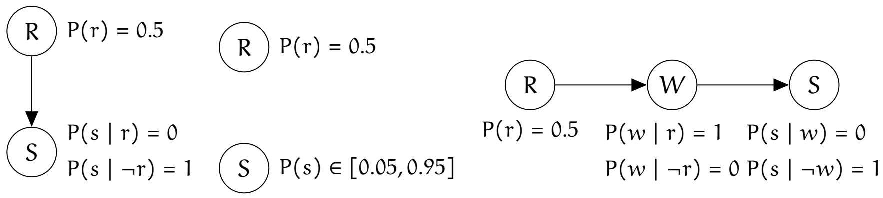
5.3 THE SUNDOWNERS EXAMPLE

The Sundowners Example (Douven & Romeijn, 2011) is a case where, intuitively, the listener's degree of belief in the antecedent does not change much, if at all.

> THE SUNDOWNERS EXAMPLE. Sarah and Marian have arranged to go for sundowners at the Westcliff hotel tomorrow. Sarah feels there is some chance that it will rain, but thinks they can always enjoy the view from inside. To make sure, Marian consults the staff at the Westcliff hotel and finds out that in the event of rain, the inside area will be occupied by a wedding party. So she tells Sarah:

(38) If it rains tomorrow, we cannot have sundowners at the Westcliff.
$\rightsquigarrow$ listener belief in antecedent remains *unchanged*

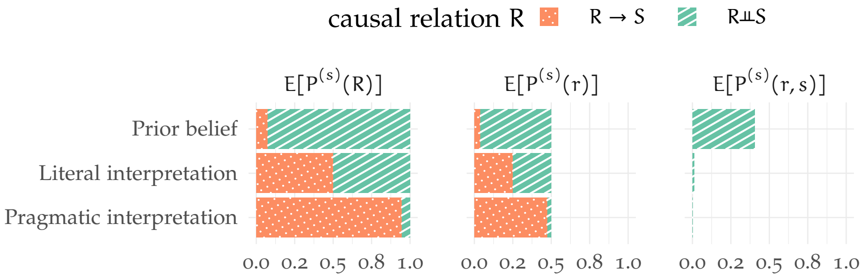
The intuition behind the formal treatment of the Sundowners Example, given below, is as follows. Even though the speaker's utterance of the conditional does not seem infelicitous, it seems less natural than the conditionals in the previous two examples. This oddness is reflected in a response from the listener that is not far to seek: why should rain prevent them from having sundowners? Put differently, we expect the listener, who has a strong prior belief in the independence of R and S, to be somewhat surprised by the speaker's utterance. The most rational explanation for the speaker's utterance choice of the conditional $R \rightarrow \neg S$ is to give up the assumption of indepen-



(a) $R \to S$; s_{dep}

(b) $R \perp\!\!\!\perp S$; $P^{(s_{ind\_low})}(s) = 0.05$, $P^{(s_{ind\_high})}(s) = 0.95$

(c) Example; R and S related via W



(d) Expected value of the listener's degree of belief in the causal relation between R and S (left), the probability of the antecedent (middle) and the joint event of antecedent and consequent (right), given the Bayes nets from Figure 16a and Figure 16b.

Figure 16: Bayes nets and results for Sundowners Example with $\mathcal{U} = \{R \to \neg S, \text{likely}S, \text{likely}\neg S, S, \neg S\}$, $\alpha = 3$ and a prior probability of 0.85 for Bayes net s_{ind_high}, which is *a priori* most likely according to the context story. The other two Bayes nets, s_{ind_low} and s_{dep}, are both assigned a prior probability of 0.075.

dence — at least if the integrity of the speaker is taken for granted. The listener's surprise may be resolved by accommodating a third, latent variable which is neither observed nor part of the speaker's utterance and acts as mediator between 'rain' and 'having sundowners'. This would justify the speaker's choice to utter the conditional $R \to \neg S$ which strongly suggests that (the speaker knows that) there is a (causal) connection between 'rain' and 'having sundowners'. In other words, we propose that, as opposed to the previous examples, the speaker's utterance in the Sundowners Example forces the listener to accommodate a mitigating variable, which forms a "causal bridge" between antecedent and consequent, so as to rationalize the speaker's utterance. As nothing further is known about this newly introduced "causal bridge" or any other relevant variable, the result is that the listener's beliefs about the variable 'rain' remains largely the same.

A formally explicit treatment of these ideas makes use of the Bayes nets shown in Figure 16a–16c. As for the Skiing and the Garden Party Example, the speaker's utterance of the conditional provokes a shift from the Bayes net where R and S are independent (Figure 16b) towards the Bayes net where they are causally related (Figure 16a). Yet, as can be seen in the leftmost panel of Figure 16d, in this example, the listener has a strong prior belief in the former which explains the listener's surprise resulting from the speaker's utterance choice. Assuming independence of R and S, the listener would rather expect the speaker to choose a more informative utterance assertable than the conditional (e.g. ,$\neg S$).

The RSA model of pragmatic language production entails a notion of *surprising utterances*. Given prior beliefs about states $P_{prior}(s)$ and the speaker's assumed production probabilities $P_S(u \mid s)$, the pragmatic listener expects utterance u with probability:

$$P_{PL}(u) = \sum_s P_{prior}(s)\, P_S(u \mid s)$$

If the pragmatic listener only considers the three states represented in Figures 16a and 16b, with a prior probability of 0.85 for the independent Bayes net in Figure 16b where $P(s) = 0.95$ and a prior probability of 0.075 respectively for the dependent Bayes net in Figure 16a, and the independent Bayes net in Figure 16b where $P(s) = 0.05$, the conditional $R \rightarrow \neg S$ is highly surprising in the sense that its expected occurrence probability is approximately 0.08 (for $\alpha = 3$).[5] This notion of listener surprise in the light of an unexpected utterances helps explain the intuition that an utterance of $R \rightarrow \neg S$ may trigger the desire to ask "why?" or to go look for an additional explanation which may rationalize the observed utterance.

One possibility of how the listener may rationalize a surprising utterance through a mitigating variable is shown in Figure 16c where W denotes the event of 'a wedding taking place inside the hotel' that represents any event which may possibly prevent the interlocutors from having sundowners at the hotel. Rather than denying the speaker's integrity due to the speaker's somewhat puzzling utterance, pragmatic reasoning, eminently causal-pragmatic reasoning, therefore allows the listener to infer a previously unexpected relation between R and S, that is able to explain the speaker's utterance. Nevertheless, the conditional still seems to be incomplete as one could expect the speaker to be more informative about the reasons why the event of rain may prevent them from having sundowners.

Contrary to the Skiing and the Garden Party Example, in the Sundowners Example, the listener does not intuitively update her belief in the probability of the antecedent. Figure 16d (middle) shows that the listener's predicted degree of belief in the antecedent remains at its prior value of 0.5 after processing the speaker's utterance, assuming a literal or a pragmatic interpretation. The crucial difference here is that the listener does not make any further observations as in the previous examples. Yet, the speaker's utterance of the conditional has a strong effect on the the listener's beliefs concerning the joint event

5 $P_S(u = R \rightarrow \neg S \mid s_{ind\_low}) \approx 0.06, P_S(u = R \rightarrow \neg S \mid s_{dep}) = 1$

of the antecedent and the consequent, shown in the right panel of Figure 16d. Contrary to her prior belief that $P(r, s)$ is almost a matter of chance, the listener judges it almost impossible that the two events both hold at the same time upon receiving the conditional information $R \to \neg S$: in the dependent Bayes net, R and S are mutually exclusive, thus $P(r, s) = 0$ and in the independent Bayes net where $R \to \neg S$ is assertable, $P(r, s)$ is close to 0.

In sum, the model is therefore able to account — by a single mechanism — for different interpretations with respect to the listener's degree of belief in the antecedent, namely by the interplay of pragmatic reasoning and an adequate, explicit representation of the interlocutors' prior probabilistic and in particular, *causal* beliefs.

5.4 SPECIAL CASES

Previous sections showed how the model of communication with conditionals presented here predicts that speakers use a conditional predominantly when there is a causal/inferential relation between antecedent and consequent and that therefore listeners will interpret conditionals accordingly. This brings up the obvious question as to how the approach advocated here positions itself with regard to cases, prominently discussed in the literature, in which antecedent and consequent are clearly *not* causally or inferentially related. Indeed, the absence of a relation between antecedent and consequent can either result in infelicity, as is the case in what we here call *missing-link conditionals* (Douven, 2017), or trigger an altogether different kind of interpretation, as in what we here address as *biscuit conditionals* (e.g., Austin, 1956; Geis & Lycan, 1993) and *concessive conditionals* () This section deals with each case in turn.

5.4.1 *Missing-link conditionals*

Missing-link conditionals have been used in support of the Inferentialist position that the requirement of a causal/inferential connection between antecedent and consequent is a necessary condition for assertability of a conditional, arguably situated in the semantics of conditionals because it is unclear how else a pragmatic account could explain the *in*felicity of missing-link conditionals (e.g. Douven, 2008, 2017; Krzyżanowska et al., 2014; Skovgaard-Olsen, 2016). This position is concretely exemplified by the contrast pair in (39), given by Douven (2008).

(39) There will be at least one heads in the first 1,000,000 tosses of this fair coin (h_{10^6}) *if*

a. there is a heads in the first ten tosses. (h_{10})

b. * Chelsea wins the Champions League. (c)

What is remarkable about the contrast between the two sentences in (39) is this: on the one hand, both (39a) and (39b) arguably pass the minimal necessary requirement for assertability, namely that the probability of the consequent given the antecedent is high — in fact the difference between the relevant conditional probabilities is minute ($P(h_{10^6} \mid h_{10}) = 1$, $P(h_{10^6} \mid c) \approx 1$); on the other hand, while (39a) is intuitively acceptable, for instance in a context where the speaker wants to highlight the entailment relation to a listener who might otherwise not attend to it sufficiently, (39b) rather clearly is odd.

To account for the infelicity of (39b) and assertability of (39a), inferentialism stipulates that a conceivable inferential relation between antecedent and consequent is a necessary requirement for assertability, howsoever the inferential link may exactly be defined. It is not limited to deductive inferences, as modern inferentialism allows less strict inferential relations such as induction or abduction (e.g. Douven, 2017; Krzyżanowska et al., 2014).[6] Inferentialists have criticized pragmatic explanations of the perceived infelicity of missing-link conditionals for remaining too vague about how the pragmatic processes may concretely look like (e.g., Douven 2017).

We argue here that (39b) is similar in kind to the Sundowners Example, yet more extreme, therefore leading to perceived infelicity. The Sundowners Example from the previous section showcases that there are contexts in which the utterance of a conditional is not questioned per se, but will neither be accepted without further ado. As there is no *obviously* conceivable connection between 'rain' and 'not having sundowners at the hotel', it seems quite natural for the listener to reply with a question asking for precisely this connection or by an anticipation of possible reasons, such as '*Why? Is there a private event taking place inside?*'. We consider infelicitous missing-link conditionals like (39b) to be similar in kind but more extreme cases of the same variety: in (39b) the speaker provides so little information that the listener does not have enough cues to make sense of the conditional utterance from world knowledge alone. In other words, we maintain that the infelicity of (39b) is a result of a failure of the listener to see a connection between antecedent and consequent that could rationalize the speaker's utterance choice.

To see how this is predicted by our model, we take the perspective of a listener, who tries to interpret a missing-link conditional and knows, from common sense world knowledge, that the antecedent

6 Several empirical studies (e.g., Krzyżanowska, Wenmackers, & Douven, 2013; Skovgaard-Olsen, 2016; Skovgaard-Olsen et al., 2017) have shown that the link between antecedent and consequent has an influence on whether conditionals are accepted. Participants in a study from Douven and Verbrugge (2010) for instance interpreted conditionals differently depending on the type of the link (deductive, inductive or abductive). Yet, there is also empirical evidence supporting the view that the link between antecedent and consequent is attributable to discourse pragmatics rather than conventional semantics (e.g. see Cruz et al., 2016; Lassiter, 2022).

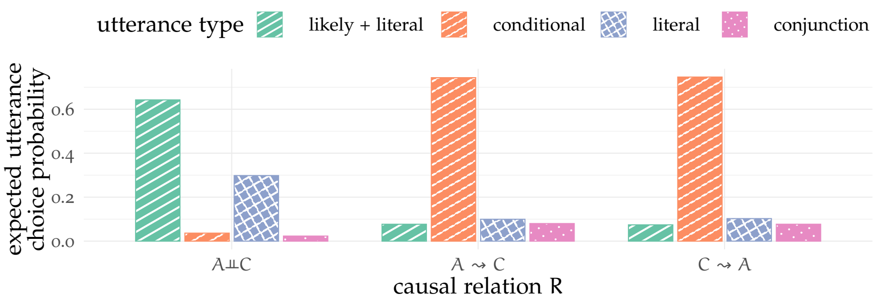


Figure 17: Speaker's expected utterance choice probabilities computed on speaker's predictions, where $\alpha = 3$, for a set of 10,000 states sampled from the default context prior, grouped by utterance type and causal relation.

and the consequent are two independent events. Given the assumption of independence, our hyperrational speaker would never choose any of the conditional utterances (see Figure 7), and even for a speaker with a less optimal rationality parameter ($\alpha = 3$), we observe that conditional utterances only have an expected utterance choice probability $< 5\%$ when considering the speaker's predictions for states with $r = A \perp\!\!\!\perp C$ from a set of 10,000 samples from the default context prior, as shown in Figure 17.

The reason for this low probability of choosing a conditional when A and C are assumed to be independent, is that very likely there will be an assertable, more informative utterance available; for the independent Bayes net, the joint probability tables where $A \rightarrow C$ is assertable (i.e., $P^{(s)}(c \mid a) \geq \theta$) also satisfy the assertability condition for a literal or a conjunction. Given the background knowledge of independence of the antecedent and the consequent and given that in these cases the speaker hardly ever chooses conditionals, the listener will naturally be highly surprised by the speaker's utterance of a conditional. It is this notion of surprise on the part of the listener, based on the speaker's expected utterance choice probabilities, that allows us to infer the infelicity of missing-link conditionals: due to the listener's surprise, she will, arguably, want to look for another way of rationalizing the assertion. If no such option is forthcoming, the utterance feels infelicitous. Missing-link conditionals are therefore *not* accepted by the listener — contrary to the conditional in the Sundowners Example where the listener's initial surprise can be repaired at the content-level since the listener is, after all, able to accommodate a connection between the antecedent and the consequent.

The infelicity of missing-link conditionals as we explain it here, may further be couched within the resource-rational, sampling-based approach to explain causal reasoning.[7] For example, Dasgupta et al. (2017) show that cognitive biases like *subadditivity* or *superadditivity* can be explained by a comparison of people's search for a plausible

7 Thanks to an anonymous reviewer for highlighting this interpretation.

explanation of observed data with sampling algorithms, in particular MCMC sampling.[8] They argue that due to cognitive load and time constraints, the number of samples that people can draw is restricted, such that the optimal answer is not found, giving rise to common cognitive biases. As regards to missing-link conditionals, the listener's search for a good explanation of the speaker's utterance, that is, the search for plausible Bayes nets with sufficient explanatory value to rationalize the speaker's utterance, seems to fail (e.g., in Example 39b). Yet, when it comes to the interpretation of Biscuit conditionals (see Section 5.4.2 below) or to conditionals like in the Sundowners example, it seems reasonable for a listener to be able to quickly come up with more or less satisfying candidates.

5.4.2 *Biscuit conditionals*

Similar to missing-link conditionals, biscuit conditionals (BCs) are commonly considered a special kind of conditional as both lack the probably most characteristic feature of conditionals, the relation between antecedent and consequent. Unlike missing-link conditionals, biscuit conditionals are however felicitous, indeed quite common; see (40) for the classical example from Austin (1956).

(40) If you're hungry, there are biscuits on the sideboard.

While Inferentialists have mostly excluded BCs from their analysis of conditionals altogether, we here argue that the felicity of BCs can indeed be explained as a pragmatic phenomenon.[9] The model presented here helps explain why listeners, who will likely assume *a priori* that the consequent and the antecedent in example (40) are independent, would be surprised by an utterance of (40) if it were to be interpreted like a normal conditional. It is plausible that for some conditionals, like BCs, the listener will see a repair strategy, unlike for missing-link conditionals, which enables a rationalization of the conditional after all. However, this rationalization does not take place at the content-level by an attempt to find a relation between the antecedent and the consequent as (unsuccessfully) in missing-link conditionals or (successfully) in the Sundowners Example, but at a different level, for instance at the level of speech-acts. Concretely, this

8 The definitions of *subadditivity* and *superadditivity* as given by Dasgupta et al. (2017, Table 1, p.3) are as follows (emphasis added).
Subadditivity: "Perceived probability of a hypothesis is *higher* when the hypothesis is described as a disjunction of *typical* component hypotheses (unpacked to typical examples)."
Superadditivity: "Perceived probability of a hypothesis is *lower* when the hypothesis is described as a disjunction of *atypical* component hypotheses (unpacked to atypical examples)".

9 Recently, van Rooij and Schulz (2020, 2021) proposed a generalization of their account for the assertability of conditionals which is able to account for BCs as well.

idea could be integrated into our model by making the speaker's utterance choice additionally dependent on communicative goals, for example, whether the speaker wants to perform a speech-act (e.g., offer the listener some cookies to eat), or just wants to make a plain assertion (e.g., inform the listener about the existence of biscuits on the sideboard).

Admittedly this sketch is just a first step towards a satisfying pragmatic account of BCs, as there is certainly more needed to fully explain their use and interpretation. We also do not, with emphasis, suggest that this kind of "surprise-repair" interpretation is actively entertained during each reception of a BC: BCs arguably provide sufficient secondary cues for the listener to trigger a BC-like interpretation, including intonation or, in some languages, word order (e.g., in German, where BCs can occur with verb-third (V3) verb order and not with the usual verb-second (V2) as seen in simple indicative conditionals).

5.4.3 *Conditionals to communicate* in*dependence*

We assumed that, first and foremost, the speaker's aim is to communicate (probabilistic) beliefs about certain events and saw that, based on this assumption, the speaker should only choose to utter a conditional if no other, more informative utterance (e.g., a literal or a conjunction) is assertable. However, in some circumstances, conditionals may be used in order to put emphasis on the fact that the consequent is *not* dependent on the antecedent. Consider the example given in (41):

(41) If you study, you will pass and if you don't study, you will pass, so don't worry, you will pass!

Along the lines that we have been arguing, a conjunction of conditionals alike should be avoided since evidently the speaker could say the same thing with a more informative, shorter alternative, e.g., "*You will pass (no matter what)!*". Notwithstanding, the conditional in (41) seems a natural utterance to say, in particular when the speaker wants to emphasize the relation, here the *independence* between studying and passing an exam. If we extended the set of alternative utterances available to our speaker by including this kind of combined conditionals ($A \rightarrow C \wedge \neg A \rightarrow C$), the listener would infer that the consequent holds true independent of the antecedent. We saw previously that our listener infers from the utterance of a conditional, that the speaker is likely uncertain about the truth or falsity of the antecedent. For simplicity, let us assume the speaker refers to a state s where $P^{(s)}(a) = P^{(s)}(\neg a) = 0.5$. Since the speaker uttered $A \rightarrow C$, $P^{(s)}(c \mid a) \geq \theta$ must hold, and additionally, due to the speaker's utterance of the second conditional, $\neg A \rightarrow C$, $P^{(s)}(c \mid \neg a) \geq \theta$. Then, $P^{(s)}(c) = 0.5 \cdot P^{(s)}(c \mid$

$a) + 0.5 \cdot P^{(s)}(c \mid \neg a) = 0.5 \cdot (P^{(s)}(c \mid a) + P^{(s)}(c \mid \neg a)) \geq \theta$. Therefore, in this state, it is also possible to assert the consequent straightaway. A speaker whose aim it is to communicate her uncertain beliefs should therefore prefer the simpler utterance, assuming higher cost for combined utterances like $A \to C \wedge \neg A \to C$. However, if the speaker's aim was two-minded, including the communication of her uncertain beliefs as well as highlighting the relationship among the variables at hand, the combined conditional should become a likely utterance choice for the speaker: for almost all states in which both conditionals are assertable, $R = A \perp\!\!\!\perp C$ whereas C is also assertable in many states where $R = A \rightsquigarrow C$ or $R = C \rightsquigarrow A$.

Another example where conditionals are used to communicate the independence between antecedent and consequent are so-called *concessive conditionals*; two examples are given in (42) and (43).

(42) Even if you are very polite, she will not help you. [Even if $A, \neg C$]

(43) Even if it rains, they will go hiking. [Even if A, C]

A conditional "Even if $A, \neg C$" like (42) or a conditional "Even if A, C" like (43) seem to communicate a clash between how the world is *observed* to be and how it is expected to be in a *stereotypical* world.

Concerning the former example, one usually expects that being very polite will at least make it more likely that one will get help, corresponding to the relation $A \overset{++}{\rightsquigarrow} C$. Therefore, we may consider the utterance context of (42) to be one where the speaker has a strong belief in $A \overset{++}{\rightsquigarrow} C$, in the extreme case $P_{\text{prior}}(r = A \overset{++}{\rightsquigarrow} C) = 1$, where A and C denote '*x asks y very politely*', respectively '*y helps x*', and x, y refer to any two distinct persons. In this (stereotypical) context, the conditional $A \to \neg C$ is *not* assertable which seems to be signaled by the speaker's use of 'even if' instead of uttering a simple 'if'. That is, with the speaker's choice to use 'even if', she seems to communicate that there is a clash with respect to what one expects based on common world knowledge and what is actually observed. Therefore, the addressee of (42) should reject the assumption that being very polite would make it more likely to get help, which can be considered a shift from a prior utterance context where $P_{\text{prior}}(A \overset{++}{\rightsquigarrow} C) \rightsquigarrow 1$ to a context where $P_{\text{prior}}(A \perp\!\!\!\perp C) \rightsquigarrow 1$. Together with the assertability conditions for the conditional $A \to \neg C$, that is, $P(C = \neg c \mid A = a) \geqslant \theta$, this will make the listener infer that the consequent is most likely true; given that A and C are assumed to be independent, $P(C = \neg c \mid A = a) = P(C = \neg c)$.

In terms of our model, we can interpret this inference to arise by the following steps:

1. The listener tries to make sense of the speaker's utterance $A \to \neg C$, knowing that usually $r = A \overset{++}{\rightsquigarrow} C$.

2. Since in such a world the speaker's utterance is not assertable, the listener tries to make sense of it by reconsidering the speaker's utterance *without* assuming the stereotypical context where A and C are dependent (e.g., $r = A \overset{++}{\rightsquigarrow} C$). Instead, the listener considers the alternative possibility that A and C are independent ($r = A \perp\!\!\!\perp C$).

3. Considering the utterance $A \rightarrow \neg C$ under the assumption that A and C are independent, then yields the conclusion that C is most likely false (i.e., $\neg C$ is true).

The concessive conditional in (43) can be treated analogously; in this case the stereotypical utterance context is such that rainy whether usually prevents people from going hiking, that is $r = A \overset{+-}{\rightsquigarrow} C$. The speaker's utterance 'Even if A, C' is a contradiction with respect to this stereotypical context, which makes the listener reconsider this assumption, yielding the inference that C is independent of A and thus the consequent (C) is most likely *true*.[10]

Douven and Verbrugge (2012) proposed what they call the *Concessive absence of support thesis* (CAST):

> *A concessive indicative conditional "Even if A, B" (or "If A, then still B") is assertable / acceptable if and only if* $\mathrm{Pr}(B \mid A)$ *is less than or equal to* $\mathrm{Pr}(B)$ *but* $\mathrm{Pr}(B \mid A)$ *remains high.* (Douven & Verbrugge, 2012, p.486)

Instead of requiring a second acceptability condition for concessive conditionals like CAST does, we argue that our model is able to account for the intuitive interpretation of "normal" indicatives and concessive conditionals alike — by modeling the utterance context appropriately, but without the need to change the assertability condition of conditionals. For concessive conditionals, we assume — unlike for ordinary indicative conditionals — that A and C are independent, which explains why the listener infers the truth of the consequent from the speaker's utterance of this kind of conditional. Modeling the pragmatic reasoning between listener and speaker helps to explain where this assumption may come from: since the listener cannot make sense of the speaker's utterance assuming a stereotypical world, the speaker's utterance, that is, the concessive conditional, is reconsidered in light of a non-stereotypical world where A and C are independent. The speaker's use of 'even if' instead of 'if' seems to reinforce the listener's reconsideration of the underlying causal structure, on the one hand, which then yields what we take to be the

10 Note that we did not include the relation $r = A \overset{+-}{\rightsquigarrow} C$ in our model since it yields very similar probability tables as $r = C \overset{-+}{\rightsquigarrow} A$. In order to account for the correct causal link, it is, however, necessary to include $r = A \overset{+-}{\rightsquigarrow} C$ as a further relation (here the *causal* direction clearly goes from 'rain' to 'not going hiking', not from 'hiking' to 'not raining').

intuitive interpretation of concessive conditionals (the truth of the consequent). On the other hand, with the use of '*even* if' the speaker seems to communicate to be aware of how one would expect things to be in a stereotypical world, as if a misunderstanding was anticipated otherwise.

5.5 SUMMARY

In this chapter, we showed that when our model is supplied with particular prior beliefs representing concrete utterance contexts, it makes predictions corresponding to what is considered their intuitive interpretation; we showed that the model helps to explain the listener's varying inferences with respect to the probability of the antecedent in three concrete utterance contexts from Douven (2012).

Moreover, we considered how our model positions itself with respect to special kinds of conditionals that lack a dependency relation between antecedent and consequent. On the one hand, our model vindicates the infelicity of missing-link conditionals by considering a notion of listener-surprisal while on the other hand, it is able to explain the felicity of conditionals that communicate the truth of the consequent such as concessive or biscuit conditionals. Yet, for a full account of the latter, that further explains why the speaker chooses to utter a biscuit or a concessive conditional in the first place instead of choosing an utterance without conditional structure, the model would need to be extended; for example by adopting the speaker's intentions to go beyond the communication of her probabilistic beliefs.

6 BEHAVIORAL EXPERIMENT: WHEN DO SPEAKERS UTTER CONDITIONALS?

Up to this point our considerations have been of a theoretical nature. In this chapter, I will now turn to the empirical data. After a short introduction in Section 6.1, I will describe the behavioral experiment that we designed in Sections 6.2 and 6.3 and present and discuss the results in Sections 6.4 and 6.5.

6.1 INTRODUCTION

As shown in the previous chapters, conditionals like '*If* A, *then* C' can be used, among others, to convey important knowledge about rules, dependencies and causal relationships. Much work has been devoted to the *interpretation* of conditional sentences, but much less is known about when speakers choose to use a conditional over another type of utterance in communication. The majority of existent theories on the pragmatics of conditionals concern their interpretation, a prominent one being *Mental Model Theory* (Johnson-Laird & Byrne, 2002). Other pragmatic accounts target particular phenomena observed in the communication with conditionals, such as the interpretation of '*if*' as '*only if*' (e.g. Geis and Zwicky (1971), Horn (2000)). Yet, on the production side, pragmatic accounts have remained rather vague about the reasons why speakers use a conditional sentence rather than an utterance without conditional structure. Grice (1989), for instance, argued that the utterance of a conditional commits a speaker to, what he called an *Indirectness condition*, a relation between antecedent and consequent that was yet not specified further (for recent semantic accounts, see Douven (2017), Douven et al. (2018)). We aim to fill this gap, and make a step towards a systematic, quantitative investigation of the situations that do or do not elicit the natural use of conditionals, namely with the RSA-model presented in Chapter 3 backed up with empirical data.

To this end, we run a behavioral experiment, where we take advantage of peoples' intuitive understanding of physics to manipulate participants' relevant probabilistic and causal beliefs that may influence — according to the RSA-model presented in Chapter 3 — whether utterances with conditional structure are preferred over utterances without conditional structure. Peoples' intuitive understanding of physics has been used to investigate other aspects of language use previously. To test their RSA-model for the pragmatics of causal language (including counterfactual, but not indicative conditionals), Beller et al.

(2020), for instance, conducted an experiment where participants had to judge the movement of billiard balls.

We show participants visual scenes of toy blocks which are created in such a way as to be able to systematically induce a wide range of uncertain belief states in human participants, so that, according to the RSA-model there may be more or less of an incentive to use a conditional as a description. The shown blocks are more or less likely to fall, possibly as the result of another object falling, thus tapping into participants' intuitive grasp of physics to induce uncertain belief states. Our experimental setup further differs from most behavioral experiments investigating the meaning of conditionals in that participants do not have to provide, for instance, acceptance or naturalness ratings, but have the choice to actively create a variety of different conditional or non-conditional utterances.

6.2 PARTICIPANTS & MATERIALS

The code for the experiment, all stimuli and the analysis are publicly available: https://tinyurl.com/pknmm9z9.

PARTICIPANTS. We collected data from 100 English native speakers via the online crowd-sourcing platform Prolific.[1] Only participants who had not participated in any of our pilot studies and had an average approval rate of at least 50% were admitted.[2] Participants were reimbursed with 3.13 £ for their voluntary participation.

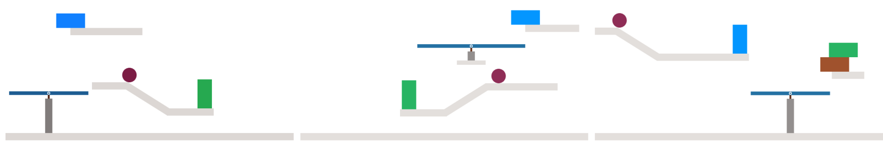
MATERIALS. The experiment consisted of 15 animations in the training phase and 13 static pictures in the test phase. All stimuli from the test phase are shown in Appendix A (Figure 44 and Figure 45).

Situations differed systematically along three dimensions: 'relation', 'prior-antecedent' and 'prior-consequent', specified in Table 9. The two prior-conditions refer to the antecedent- and consequent-block respectively; the former is the block that is shown on the upper platform, the latter is shown on the lower platform. For presentation purposes, the blue block is always shown as the antecedent- and the green block as the consequent-block here, during the experiment colors were, however, randomly assigned to blocks. The prior dimensions specify how likely it is for the blocks to fall initially, without considering the respectively other block. The relation dimension specifies whether there is a causal relation between the two block's falling. An example stimulus for each of the three relations is shown in Figure 18. In situations labeled as independent (ind), stimuli were created such that there is no interaction between the two blocks, whereas there

1 www.prolific.co
2 Note that the final stimuli and setup used in this experiment were created based on several pilot studies, the data of which we did not use further.

relation	prior-antecedent $\triangleq P(b \mid \neg g)$	prior-consequent $P \triangleq (g \mid \neg b)$	$P(g \mid b)$
if_1	$[L, U^-, U, H]$	$[I]$	$[H]$
if_2	$[L, U^-, U, H]$	$[L]$	$[H]$
ind	$[L, U, H]$	$[L, H]$	$P(g \mid b) = P(g \mid \neg b)$

Table 9: Conditions for test stimuli. Letter b denotes proposition 'the blue block falls' (antecedent-block), g stands for 'the green block falls' (consequent-block). I, L, H respectively refer to an impossible event, an event that has low probability, and an event with high probability. U^-, U denote uncertain events, meaning that they are expected to occur approximately at chance level, where U^- denotes events that are slightly less likely than those denoted by U.



(a) ind:UH (b) if_1:UI (c) if_2:UL

Figure 18: Example test stimulus for each relation, where the prior of the antecedent-block to fall is at chance level (U). In Figures 18a and 18c, the ball will roll due to its position over the edge whereas in Figure 18b, it only moves if the blue block falls on the seesaw.

is an interaction in situations with *if*-labels. The difference between if_1- and if_2-trials is that in the former there is only one conceivable possibility for the consequent-block to fall, namely by the rolling ball (which only moves in case that the antecedent-block falls), whereas in the latter it is *also* possible (although not likely) that it falls due to its position on the edge of another block.

From all possible combinations of priors and relations, we choose the 13 most relevant for our purposes. The dependent test situations (if_1, if_2) were chosen to include one trial where the prior of the antecedent-block to fall is low (L), one where it is high (H) and two where it is approximately at chance level (U, U^-). The independent test situations (ind), include two trials were both blocks have the same prior probability to fall, which is either high or low (ind:HH, ind:LL), and two trials where the prior probability of the antecedent-block to fall is at chance level while the prior probability of the consequent-block is again either high or low (ind:UH, ind:UL). We further included a fifth independent situation, in which the antecedent-block is likely to fall whereas the consequent-block is unlikely to fall (ind:HL).

The scenes shown in the test phase are slightly different instantiations of the same kind of scenes as used in the animations of the training phase. All stimuli were created with ‘matter.js’, a rigid body physics engine.[3] The static pictures in the test phase are 820x450 pixel screenshots of animations frozen at their initial state.

6.3 EXPERIMENTAL SETUP

The experiment consisted of a training phase and a subsequent test phase.

TRAINING PHASE. The purpose of the training phase was, on the one hand, to familiarize participants with the stimuli and make them acquire a good sense of the physical properties of the blocks in the simulated world. On the other hand, participants should also become familiar with the use of the sliders so that they would be able to indicate their beliefs appropriately. In the beginning of the training phase, three comprehension questions were used to ensure that participants understood the instructions, in particular the meaning of the four icons that represent the four possible outcomes of a trial (each of the two blocks falls/does not fall). In the ten subsequent preparatory trials (slider-choice trials), participants were shown pictures of slider ratings and were asked whether a given statement was an adequate description of the beliefs represented by the slider ratings (see Figure 43 in Appendix A).

Then the actual training phase started in which participants were shown fifteen animated situations. Before they were able to run the animations, they had to indicate how likely they believed the green and the blue block were to fall. We asked them to adjust four sliders, one for each of the four possible outcomes (only green falls, only blue falls, both fall, neither falls). When participants had estimated the probability of all four events, their ratings were automatically adjusted to sum up to one and they were shown the result of this normalization.[4] Participants then had the chance to update their slider ratings for as long as they liked. After each stage of selection, the current normalized probabilities were shown numerically and, as further visual help, as a blue and a green pie chart, representing the marginal probability assigned to each of the two blocks to fall. When participants were satisfied with their slider adjustments, they would click on a “RUN" button which started the animation.

Before participants could move on to the next trial, they were given feedback about which event actually occurred and how much probability they had assigned to this event. Instructions made clear that

3 https://brm.io/matter-js/

4 The normalization is done with respect to the first slider that participants moved that was not set to 0. For example, if a participant set the first slider to 0.8, the second and third to 0 and the fourth to 0.4 in this order, the value of the first slider is used as reference and the adjusted slider rating in this example is thus $\langle 0.8/(0.8{+}0.4), 0, 0, 0.4/(0.8{+}0.4)\rangle = \langle 2/3, 0, 0, 1/3\rangle$.

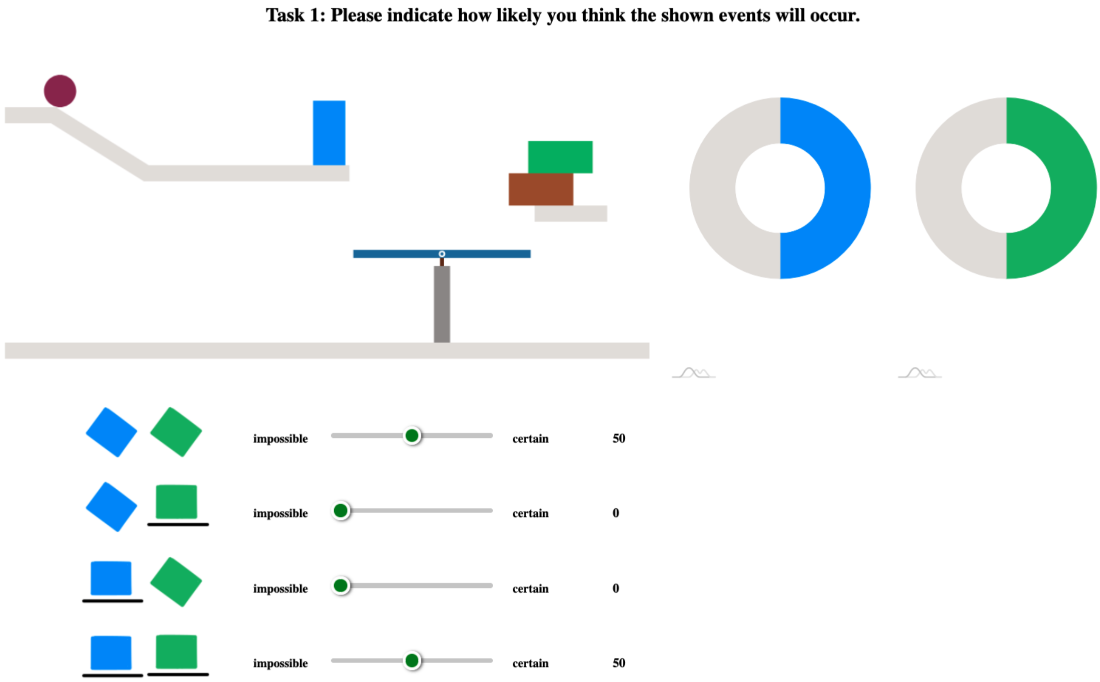


Figure 19: Screenshot of a PE-task trial for condition if_2:UL; the prior probability of the antecedent-block (blue) to fall is at chance level, and the prior probability of the consequent-block (green) to fall is low. The colored circles on the right represent the probability of the blue, respectively the green block, to fall as it is currently indicated by the sliders (here: either the green and the blue block fall or neither does). The icons next to the sliders represent the four events (top: both blocks fall, bottom: neither block falls).

assigning a low probability to the eventual outcome might still have been a reasonable choice due to chance, and participants were encouraged to continue indicating their genuine beliefs and uncertainties. All training trials were pseudo-randomized such that the number of blocks that fall per trial was approximately evenly distributed across all trials.

TEST PHASE. In the test phase, each participant saw each of the 13 test situations once in pseudo-randomized order, such that *high*, *low* and *uncertain* prior conditions were approximately evenly distributed and no subsequent trials had identical relation conditions. Further, an attention check trial was put after each second test trial, asking for the color of a block in a shown picture. For each situation, participants worked on two tasks in direct sequence.

The first task, called the PE-task (prior elicitation), elicited belief judgments about likely outcomes of each physical arrangement. An outcome always concerned two blocks, the blue and the green block, so that participants had to judge the probability of four possible events; both blocks fall / only the green block falls / only the blue block falls / no block falls, which are respectively denoted by w_{bg},

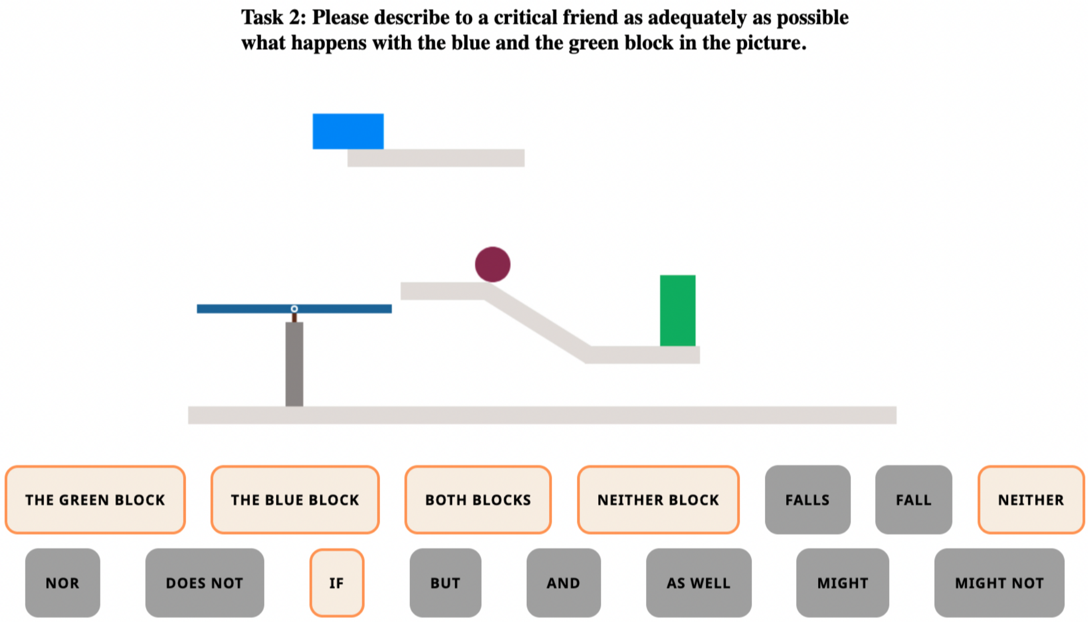


Figure 20: Screenshot of a UC-task trial for condition ind:UH; the falling of the blue and the green block is independent of each other, whereby the prior probability to fall is at chance level for the blue and high for the green block.

w_b,w_g and $w_\varnothing$. Unlike before, during the training phase, participants were now shown static pictures instead of animations and did not get any feedback on their ratings; Figure 19 shows an example trial for the PE-task. We denote the probability judgments of a participant i for trial j in the PE-task by $d^{PE}_{i,j}$, specified in Equation [18].[5]

$$d^{PE}_{i,j} = \langle w_{bg}, w_b, w_g, w_\varnothing \rangle \qquad [18]$$

The second task, called the UC-task (utterance choice), reused the same pictures shown previously in the PE-task. Participants were now asked to "describe to a critical friend as adequately as possible what happens with the blue and the green block in the picture". Participants were instructed not to say what they were not sufficiently convinced of (as the friend is assumed to be critical). The choice of descriptions that participants could possibly create was limited: they were shown a set of buttons with words that had to be clicked on in order to concatenate them to form sentences; see Figure 20 for an example trial of the UC-task. The experiment only allowed concatenations of sentence chunks that formed grammatical sentences, the buttons of those words that could not follow a previously chosen word were grayed out. That is, which buttons were grayed out changed after each selection participants made. The created utterances were incrementally shown in a box in the lower left of the screen and participants had the possibility to make corrections. After submitting the

5 When we refer to the events corresponding to the sliders of the respective probability judgments, we will write $\langle bg, b\neg g, \neg bg, \neg b\neg g \rangle$.

sentence, participants could further give a custom response by freely typing a sentence. They were encouraged to do so if they did not consider any of the given possibilities to be adequate.

The available utterances can be categorized into four different types: conditionals, which is the category that we are mainly interested in, conjunctions, which allow participants to explicitly mention two events with an utterance other than a conditional, and simple assertions (henceforth referred to as *literals*), like 'the blue block falls' or 'the green block does not fall', describing possible outcomes for the green and blue block separately. Each of the four literals could further be combined with 'might' such that participants had another possibility to express uncertainty other than by using conditionals.[6] This yields a total of 20 sentences with distinct meaning, some realizable in multiple ways; 'the blue block falls and the green block falls', for instance, denotes the same distinct meaning as 'both blocks fall'.

6.4 RESULTS

DATA CLEANING. The entire data set of a participant was excluded if (i) they failed any of the attention check questions in the test phase (4) or (ii) got more than half of the ten slider-choice trials in the training phase wrong (1), and if (iii) the average squared differences between a participant's ratings in the PE-task and the mean response of all other participants across all test situations was larger than 0.5 (3).[7] Further, we excluded all data from three participants due to their comments in the end of the study which indicated that they had technical problems or difficulties with the task. Lastly, six single trials were excluded where the event described by the created utterance in the UC-task had been assigned a probability of 0 in the PE-task. After cleaning the data, 88.7% of all trials were included in the analysis, from 89 participants (47 men, 40 women, 2 other) with an average age of ≈ 30 years (range 18-62);[8] 35 participants reported that they have finished high school, 34 have graduated from college and 20 reported to have a higher degree. The experiment was finished within 15 to 60 minutes, on average participants needed 34 minutes.

6 Note that 'might' could not be used within conjunctions, e.g. 'blue might fall, but green does not fall' could not be created.

7 This is a measure of the quality of the data that reflects how often the responses of a participant deviate from the average response of all other participants. For each of the four rated probabilities of participant i in trial j (vector $d^{PE}_{i,j}$), we compute the squared difference to the mean value of the rated probability of all participants except participant i (denoted by $\mu^{PE}_{\bar{i},j}$): $(d^{PE}_{i,j} - \mu^{PE}_{\bar{i},j})^2$, computed element-wise for each of the 4 vector entries. Then, we sum up the resulting four values to get a single value per participant and trial and take the mean value for each participant across all 13 trials which yields a single value per participant, ν_i, and exclude the data of participant i if $\nu_i > 0.5$.

8 One of the participants whose data was entirely excluded failed in more than one criteria.

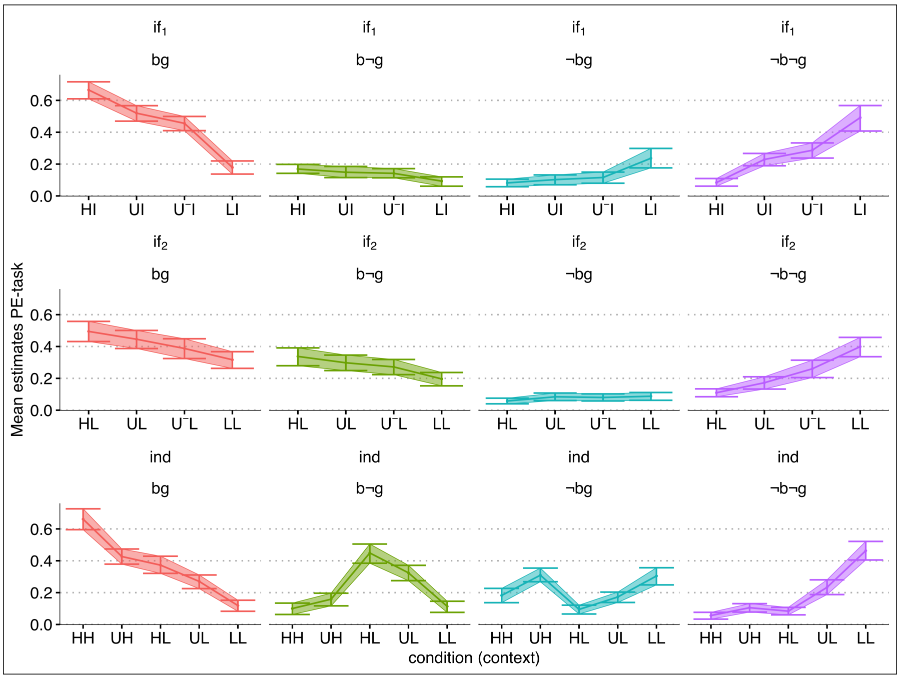


Figure 21: Mean estimates of participants' ratings for each of the four possible outcomes (worlds) in the PE-task of each of the 13 test trials; errorbars and ribbons are 95% bootstrapped confidence intervals.

BEHAVIORAL DATA PE-TASK. Figure 21 shows participants' mean estimates for each possible outcome in all 13 test trials. The stimuli were created in a way that we expect them to elicit different probabilistic beliefs reflected in distinct responses in the PE-task across prior conditions. An exception are the two conditions in the if_1 and if_2 relation conditions in which the prior probability of the blue block to fall is approximately at chance level (U, U^-). The corresponding stimuli differ only slightly so that we expect to see less differences in participants' estimates in these conditions. In the following, we will consider the results for each relation condition in turn.

In the two dependent relation conditions (if_1, if_2), we expect participants' slider ratings w_{bg} (Figure 21, column 1) and $w_{\varnothing}$ (Figure 21, column 4) to differ respectively across the four prior conditions (x-axis). When the prior for the blue block to fall is high (if_1:HI, if_2:HL), w_{bg} should be assigned large probabilities while we expect to see low probabilities for $w_{\varnothing}$. Analogously, we expect large probabilities for $w_{\varnothing}$ and low probabilities for w_{bg} when the prior probability of the blue block to fall is low (if_1:LI, if_2:LL). When the prior for the blue block to fall is neither high nor low (i.e., if_1:UI, if_1:U^-I, if_2:UL, if_2:U^-L) participants' estimates are expected to lie in between their estimates for the low and high prior probability conditions.

Starting with relation condition if_1, eyeballing the data for w_{bg} and $w_{\varnothing}$ (Figure 21, row 1) strongly suggests that participants perceived the four prior conditions as different: a decrease of the prior probability of the the blue block to fall (moving from left to right on the x-axis) comes along with a decrease of participants' mean estimates w_{bg} for outcome bg (row 1, panel 1) and an increase of the mean estimates $w_{\varnothing}$ for outcome $\neg b \neg g$ (row 1, panel 4). Contrary to that, we do not expect the estimates for the other two outcomes (w_b, w_g) to vary across prior conditions. Ideally, w_b and w_g would not receive any probability mass at all: in the if$_1$-trials, the green block will definitely fall if the blue block falls and if the blue block does not fall the green block would not fall either. The fact that we observe mean estimates clearly above 0 for w_b and w_g, suggests that there remains some uncertainty about the relation between both blocks falling. Especially salient is the mean rating w_g for $\neg bg$ in condition if$_1$:LI (row 1, panel 3) which is close to 24%. Since this is the condition in which the blue block has very low prior probability to fall, large estimates for the event that the blue block does not fall are reasonable. However, the world in which the blue block does not fall, but the green block does fall ($\neg bg$), is expected to receive very low estimates, which is — on average — not the case. Thus, this suggests that there is a certain amount of participants who believed that the green block could fall even though the blue block does not fall.

Turning to relation condition if$_2$ (Figure 21, row 2), we observe similar tendencies for the two worlds where either both blocks fall or do not fall (bg, $\neg b \neg g$) as we did in the the if$_1$ relation condition — however with greater uncertainty, reflected by less extreme values shifted more towards values in the middle of the range between 0 and 1. Contrary to that, the observed pattern for participants' estimates w_g and w_b for worlds $\neg bg$ and $b \neg g$ seem slightly different in the if$_2$-trials than those observed in the if$_1$-trials. Across all prior conditions, the mean estimates for the case that the blue block falls but the green block does not fall ($b \neg g$) seem larger in the if$_2$-trials than in the if$_1$-trials (row 2, panel 2 vs. row 1, panel 2). That is, in the if$_2$-trials, participants seem to judge it more likely than in the if$_1$-trials that the blue block falls while the green block does not fall. Further, participants estimated the case where only the green block falls ($\neg bg$) as quite unlikely in relation condition if$_2$:LL, which, however, received rather large mean estimates in the corresponding if$_1$ condition (if$_1$:LI). In the condition where the prior probability of the blue block to fall is low (if$_1$:LI, if$_2$:LL), participants thus seem to judge it *more* likely in condition if$_1$ than in condition if$_2$ that the green block falls while the blue block does not fall. This is surprising since particularly in the if$_1$-trials, we would expect participants to believe that it is impossible for the green block to fall without that the blue block falls. A possible explanation might be given by the ball in the if$_1$-trials. Some participants might have erroneously believed that the ball moves independently of the blue block, which is the case in the independent- and if$_2$-trials, in which the ball is positioned on the edge. In condition

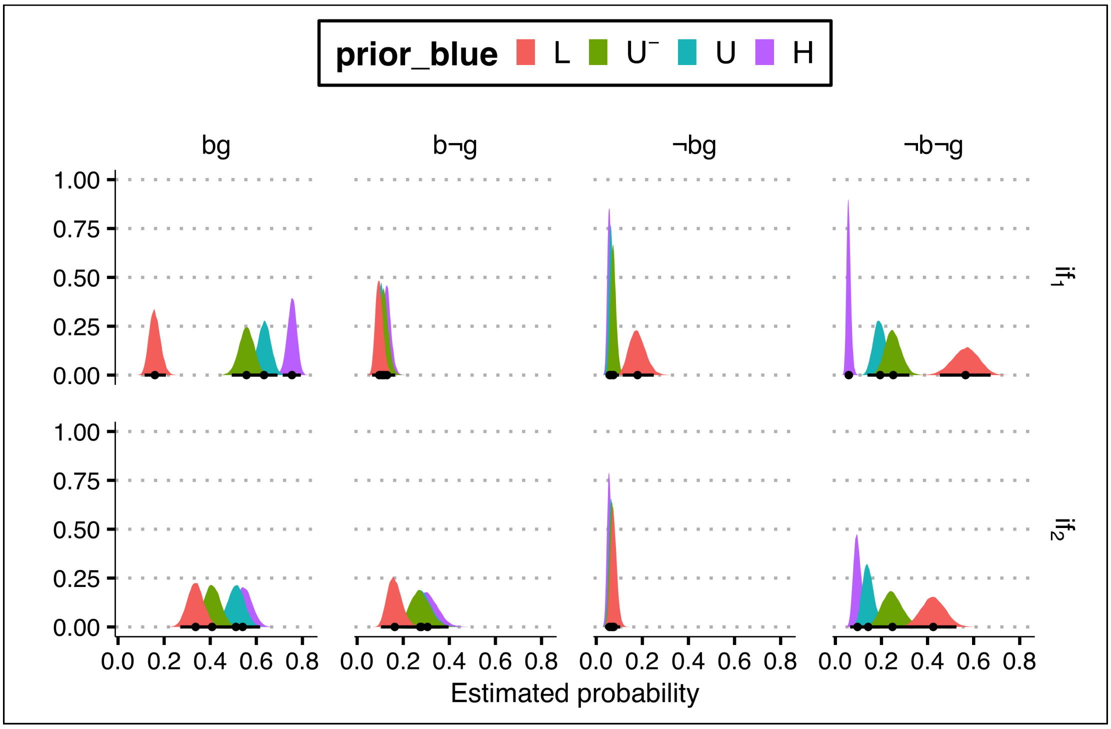


Figure 22: Posterior probability from the Dirichlet-regression model for the probability estimate of the four sliders in the two dependent conditions of the PE-task. Bars are 95% highest density intervals (HDI) of the mean.

if_1, the ball would, however, only start rolling down the ramp if the seesaw moves, that is, if the blue block falls.

For the independent relation condition (Figure 21, row 3), as expected, the different prior conditions seem to provoke different estimates for all four worlds, including b¬g and ¬bg.

To test statistically whether the different conditions elicit different beliefs — manifested in distinguishable slider ratings across conditions — we ran two Dirichlet Bayesian regression models with the R-package brms (Bürkner, 2017), one for the dependent and one for the independent relation condition. The latter predicts participants' slider ratings by the prior condition of the blue, and respectively the green block (H,U,L) as well as their interaction, with the maximal random effect structure, that is, using by-participant random effects for all regression parameters (intercept, slopes and interactions). The predictors in the model for the dependent relation condition, are the prior condition of the blue block (H, U, U^{-},L), the specific relation (if_1, if_2), and their interaction, again with the maximal random effect structure.[9] We use the default prior from brms for the Intercept (a student_t distribution with parameters (3, 0, 2.5)), but custom priors for the regression coefficients. Since these are estimated on the logit-scale,

9 Brms formula for the dependent model: y ~ pblue ∗ relation + (1 + pblue ∗ relation | subj) and the independent model: y ~ pblue ∗ pgreen + (1 + pblue ∗ pgreen|subj)).

we constrain the considered values by using a normal distribution with $\mu = 0, \sigma = 2.5$ (instead of the default flat priors), so that most values will lie in the interval [-4.9, 4.9] (95% highest density interval). The probabilities that are represented by these values range from 0.007 to 0.993 (applying the inverse logit function, logit^{-1} = logistic). That is, larger or smaller logit-values than those that are likely given the assumed prior only yield marginal differences in the corresponding probabilities and can thus be ignored.

Figure 22 shows the posterior distributions for the estimated probabilities of each of the four sliders for the dependent relation condition.[10] For condition if_1, the model provides strong evidence for the hypothesis that participants' estimates for w_{bg} differ across prior conditions of the blue block; the posterior probabilities for w_{bg} to be estimated as more likely in (i) prior condition H than in prior condition U is 1, (ii) prior condition U than in prior condition U^- is 0.988, and (iii) prior condition U^- than in prior condition L is 1. Similarly, there is strong evidence (posterior probability ≈ 1) for the respective comparisons between different prior conditions for the estimates for $w_{\varnothing}$; only the posterior probability of prior condition U^- to yield larger estimates for $w_{\varnothing}$ than prior condition U is with a value of 0.96 slightly smaller than 1.

As anticipated, the posterior distributions of relation condition if_2 are less distinguishable across prior conditions than they are for relation condition if_1. There is strong evidence for the two uncertain prior conditions (U, U^-) to yield different estimates w_{bg} and $w_{\varnothing}$ (posterior probabilities ≈ 1). Yet, only for condition U^-, not for U, there is strong evidence to yield higher estimates w_{bg} than prior condition L (posterior probability 0.959) *and* lower estimates w_{bg} than prior condition H (posterior probability 0.999). Contrary to that, for condition U, the posterior probability for outcome bg to be estimated smaller than in condition H is 0.781. That is, condition U^- provides more distinguishable responses from the responses in condition H than U does.

Lastly, the posterior distributions of the independent relation condition are shown in Figure 23. We will not consider them in more detail here since we are mainly interested in checking that prior conditions H, L and U and/or U^- for the blue block in the dependent relation conditions yield distinguishable slider ratings to see whether participants utterance choice depends on these categories. More precisely, we expect to see more conditionals in the UC-task for relation $\in [\text{if}_1, \text{if}_2]$ and prior condition of the blue block $\in [U, U^-]$ as com-

10 Since posterior predictive checks (see Appendix A) reveal that some aspects of the data are not fully captured by the model, in particular for the data where the prior condition of the blue block $\in [U, U^-]$, these results need to be interpreted with caution. Indeed, the Dirichlet regression is not ideal to model our slider rating data (but sufficient for our purposes here) since it requires to smooth the data such that all values lie in the open interval (0,1) whereas participants had the option to select 0 and 1.

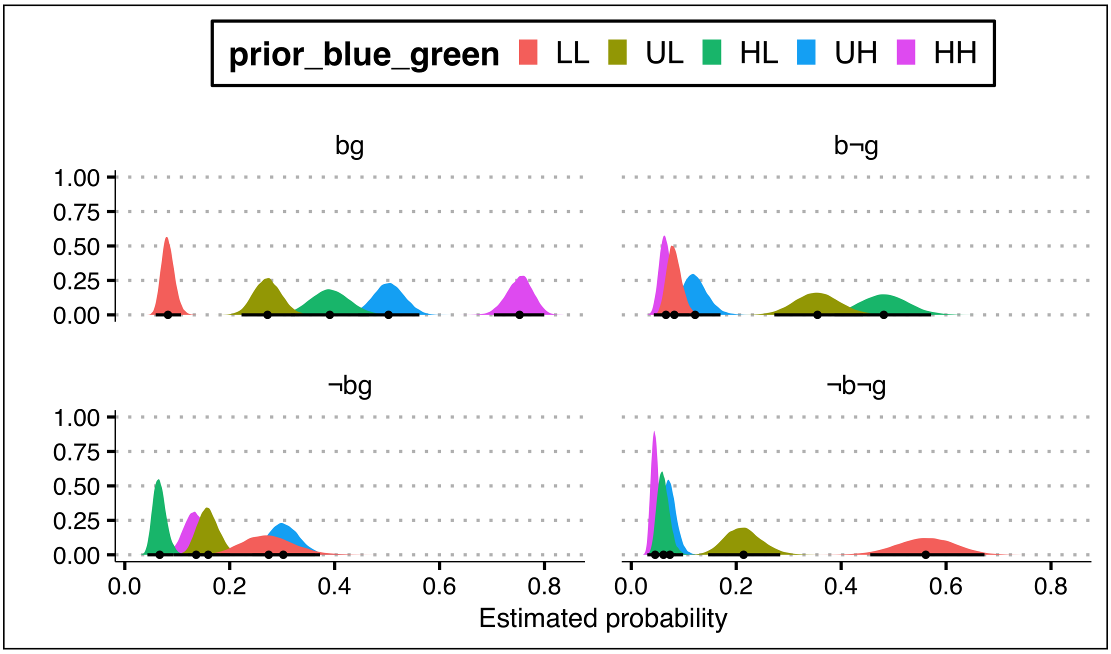


Figure 23: Posterior probability from the Dirichlet-regression model for the probability estimate of the four sliders in the independent condition of the PE-task. Bars are 95% highest density intervals (HDI) of the mean.

pared to $[H, L]$, while in the independent condition we do not expect participants to select conditionals at all, no matter the prior condition of the two blocks.

BEHAVIORAL DATA UC-TASK. In the UC-task, participants were asked to describe the shown scene of block arrangements. To create an utterance participants had to concatenate chunks of words by clicking on the respective buttons (see Figure 20). In total, they could create 20 utterances with different meanings, some of which were realizable in several ways. The utterance 'both blocks fall' was for instance realized by seven different utterances, the most frequent ones being 'both blocks fall' (283) and 'the blue block falls and the green block falls as well' (39); we will discuss the influence of the length of the selected utterance later on in Section 6.5. For the purpose of a clean presentation, we use 20 standardized utterances, which respectively summarize all their possible different realizations. In the following, we refer to these 20 standardized utterances if not indicated otherwise.

Figure 24 shows the overall number of selections of each of the 20 utterances. We observe a strong preference for conjunctions, in particular for the utterance 'both blocks fall'. Further, we observe a strong preference for positive affirmations when considering participants' use of utterances with 'might'. Overall, 'blue might fall' and 'green might fall' are selected approximately 6, respectively 4, times as often (85 and 51 selections respectively) as their negative counter-

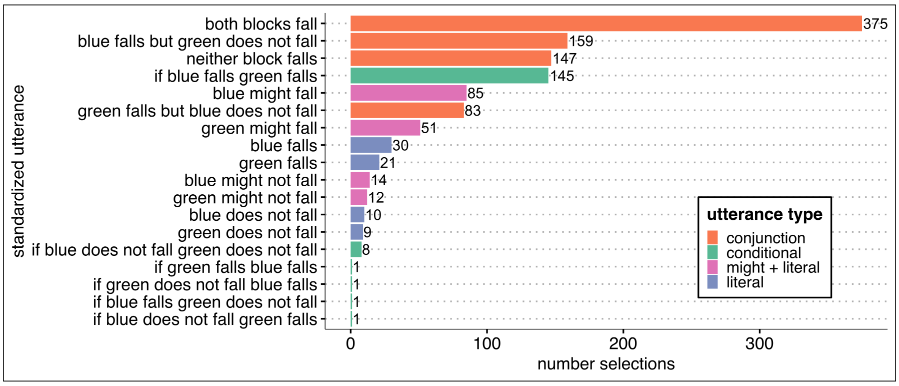


Figure 24: Total number of selections of each (standardized) utterance in the UC-task. For better readability, utterances are shortened here (e.g., 'green falls' is short for 'the green block falls').

parts, 'blue might not fall' (14) and 'green might not fall' (12) . Concerning participants' use of literals, a similar, although less strong tendency is observed: 'blue falls' and 'green falls' are selected 30, respectively 21 times whereas 'blue does not fall' and 'green does not fall' are selected 10, respectively 9 times. But first and foremost we are interested in participants' use of conditionals. In accordance with our stimuli, we did not expect participants to create conditionals with only one of antecedent and consequent being negated (e.g., 'If the blue block falls, the green block does not fall'). This is confirmed by our data: three of the four possible conditionals of this kind were selected just once, and the conditional 'If the green block falls, the blue block does not fall' was not selected at all. With respect to the other four conditionals available, where either both or neither of antecedent and consequent are negated, we observe a strong preference for the conditional 'If the blue block falls, the green block falls'. It is with 145 selections by far the most selected conditional, followed by 'If the blue block does not fall, the green block does not fall' which was, however, selected only eight times. The conditional 'If the green block falls, the blue block falls' was selected just once and the conditional 'If the green block does not fall, the blue block does not fall' was not selected at all.

We expected the number of conditionals selected in the UC-task (as compared to non-conditional utterances) to be dependent on the relation condition, where the dependent conditions are expected to trigger conditionals more than the independent conditions. In fact, we do not expect conditionals to be selected in the independent relation condition at all. Further, we expect the selection of conditional utterances to be dependent on the prior probability of the antecedent-block to fall: in the prior conditions with uncertainty in the falling of the antecedent-block (e.g., if_1:UI) conditionals are expected to be selected more often than in the high and low prior conditions for the antecedent-block (e.g. if_1:HI).

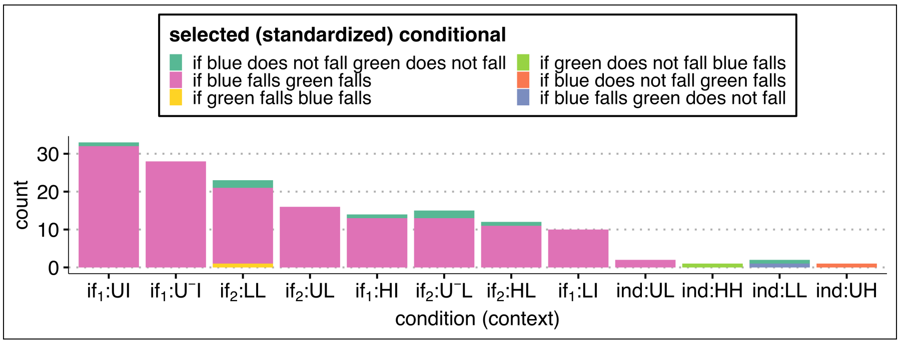


Figure 25: Conditionals that participants selected in the UC-task.

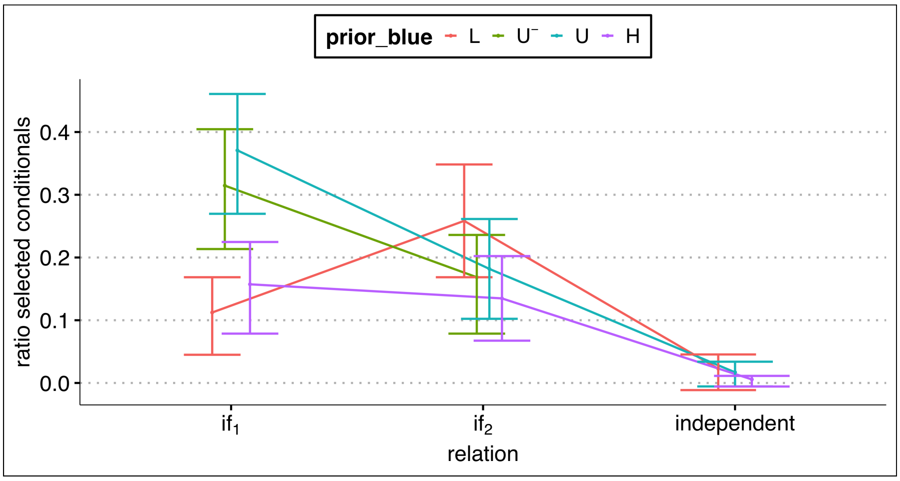


Figure 26: Proportion of conditionals that participants selected in the UC-task with 95% bootstrapped confidence intervals; color code is prior condition of the antecedent-block.

Before we come to the statistics testing these predictions, let us shortly eyeball the data. Figure 25 shows the overall number of selections of each conditional in each of the 13 test trials. It is immediately apparent that — in line with our expectations — participants hardly ever selected conditionals in the independent-trials. The same data is shown as relative proportions in Figure 26. In the if_1-trials the two conditions that were supposed to elicit uncertainty about the falling of the antecedent-block, as expected, triggered the selection of conditionals most often. On the contrary, in case of the if_2-trials, it is the condition where the prior of the antecedent-block is supposed to be low (if_2:LL) that triggered most conditionals, followed by the expected two conditions that involve uncertainty with respect to the prior probability of the antecedent-block to fall (if_2:UL and if_2:U¯L).

Turning to the statistical tests, we ran a logistic regression model using the R-package brms (Bürkner, 2017) with varying intercepts and slopes per participant to predict participants' choice of conditional vs. non-conditional utterances based on the relation (if_1, if_2) and the prior of the antecedent-block to fall (H/L vs. uncertain, where the latter comprises U, U^-).[11] We hypothesized that in the two dependent relation conditions uncertain situations (i.e., beliefs neither close to 1 nor close to 0) would be described more often with conditionals than situations in which participants are quite certain which blocks would fall (i.e., beliefs close to 0 or 1). Since the regression model of the PE-task data provided evidence that prior conditions U, U^- indeed elicited different probabilistic beliefs in conditions if_1 and if_2 than prior conditions H and L, it is reasonable for us to test whether the latter two prior conditions yield less selections of conditional utterances than the former, the uncertain prior conditions.

For relation condition if_1, our hypothesis is confirmed: the posterior probability of the probability of a conditional utterance to be larger in relation condition if_1 when the prior of the antecedent-block is uncertain as compared to (i) high (H) is 0.98 and (ii) low (L) is 1.

For relation condition if_2, the picture looks slightly different. There is no evidence for the proportion of conditionals selected in the uncertain prior conditions to be larger than in prior conditions H and L. To the contrary, the posterior probability for the proportion of conditionals selected in the low prior condition to be *larger* than in the uncertain conditions amounts to 0.92. However, this fits with the results from the PE-task. Participants' probability estimates in the PE-task already suggested that in condition if_2, the prior conditions of the antecedent-block were perceived as less distinct than for condition if_1. In particular, the range across all four prior conditions of participants' estimates for w_{bg} was smaller for condition if_2 than for if_1. The bootstrapped confidence intervals of w_{bg} in condition if_2 range from a minimum of a relatively large value of 0.26 (prior condition antecedent-block: L) to a relatively low maximal value of 0.558 (prior condition antecedent-block: H) whereas this range is more widespread for relation if_1, ranging from a minimum of 0.137 (prior condition antecedent-block: L) to a maximum of 0.717 (prior condition antecedent-block: H). Thus, participants were more uncertain in condition if_2 for prior condition L of the antecedent-block (95% HDI of mean: [0.27, 0.41]), the condition for which we observe most conditionals in relation if_2, than they were uncertain in the corresponding prior condition for if_1 (95% HDI: [0.12, 0.21]); see Figure 22, first column.

Concerning the independent relation condition, our expectations — namely no selections of conditionals — are confirmed by the model: the posterior probability of the proportion of selected conditional utterances to equal 0 in the independent relation condition exceeds 95% in all three prior conditions (H, L, U).[12]

11 brms model formula: y ~ 1 + priorblue ∗ relation + (1 + priorblue ∗ relation | subj). Posterior is approximated by 8000 MCMC-samples, including 1000 warmup-samples, for 4 chains.

12 R code: hypothesis("Intercept + β_{ind} + β_X + $\beta_{X:\text{ind}}$ = 0") with $X \in [H, U, L]$

CUSTOM RESPONSES UC-TASK. From a total of 1153 trials, custom responses were given in 85 trials (≈ 7.3% of trials) by 23 participants (≈ 25.8% of included 89 participants).[13] Surprisingly, some of the custom responses were identical to participants' selected responses (14/85). The large majority of the remaining custom responses can be summarized to belong to two groups. Participants in the first group used different words to effectively describe what they had communicated with their selected utterances. The second group covers those cases where participants would have preferred a conjunction where one conjunct (or both) contains the word 'might'.

In group 1, some of the responses were simple replacements of single words: in two cases 'and' was replaced by 'but' (which was in fact a proper option) and in one case 'and' was replaced by 'while' (which was not a proper option). In other cases, the selected utterance was shortened in the custom response. The utterance 'The blue and the green block fall' was, for instance, shortened several times in the custom response to 'Both blocks fall' (which was also a proper option). Similarly, 'Both blocks fall' and 'Neither block falls' were respectively shortened to 'Both fall' and 'Neither falls' or 'Neither fall', which are all wordings that were no proper options in the main task. Instead of shortening a responses, one participant also refined the selected response, replacing 'both blocks fall' by 'both the blue and the green block will fall'. Another interesting observation is the reference to causal language in the custom response replacing a selected conditional. One participant, for instance, rephrased 'The green block falls if the blue block falls' three times by using the word 'cause' in the consequent: 'If the blue block falls, it'll cause the green block to fall too' (in conditions if_1:UI, if_2:UL and if_2:LL). Similarly, a participant selected the utterance 'The blue block might fall' and replaced it by the utterance 'The blue block might fall which will make the green block fall' (in conditions if_2:LL and if_1:UI).

The larger group is the second group where participants' custom responses were conjunctions including the word 'might'. The selected utterance 'the blue block might fall' was, for instance, rephrased by the utterance 'both blocks might fall'. Some participants selected a conjunction (e.g., 'the blue block falls but the green block does not') but indicated in their custom response that they would have preferred a conjunction using 'might' in one conjunct (e.g., 'the blue block falls but the green block *might* not'). Similarly, three participants would have preferred to use 'might' in the consequent of a conditional (e.g., 'if the blue block falls, the green block might fall'). All three cases were observed for if_2-trials, the setting of which was a little more complex than it was for if_1-trials. Further, there were eight cases where participants' selected response was an utterance with might (e.g., 'the blue block might fall') and the corresponding custom response a conjunc-

13 See Table 16 in Appendix A for a list of all custom responses.

tion that either makes explicit that the other block will fall (e.g., 'the green block falls and the blue block might fall') or will not fall (e.g., 'the blue block might fall but the green block won't fall').

Most of the custom responses were created in independent trials (34), followed by if_2-trials (27) and least of them fall onto if_1-trials (10). In the independent trials participants could not describe their uncertainty about one block and at the same time explicitly say something about the other block which is reflected in the large number of custom responses that use 'might' within a conjunction. In the if_1-trials participants do not seem to feel the need for a different kind of utterance as much as in the if_2-trials for which more than 2.5 as many custom responses were created. This reflects the additional complexity of the if_2-stimuli; here participants seem to be less convinced about the fact that the blue block will make the green block fall and that it will not fall in case that the blue block does not fall.

JOINT DATA FROM PE- AND UC-TASK. Now that we have considered the data from the PE-task and the UC-task in isolation, we will turn to consider the data jointly to get a sense of the relation between participants' concrete prior ratings for a given scene and their choice of description for that same scene. Recall that the stimuli were shown twice in direct sequence, first in the PE-task, in which participants were asked to indicate their beliefs regarding the (falling) behavior of the two blocks, and consequently in the UC-task, in which participants were asked to create a sentence that described the visual scene.

We expect participants to indicate a strong belief — in terms of the probabilities provided in the PE-task (or derived thereof) — in the event that they describe in the corresponding UC-task. The only exception are those cases where participants select an utterance with 'might' (e.g., 'the blue block might fall') in which we expect to see beliefs ≈ 0.5 that the respective block falls or does not fall. According to the semantics discussed in Section 3.2.1 (see Table 3), the considered probabilities depend on the selected utterance: when participants describe the visual scene with a conjunction, we consider the corresponding joint probability (e.g., $P(b, g)$ for 'both blocks fall'), when they select a literal or a literal combined with 'might', we consider the corresponding marginal probability (e.g., $P(b)$ for 'the blue block might fall' / 'the blue block falls') and when they select a conditional, we consider the corresponding conditional probability (e.g., $P(g \mid b)$ for 'if the blue block falls, the green block falls'). The mean estimates (computed from participants' slider ratings in the PE-task) of the probability of the events described by the utterance that participants selected in the UC-task are shown in Figure 27. Let us consider participants' indicated probabilistic beliefs for each of the four different utterance types in turn.

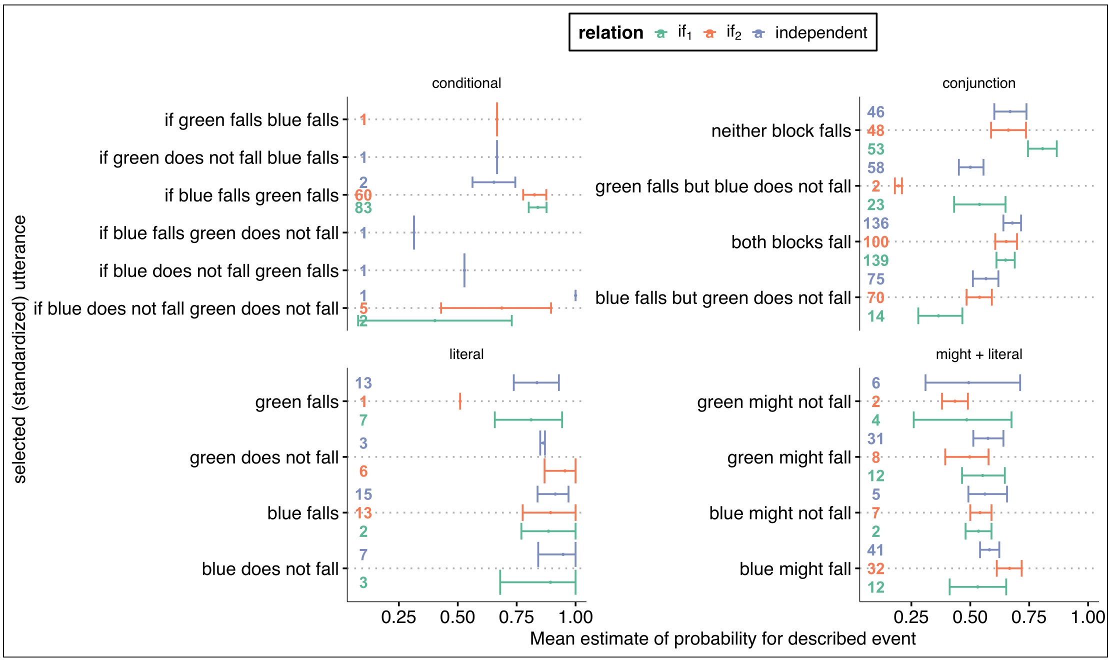


Figure 27: Mean estimates of the probability for the event described in the UC-task with 95% bootstrapped confidence intervals. The relation condition is color coded; colored numbers are the number of times that the respective utterance was selected in the respective relation condition.

When participants selected a literal ('green falls/does not fall', 'blue falls/does not fall', Figure 27 lower left panel), their estimates of the corresponding marginal probabilities tend to be large (average estimates ≥ 0.81, excluding the single case where 'green falls' was selected in an if_2-trial in which $P(green)$ was assigned a probability of 0.51.

Concerning participants' selection of conditional utterances, 'If blue falls, green falls' is the only conditional that was selected more than a handful of times (Figure 27 upper left panel). In these cases, participants' estimates of the corresponding conditional probabilities also tend to be large (mean estimate of 0.84 for if_1- and 0.826 for if_2-trials).

As expected, literals combined with 'might' are assigned much lower values, clustered more or less around 0.5 (Figure 27 upper left panel).

In those cases where participants selected a conjunction to describe the scene, their estimated probabilities of the corresponding events are more surprising: contrary to our expectations, these are, on average, assigned quite low probabilities with mean values between 0.195 (utterance 'green falls but not blue', relation if_2) and 0.807 (utterance 'neither green nor blue' for relation if_1). The mean (median) estimates for the sliders corresponding to the selected utterance (i.e., conjunction) across relation conditions are 0.49 (0.54) for 'blue falls but not green', 0.66 (0.65) for 'both fall', 0.41 (0.50) for 'green but not blue' and 0.71 (0.67) for 'neither green nor blue falls'.

6.5 DISCUSSION

While most experimental studies on conditionals focus on their interpretation, our experiment focused on the speaker part instead. We manipulated scenes of block arrangements with respect to the causal relation of two target blocks and their prior probability to fall to investigate whether these factors influence participants' utterance choices to describe these scenes. A Dirichlet regression model confirmed that our manipulations have likely elicited different probabilistic beliefs in participants — as measured by the slider ratings in the PE-task — which mostly correspond to how we intended them to be (an exception are the observed estimates for w_g in $\text{if}_1$1:UI as discussed above). The production data from the UC-task showed (using a Bernoulli regression model) that, as anticipated, the independent relation condition did not trigger a proportion significantly above 0 of conditional utterances, whereas relation conditions if_1 and if_2 did. Further, our expectations with respect to the relation between the prior conditions of the antecedent-block and the proportion of selected conditional utterances was confirmed, at least in relation condition if_1: participants' selected more conditional utterances in the uncertain prior conditions (probability of the antecedent-block to fall approx. at chance level) as compared to high and low prior conditions. We did not observe this relation in condition if_2, in which participants estimated beliefs measured by the slider ratings were distinct across prior conditions, yet closer together than in condition if_1.

A surprising aspect of our data that draws our attention, in particular with respect to our RSA speaker model, is participants' observed readiness to select conjunctions in the UC-task despite indicating quite low beliefs in the described event in the PE-task. We can only speculate about the reasons that may have contributed to this pattern. Participants might, for instance, have looked at the scene and instead of expressing their apparent uncertainty about the falling of one or both blocks that we observe in the slider ratings from the PE-task, they made a decision which among the four possible outcomes is most likely and described it with a conjunction in the UC-task. We wanted to avoid this by encouraging participants in the instructions to make only claims that they were convinced to be true. More precisely, we asked them that the friend to whom they describe the shown scene was critical and expects them to say only things that they were confident about. A stronger incentive may, however, be necessary to ensure that participants only select utterances that they really believe to be true. This may be achieved by introducing consequences for participants' utterance choice, for example by a slightly different experi-

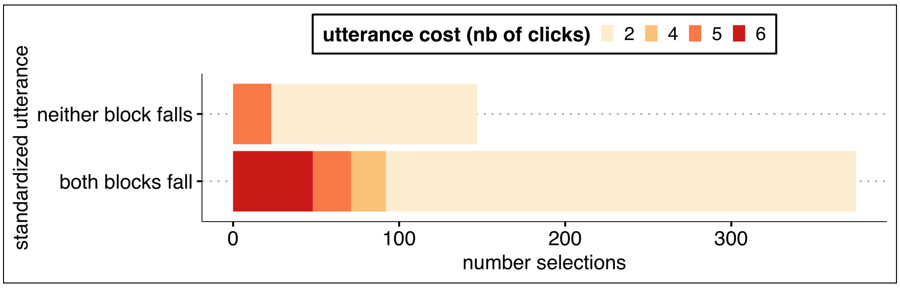


Figure 28: Number of selections of the two utterances 'neither block falls' and 'both blocks fall' that could be realized in different ways requiring a different number of clicks, indicated by the color-coded utterance cost.

mental setup of the UC-task. One possibility would be to formulate it in terms of a classical reference game where the participant has to describe a target scene which the interlocutor has to find among a set of different scenes, based on the participant's description. The set of scenes may comprise two dependent scenes, for example if_1:UI and if_1:HI, and the two independent scenes with corresponding prior probabilities, in this example ind:UH and ind:HH. When the target scene is if_1:UI, we would expect to see a preference for the conditional 'if the blue block falls, the green block falls' over the conjunction 'both blocks fall' since there is an alternative scene that is better described by the conjunction, and would thus, probably be more likely chosen by the interlocutor.

Especially with regard to the independent scenes, it would further be reasonable to allow conjunctions that include 'might', like 'the blue block might fall, but the green block does not fall' which is also suggested by the custom responses that participants gave. This would allow participants to explicitly say something about both blocks even though they are not sure whether or not each of the two blocks will fall.

Another aspect that might play a role in participants' extensive use of conjunctions (even despite rather low beliefs in the corresponding event) are the utterance cost in terms of the number of clicks that were necessary to create an utterance. While the creation of a conditional utterance required at least five clicks ('if the blue block falls the green block falls [as well]'), the conjunction 'both blocks fall' required only 2 clicks, or 4 when expressed as ' the blue/green block and the green/blue block fall' (underlined words correspond to buttons). Similarly, at least 7 clicks were necessary to create the conditional 'if the blue block does not fall the green block does not fall [either]' whereas a minimum of 2 clicks were necessary to create the conjunction 'neither block falls' in its shortest version. If participants particularly aimed to create short sentences — which is not unreasonable to assume, especially in an online experiment — they may have prioritized utterance cost over the requirement to only say what

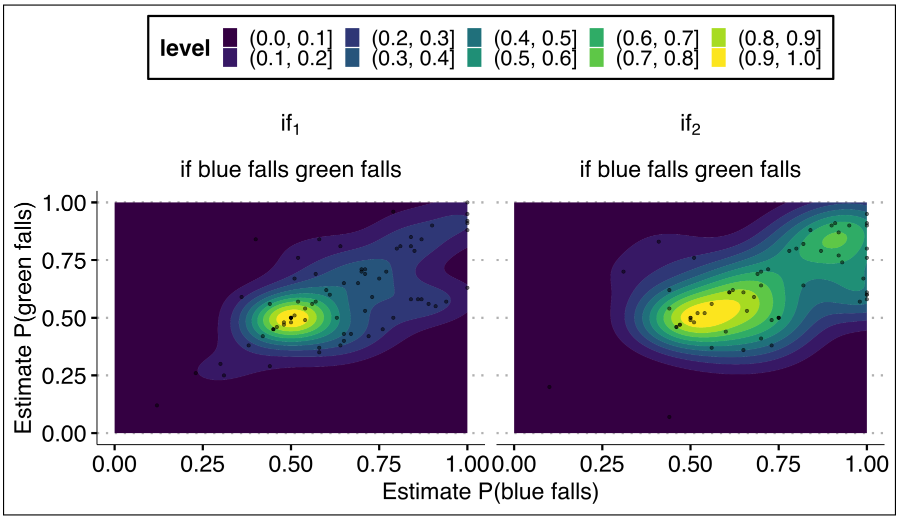


Figure 29: Density estimates of participants' estimates of the probability that the blue (x-axis) and the green (y-axis) block fall in the if_1- and if_2-trials where they selected the conditional 'if the blue block falls, the green block falls' in the UC-task.

they are really confident about, especially since their utterance choice did not have any consequences in the experiment. Figure 28 shows how often the two utterances 'neither block falls' and 'both blocks fall' were expressed with 2, 4, 5, and 6 clicks. For both utterances, short expressions that require only 2 clicks are clearly preferred.

There is one last aspect of the joint data from PE and the UC-task that is worth mentioning. According to Grice (1989), a speaker will not choose to utter $A \rightarrow C$ in case that she is in a position to say something more informative (e.g., C). Under this assumption, we should expect participants' estimates of the probability that the blue block falls, as well as of the probability that the green block falls, to cluster around 0.5 in trials where they select the conditional 'if the blue block falls, the green block falls' as description of the scene. This is pretty much what we observe in Figure 29 which shows density estimates of the joint probability of the blue and the green block to fall in if_1- and if_2-trials where participants selected 'if the blue block falls, the green block falls' in the UC-task. Most of the density is centered round (x=0.5, y=0.5). For both, if_1- and if_2-trials, the higher density region spreads from the center towards (x=1, y=1). In particular in the if_2-trials, participants selected the conditional 'If blue falls green falls' despite indicating a rather strong belief in the antecedent and the consequent (upper right corner, right panel Figure 29). However, we do not observe selections of this conditional in combination with low beliefs in the antecedent and the consequent (lower left corner, Figure 29).[14] This is an interesting observation as one might expect that,

14 Note that we do generally observe beliefs below 0.5 in the antecedent (that the blue block falls) in the dependent relation conditions which were yet less frequent (n =

given that participants do select the conditional 'If blue falls green falls' even though they indicate a large belief in both, $P(b)$ and $P(g)$, they would also select the conditional when they indicate to have low beliefs in $P(b)$ and $P(g)$ — which we do not observe in our data, however. One aspect that might contribute to this observation and which might be worth looking at is the fact that only for low beliefs in $P(b)$ and $P(g)$ the more informative utterances contain negations.

174) than beliefs above 0.5 ($n = 497$). When participants indicated low beliefs in the antecedent and the consequent, they selected other utterances than 'If the blue block falls, the green block falls'.

7 MODELING THE EMPIRICAL DATA WITH RSA

7.1 MODEL DEFINITION

We aim to test the RSA-model presented in Chapter 3, a vanilla RSA-model (Franke & Jäger, 2016; Goodman & Frank, 2016) adapted to be applicable to communication of stochastic/causal dependencies, for its ability to predict empirical data on the speaker's choice of conditional vs. non-conditional utterances.[1] I will summarize the most important aspects of the model along the way where I consider it helpful to follow this chapter more easily without the the need to go back to Chapter 3 in which it is explained in detail, starting with a short overview of the model.

7.1.1 *Background RSA-model*

RSA-models are probabilistic models that formalize Gricean pragmatic reasoning: the speaker's utterance choice is predicted to depend on the utility of an utterance for communicating a state, in relation to the utility of plausible alternative utterances available to the speaker. As the relevant data we consider here is for the choice of a suitable description, we focus on the speaker part of vanilla RSA:

$$P_S(u \mid s) \propto \exp(\alpha \cdot U(u; s)) \cdot P(u) \qquad \text{[6 revisited]}$$

The free parameter α tweaks the extent of 'rationality' of the speaker; larger values of α correspond to stronger pragmatic inferences, that is, the larger α, the more the speaker's predicted distribution will be peaked on the utterance with the largest utility. The utility of an utterance u for a state s, $U(u; s)$, corresponds to its degree of informativeness, defined in terms of the literal meaning of u. Further, utterance utilities are possibly attenuated by utterance costs.[2]

$$U(u; s) = \log P_{lit}(s \mid u) - \text{cost}(u) \qquad [19]$$

Whether an utterance u is literally true/assertable for a given state s, is defined by the denotation function $[\![u]\!]$, that maps from an utter-

1 The code and all data from this chapter are publicly available on OSF: https://osf.io/acny6/?view_only=6deab67f1aaf494aaae0c12466670487.

2 Instead of using additive utterance cost (added to the utility of an utterance), it is also possible to induce utterance preferences for the speaker via the prior over utterances, $P(u)$. For example by setting $P(u; \text{cost}(u)) \propto \exp(-\text{cost}(u))$ where $\text{cost}(u)$ returns non-negative values. The choice between these two options influences whether the induced utterance preferences are influenced by the rationality parameter α (for details see Scontras et al., 2018, Appendix Ch.3).

ance u to the set of states in which u is assertable. Details on the conditions that need to be fulfilled for state s so that u is true/assertable in s, and thus $s \in [\![u]\!]$, are given below. The easier it is for a literal interpreter (Equation [4]) to determine the speaker's intended state, the larger the utility of an utterance is, and so, the more likely the speaker is to choose the respective utterance as description of the given state.

$$P_{\text{lit}}(s \mid u) \propto \delta_{s \in [\![u]\!]} \cdot P_{\text{prior}}(s) \qquad [4 \text{ revisited}]$$

Put differently, an utterance that applies to *many* states makes it more difficult for an interpreter to discriminate the speaker's intended state among all those states that the speaker possibly refers to, which is the set of states returned by the denotation function of the selected utterance ($[\![u]\!]$) . Contrary to that, an utterance that only applies to *few* states, leaves much less options to the interpreter, who will thus be more likely to infer the speaker's intended state.[3]

Note that for conceptual reasons, we slightly adapted the utility of utterances that we just described. As this adaptation concerns a detail that is not essential for the general understanding of the model, it is explained in the infobox below.

Utility of utterances

As mentioned in the main text, the speaker's utility of an utterance u (for a state s to be described) is greater, the fewer the number of states that can truthfully be described by u, due to $P_{\text{lit}}(s \mid u)$ being larger. The way that we sample model states implies that conjunctions will be the most informative type of utterance, followed by literals, which are in turn followed by conditionals (literals with 'might' are assertable for any state). Although the number of states for which an utterance of a certain type is assertable will be in the same range as the number of states that may truthfully be described by another utterance of the same type (e.g., 'if the green block falls, the blue block falls' vs. 'if the blue block falls, the green block falls'), these numbers will most likely not be identical. That is, the speaker would likely have a preference between utterances of the same type which is, however, not consistent when repeatedly sampling sets of model states of the same size. Therefore, we adapt utterance utilities as follows:

$$U(u; s) = \log P_{\text{lit}}(s \mid u') - \text{cost}(u)$$

3 For an example consider the utterance 'all' vs. 'some'. Saying 'I ate all cookies' excludes all possibilities except for me having eaten all n available cookies whereas when saying 'I ate some cookies' I may have eaten any number of cookies between 1 and n. That is, 'all' is more informative than 'some'.

where u' denotes the most informative utterance of the same type as u that is assertable in state s. Therefore, if any two conditionals are assertable in a state s (e.g., $B \rightarrow G$ and $G \rightarrow B$), the predicted utterance choice probability for the speaker will be identical for both conditionals.

It remains to specify the set of alternative utterances, their assertability conditions (the literal meaning) and the definition of states. A state is defined as a pair of a probability table (t) over two binary variables, B and G, that denote whether or not the blue, respectively the green block, will fall in a given situation, and a (latent) causal relation (r). The latter defines the shape of the considered probability tables. When $r = B \perp\!\!\!\perp G$, t is a probability distribution over two probabilistic *independent* random variables, B and G, that is, $P(Bb \mid g) = P(b \mid \neg g) = P(b)$. Contrary to that, when r refers to a dependent relation, the falling of the blue block may, for instance, make the green block fall as well ($r = B \overset{++}{\rightsquigarrow} G$), and thus, $P(b \mid g) \neq P(b \mid \neg g)$. That is, a state represents a relation r and (probabilistic) beliefs about the four possible combinations of outcomes (as judged in the PE-task), corresponding to a vector t of probabilities (Equation [20]) that sum up to one.

$$t = \langle P(b,g), P(b,\neg g), P(\neg b,g), P(\neg b,\neg g)\rangle = \langle w_{bg}, w_b, w_g, w_\varnothing \rangle \quad [20]$$

Note that the speaker is modeled to communicate only t explicitly while r is communicated implicitly: utterance utilities are only dependent on probability tables t, not on relations r, since the assertability conditions, given by the denotation function $[\![u]\!]$ (details below), do not depend on r.

Following the experimental setup, the speaker model includes all 20 utterances that participants could form in the UC-task (without using the free typing option). There is one little difference with respect to the set of alternative utterances used in Chapter 3. Here, we do not combine literals with 'likely' but with the weaker expression 'might' (e.g., 'the blue block might fall') which operates like a fall-back option; 'might ϕ' is assertable whenever $P^{(s)}(\phi) > 0$. The set of alternative utterances available in our speaker model thus comprises four different types: conditionals, conjunctions, literals, and literals combined with 'might'. Utterances are defined to be literally true or assertable with respect to a given state s, when the probability corresponding to that utterance and derived from the probability table t of state s is larger than a threshold θ, a free parameter of the model. For the definition of the semantics (i.e., the assertability conditions of utterances), see Table 10, adapted from Table 3 in Chapter 3. The conjunction, 'both blocks fall', for instance, corresponds to the joint probability

utterance type	assertability in state s	example: utterance u	assertability u in state s
conjunction	$P^{(s)}(\phi, \psi) \geq \theta$	'blue falls but green does not'	$P^{(s)}(B = b, G = \neg g) \geq \theta$
literal	$P^{(s)}(\phi) \geq \theta$	'blue falls'	$P^{(s)}(B = b) \geq \theta$
conditional	$P^{(s)}(\psi \mid \phi) \geq \theta$	'if blue falls green does not fall'	$P^{(s)}(G = \neg g \mid B = b) \geq \theta$
might + literal	$P^{(s)}(\phi) > 0.5$	'green might not fall'	$P^{(s)}(G = \neg g) > 0$

Table 10: Types of utterances with corresponding assertability conditions and an abbreviated example (e.g., 'green' = 'the green block'), ordered from most informative utterance on top to least informative at the bottom. For conditionals and conjunctions, $\phi \neq \psi$.

$P(B = b, G = g)$. For a state $s = \langle r, t \rangle$, we write $P^{(s)}(B = b, G = g)$ to refer to this probability as given by the probability table t of state s. For literals, with and without 'might', we consider the relevant marginal probabilities given by s (e.g., $P^{(s)}(B = b)$ for utterance 'the blue block falls/might fall') and for conditionals we consider the relevant conditional probabilities (e.g., $P^{(s)}(G = g \mid B = b)$ for utterance 'if blue falls green falls').

Having defined model states — pairs of probability tables t and an associated a causal relation r — the set of alternative utterances ($\mathcal{U}$) and the assertability conditions in form of the denotation function $[\![u]\!]$ (for each $u \in \mathcal{U}$), we are good to compute speaker predictions for each state, $P_S(u \mid s)\ \forall s \in S$, given fixed values for the free parameters θ and α. What remains to be defined is a prior distribution over model states from which we sample a finite set, the set of model states S.

We assume that, when communicating their beliefs about a shown block arrangement, speakers mainly draw on their general experiences with objects similar to the blocks shown in the experiment. That is, we essentially stick to the default state prior as defined in Chapter 3, corresponding to a generic case of communication about two binary variables which may or may not stand in a causal relationship ($B \perp\!\!\!\perp G$: independence, $\rightsquigarrow$: dependence, superscript indicates truth/falsity, e.g., $B \overset{-+}{\rightsquigarrow} G$ means that the falsity of B ($\neg b$) makes the truth of G (g) more likely). Each pair $\langle r, t \rangle$ is generated by first sampling a causal relation r and then the probability table t is sampled based on r. More precisely, for the independent relation, we sample two probabilities, $P(B = b)$ and $P(G = g)$ and for the dependent relations, we sample three values, causal power (of the cause to provoke the effect), noise (the probability of the effect to be provoked in absence of the cause) and the marginal probability of the cause. These two, respectively three values are then combined into a joint probability distribution $P(B, G)$. For convenience, Figure 5 and Table 6 from Chapter 3, which show the corresponding sampling procedure, are repeated here as Figure 30 and Table 11.

While in Chapter 3, dependent states (e.g., with $r = A \overset{++}{\rightsquigarrow} C$) were sampled with high causal power ($\sim$ beta(10, 1)) and low noise ($\sim$

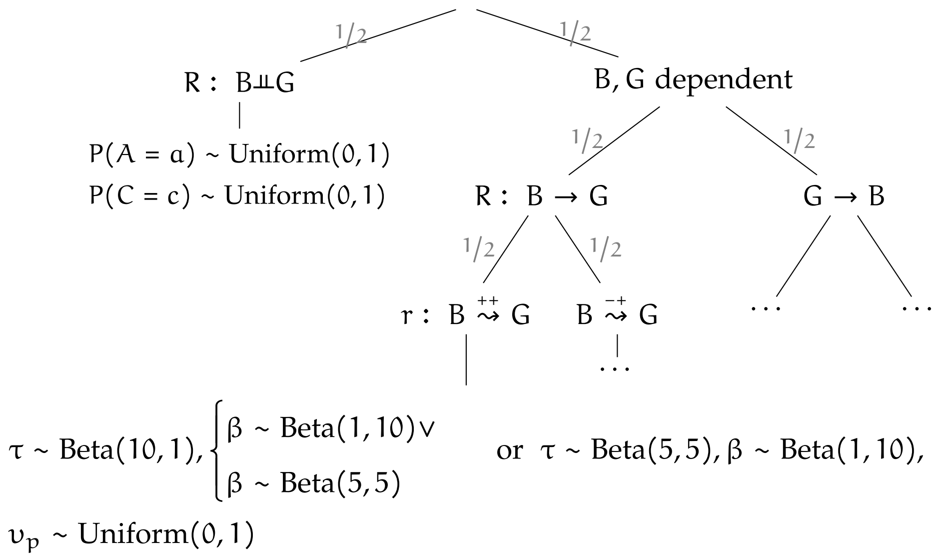


Figure 30: Graphical representation of the procedure for sampling a state s from the prior in the default context. τ denotes causal power, β noise and υ_p the respective marginal probability of the cause to be present.

instance causal relation (r)	υ_p	υ_c	β
$B \overset{++}{\rightsquigarrow} G$	$P^{(s)}(B = b)$	$P^{(s)}(G = g \mid B = b)$	$P^{(s)}(G = g \mid B = \neg b)$
$B \overset{-+}{\rightsquigarrow} G$	$P^{(s)}(B = \neg b)$	$P^{(s)}(G = g \mid B = \neg b)$	$P^{(s)}(G = g \mid B = b)$
$G \overset{++}{\rightsquigarrow} B$	$P^{(s)}(G = g)$	$P^{(s)}(B = b \mid G = g)$	$P^{(s)}(B = b \mid G = \neg g)$
$G \overset{-+}{\rightsquigarrow} B$	$P^{(s)}(G = \neg g)$	$P^{(s)}(B = b \mid G = \neg g)$	$P^{(s)}(B = b \mid G = g)$

Table 11: Probabilities $(\upsilon_p, \upsilon_c, \beta)$ that define the joint probability distribution of a state s, $P^{(s)}(B, G)$, for each instance of a dependent causal relation. υ_p is the prior probability of the cause, υ_c is the conditional probability of the effect to be true when the cause is true and β is the power of the unmodeled variables to provoke the effect which corresponds to the conditional probability of the effect to be true when the explicitly modeled cause is false.

beta(1, 10)), we here also sample dependent states with values for causal power ≈ 0.5 (~ beta(5, 5)) while noise remains small (~ beta(1, 10)) and states with large causal power (~ beta(10,1)) while values for noise are ≈ 0.5 (~ beta(5, 5)). The reason for this choice is that participants' beliefs (as given by the slider ratings from the PE-task) sometimes deviated from the default case (e.g., in if_1-trials, some participants indicated that the green block would fall if the blue block falls but that it might also fall if the blue block does not fall, which corresponds to larger noise values). Concretely, we forward sample 500 states, forming the overall set of model states, S.

The last bit that is missing in order to compute the model's prediction for particular contexts, each corresponding to an experimental condition, that is, a scene of block arrangements, is the (sub-)set of model states that will be considered — and if so, to what extent — as a basis for the model's prediction for a particular context. We will consider this next.

7.1.2 *Model Predictions*

For each context (i.e, the 13 test stimuli), we aim to predict a categorical distribution over utterances, $P_S(u \mid C_i)$. The model's prediction may be defined as the average prediction across a set of model states, S_{C_i}, that is representative for context C_i, displayed in Equation [21].

$$P_S(u \mid C_i) = {}^1/_{|S_{C_i}|} \cdot \sum_{s \in S_{C_i}} P_S(u \mid s) \qquad [21]$$

The most straightforward option is to define S_{C_i} as the set probability tables observed in context C_i (D_i^{PE}) which is reasonably the best representation of our contexts that we have. In that case the model's prediction for C_i would come down to the average prediction across all $d_{i,j}^{PE} \in D_i^{PE}$, shown in Equation [22], where the sum iterates over participants.

$$P_S(u \mid C_i) = {}^1/_{|D_i^{PE}|} \cdot \sum_{j=1}^{|D_i^{PE}|} P_S(u \mid d_{i,j}^{PE}) \qquad [22]$$

However, it cannot be taken for granted that $D_i^{PE} \subset S$ since S is a discrete set of n forward sampled model states.

One possibility how to circumvent this problem would be to use for each empirically observed probability table $d_{i,j}^{PE}$ the single most similar model state. Similarity may for instance be defined in terms of the Kulback-Leibler divergence (Kullback & Leibler, 1951), a distance measure for probability distributions. Since we here adopt a population-level approach, meaning that we aim to model the average probability of each utterance to be selected given a context C_i, instead of modeling utterance choices for each participant individually, we define model predictions as shown in Equation [23].

$$P_S(u \mid C_i) = \sum_{s \in S} P(s \mid C_i) \cdot P_S(u \mid s) \qquad [23]$$

That is, to make an RSA-speaker prediction for a context C_i, we will consider all model states (i.e., $S_{C_i} = S$) and weigh the model's predictions for each state to the extent that it represents context C_i. This means that we need to focus on the likelihood $P(s \mid C_i)$, defined in terms of the data-generating model. In Chapter 6 we modeled the data from the PE-task, that is, participants' slider ratings, by means of a Dirichlet regression model. Posterior predictive checks revealed that for some contexts the model was still surprised about the data it had been trained on (see Figure 46–47 in Appendix A), in particular for relation conditions if_1, if_2 and prior conditions U and U^-. Since these are, however, particularly relevant for us — it is in these conditions where we mainly expect participants to select conditionals — we will reconsider and adapt the data-generating model accordingly in the next section.

7.1.2.1 *Data-generating model to define* $P(s \mid C_i)$

The likelihood $P(s \mid C_i)$ — corresponding to the extent to which a model state $\langle r, t \rangle$ represents a certain context — is defined by Equations [24]–[26]. We will consider $P(r \mid C_i)$ and $P(t \mid C_i)$ in turn, starting with the former.

$$P(s = \langle r,t \rangle \mid C_i) = {}^{P(C_i \mid s = \langle r,t \rangle) \cdot P(s = \langle r,t \rangle)}/_{P(C_i)} \quad \text{with } s \in S \qquad [24]$$

$$P(C_i \mid r,t) \propto P(r \mid C_i) \cdot P(t \mid C_i) \cdot P(C_i) \qquad [25]$$

$$\Rightarrow P(s = \langle r,t \rangle \mid C_i) \propto P(r \mid C_i) \cdot P(t \mid C_i) \cdot P(s = \langle r,t \rangle) \qquad [26]$$

CAUSAL RELATIONS. $P(r \mid C_i)$ is the prior probability of how well a causal relation r from our RSA-model ($r \in \{B {\perp\!\!\!\perp} G, B \overset{++}{\rightsquigarrow} G, B \overset{-+}{\rightsquigarrow} G, G \overset{++}{\rightsquigarrow} B, G \overset{-+}{\rightsquigarrow} B\}$) represents a particular context C_i. The scenes were created so that in the independent trials, the falling of one block has nothing to do with the falling of the other block, and in the dependent trials, the falling of the blue block will always make the green block fall. Based on our analysis of the behavioral data, we here assume that participants fully grasped the difference between the independent and dependent conditions and, thus, set $P(r = B {\perp\!\!\!\perp} G \mid C_i) = 1$ for independent and $P(r = B {\perp\!\!\!\perp} G \mid C_i) = 0$ for dependent contexts C_i. For the dependent contexts, the probability mass is thus distributed among the four dependent causal relations. The slider ratings observed in the dependent contexts aligned at large, but not entirely, with what the dependent scenes were meant to communicate, namely that the green block would only fall if the blue block falls (in if_1 contexts) and

type C_i \ r	$B \perp\!\!\!\perp G$	$B \overset{++}{\rightsquigarrow} G$	$G \overset{++}{\rightsquigarrow} B$	$B \overset{-+}{\rightsquigarrow} G$	$G \overset{-+}{\rightsquigarrow} B$
independent	1	0	0	0	0
if_1	0	1/4	1/4	1/4	1/4
if_2	0	1/4	1/4	1/4	1/4

Table 12: Prior distributions $P(r \mid C_i)$ for the three types of contexts; the independent contexts, and the two types of dependent contexts, if_1 and if_2.

that the green block might possibly fall without that the blue block would fall (if_2 contexts). Therefore, for the dependent contexts, we choose a flat prior distribution over the four causal relations and set $P(r = B \perp\!\!\!\perp G \mid C_i) = 0$; Table 12 displays the full distributions $P(r \mid C_i)$.

PROBABILITY TABLES. Let us now turn to $P(t \mid C_i)$, the probability of a joint distribution table t given a particular experimental context which we previously modeled to be Dirichlet distributed (see Chapter 6) . One potential problem with the Dirichlet model (beside some of the posterior predictive checks being unsatisfactory) is that the observed data has to be smoothed such that all elements of the slider ratings $d^{PE}_{i,j} = \langle w_{bg}, w_b, w_g, w_\varnothing \rangle$ lie within the open interval (0,1).[4] The smoothing has the side-effect that on the one hand all conditional probabilities are defined, but on the other hand the new values of the smoothed conditional probabilities do not necessarily represent what participants might have in mind. For example, when $P(b) = 0$, $P(g \mid b)$ is not defined, but when we set $P(b, g)$ and $P(b, \neg g)$ both to 10^{-6}, that is, $P(b) \approx 0$, the conditional probability $P(g \mid b)$ becomes 0.5. It is, however, doubtful whether this is reasonable. Consider, for instance the two probability tables $t_1 = \langle w_{bg} = 0, w_b = 0, w_g = 0.5, w_\varnothing = 0.5 \rangle$ and $t_2 = \langle w_{bg} = 0.25, w_b = 0.25, w_g = 0.25, w_\varnothing = 0.25 \rangle$. When the former is smoothed, the conditional probabilities $P(g \mid b)$ of both distributions are 0.5, but this seems to be much more appropriate for t_2 than for t_1.

Therefore, instead of modeling each of the four slider ratings separately, we model the distributions $d^{PE}_{i,j}$ similarly to how we generated model states for the RSA-model. In short, for independent contexts, we assume that participants have a belief about $P(b)$ and $P(g)$ and about $P(b, g)$ which is approximately equal to $P(b) \cdot P(g)$. These three values fully define a joint probability distribution $P(B, G)$. For dependent contexts, we assume that participants on the one hand have a belief about the probability of the antecedent-block (e.g., $P(b)$) and on the other hand about the two conditional probabilities about the prob-

4 We do this by adding a small value $\epsilon = 1 \cdot 10^{-6}$ element-wise to vectors $d^{PE}_{i,j}$ and then normalize by element-wise division by $(1 + 4 \cdot 10^{-6})$.

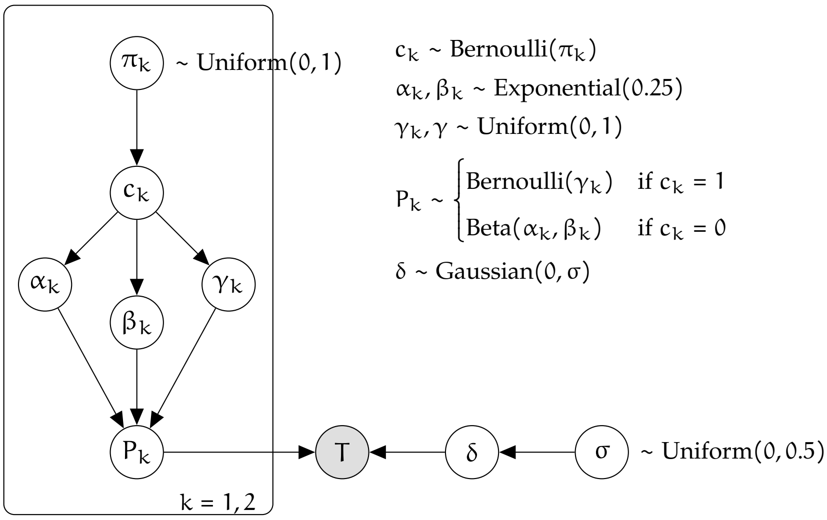


Figure 31: Graphical model of the observed probability tables (T) in the independent contexts. P_k denotes the probability of the blue (e.g., $k = 1$), respectively the green block (e.g., $k = 2$), to fall. These probabilities follow a zero-one inflated beta distribution: with probability π_k, p_k is either 0 or 1. If that is the case, the probability for that block to fall ($p_k = 1$) equals γ_k. If instead $p_k \in (0, 1)$, it is distributed according to a beta distribution with shape parameters α_k, β_k. We assume the observed probability tables t to be noisy with respect to the 'perfectly' independent values (see Equations [27]–[30]). This is realized by setting $P(b, g) - P(b) \cdot P(g)$ to a small value $\neq 0$, using random variable δ; for details see infobox in main text.

ability of the consequent-block to fall *given that* the antecedent-block falls or does not fall(e.g., $P(g \mid b), P(g \mid \neg b)$). Let us now consider both cases in more detail, starting with the independent contexts.

$P(t \mid C_i)$ FOR INDEPENDENT C_i. As mentioned above, in the independent trials, the probability of each block to fall is considered to be determined independently of the respectively other block. That is, only two values are necessary to get a completely defined joint probability distribution over the two random variables B and G, namely the probability that the blue block falls, $P(B = b)$, and the probability that the green block falls, $P(G = g)$, which were systematically manipulated in the experiment. Under the assumption that B and G are probabilistically independent, the joint probability table $t = \langle w_{bg}, w_b, w_g, w_\varnothing \rangle$, derived from $P(b)$ and $P(g)$, is given by Equations [27]–[30]:

$$w_{bg} = P(b, g) = P(b) \cdot P(g) \quad [27]$$
$$w_b = P(b, \neg g) = P(b) \cdot (1 - P(g)) \quad [28]$$
$$w_g = P(\neg b, g) = (1 - P(b)) \cdot P(g) \quad [29]$$
$$w_\varnothing = P(\neg b, \neg g) = (1 - P(b)) \cdot (1 - P(g)) \quad [30]$$

We consider the probability tables observed in the independent conditions to be noisy versions of these 'perfectly independent' proba-

	δ	P(b)				P(g)			
context	σ	π	γ	α	β	π	γ	α	β
ind:UH	0.04	0.07	0.86	3.33	2.58	0.27	0.84	2.7	1.24
ind:HH	0.03	0.28	0.88	2.57	1.13	0.42	0.95	3.91	1.33
ind:HL	0.05	0.43	0.95	4.2	1.73	0.11	0.36	2.09	2.25
ind:UL	0.05	0.09	0.66	3.13	2.22	0.12	0.17	1.81	1.9
ind:LL	0.05	0.42	0.03	1.05	1.62	0.14	0.21	1.83	2.14

Table 13: Expected values of parameters of distributions (rounded to 2 digits) fitted to the PE-data from the *in*dependent trials.

bility distributions. Figure 31 shows a graphical model of the probability tables observed in the independent trials where shaded circular nodes represent observed, continuous variables while unshaded circular nodes represent latent, continuous variables (Lee & Wagenmakers, 2014). The marginal probabilities, that is the probability of the blue, respectively the green block to fall (represented by P_k) each follow a zero-one inflated Beta distribution (ZOIB): some participants may be certain with respect to the falling of a block, which they indicate by setting the respective sliders so that they sum up to 0 or 1 (e.g., $w_{bg} = 0.5$, $w_b = 0.5 \Rightarrow P(b) = 1$). The proportion of participants who shows this behavior is modeled by π_k and *given* that 0 or 1 is selected, the probability of a success (i.e., $p_k = 1$) is modeled by γ_k. Non-extreme values within the open interval (0,1) are modeled by a beta distribution with parameters α_k and β_k. The noise added to the optimal value for $P(b, g)$ — assuming B and G are independent — depends on the normally distributed random variable δ and the two marginal probabilities, $P(b)$ and $P(g)$ (for details, see infobox below). For each of the five independent contexts, we fit the parameters of each of the two zero-one inflated beta distributions as well as the standard deviation σ of the normal distribution of random variable δ to the respectively observed data.[5] Table 13 displays the mean posterior values for each fitted parameter which we use to parameterize the likelihood functions (Figure 31). These were estimated in one swoop, using a custom model written in WebPPL.

With the definition of the likelihood functions, we can then compute the likelihood of an arbitrary probability distribution $P(B, G)$ to represent probabilistic beliefs with respect to an independent experimental condition. Note that noise variable δ can be ignored when we only consider the likelihood of model states with independent relation as possibly representing an independent context ($P(s = \langle r =$

5 The posterior distributions are computed with WebPPL (using method 'incrementalMH') and approximated by 5000 MCMC-samples drawn for 4 chains with a lag of 10 after a burn-in period of 50,000 samples. To check for convergence, we use the posterior package in R to compute $\hat{R}$ as defined in Vehtari et al. (2021). $\hat{R}$ is smaller than the commonly used threshold of 1.1 for all parameters.

$B \perp\!\!\!\perp G, t\rangle \mid C_i)$ for C_i = ind:XY). The difference, $P^{(s)}(b, g) - P^{(s)}(b) \cdot P^{(s)}(g)$ will always be 0 for these states, and, thus, the corresponding part of the likelihood will be identical across all considered states. If we did not allow noise, that is, if we assumed that in independent contexts, participants beliefs correspond to probability distributions of two probabilistic independent variables, B and G, so that $P(b, g) = P(b) \cdot P(g)$, $P(d^{PE}_{i,j} \mid C_i)$ would be 0 for all $d^{PE}_{i,j}$ where the joint probability $P(b, g)$ does *not* equal the product of the two prior probabilities, $P(b)$ and $P(g)$ — which is essentially the case for all observed $d^{PE}_{i,j}$, except when $P(b, g) = 0$ or $P(b, g) = 1$.

Let us turn to the data of the dependent contexts next.

Noisy independence

When $P(b)$ and $P(g)$ have fixed values, the values that $P(b, g)$ can possibly take on are restricted. The theoretically possible values of $P(b, g)$ are given by the equation below, where $p_b = P(b), p_g = P(g)$. Note that when the sum of the marginals exceeds 1, $P(b, g)$ does not only have a maximal, but also a minimal value (> 0) since the sum of $P(b, g)$, $P(b, \neg g)$ and $P(\neg b, g)$ must not exceed 1.

$$\min(p_b, p_g) \geqslant P(b, g) \geqslant \begin{cases} p_b + p_g - 1 & \text{if } p_b + p_g > 1 \\ 0 & \text{else} \end{cases}$$

Let m_{sub} denote the positive value that can maximally be subtracted from the 'perfectly' independent value, $P^* = P(b) \cdot P(g)$, and let m_{add} denote the positive value that can maximally be added to P^*.

EXAMPLE. $P(b) = 0.7, P(g) = 0.8 \Rightarrow P^* = 0.7 \cdot 0.8 = 0.56$, the upper bound for $P(b, g)$ is $\min(0.7, 0.8) = 0.7$, its lower bound is $0.7 + 0.8 - 1 = 0.5$. Further, $m_{sub} = P^* -$ lower bound $= 0.56 - 0.5 = 0.06$ and $m_{add} =$ upper bound $- P^* = 0.7 - 0.56 = 0.14$. Then, keeping $P(b)$ and $P(g)$ fixed, $P(b, g)$ is set as follows:

$$P(b, g) = \begin{cases} P^* + \delta & \text{if } \delta > 0 \land \delta < m_{add} \textbf{ or} \\ & \delta < 0 \land \delta > -m_{sub} \\ P^* + m_{add} & \text{else if } \delta \geqslant 0 \\ P^* - m_{sub} & \text{else } (\delta < 0 \land \delta \leqslant -m_{sub}) \end{cases}$$

For a sampled value of δ with $\delta > 0 \land \delta < m_{add}$ (e.g., $\delta = 0.03$), $P(b, g)$ would be set to $P^* - \delta$ (e.g., 0.56 - 0.03 = 0.53). For a sampled value of δ with $\delta < 0 \wedge \delta \leqslant -m_{sub}$ (e.g., $\delta = -0.07$), $P(b, g)$ would be set to $P^* - m_{sub}$ (e.g., 0.56 - 0.06 = 0.5).

$P(t \mid C_i)$ FOR DEPENDENT C_i. In the dependent contexts, three values fully define a probability distribution over variables B and G, two conditional probabilities and a marginal probability, for instance, $P(G = g \mid B = b), P(G = g \mid B = \neg b)$ and $P(B = b)$. The corresponding joint probability distribution is then derived as shown in Equation [31]:

$$\begin{array}{ll} \langle P(b,g) = P(b) \cdot P(g \mid b), & P(b,\neg g) = P(b) \cdot (1 - P(g \mid b)), \\ P(\neg b, g) = (1 - P(b)) \cdot P(g \mid \neg b), & P(\neg b, \neg g) = (1 - P(b)) \cdot (1 - P(g \mid \neg b))\rangle \end{array} \quad [31]$$

Like the prior probabilities $P(b)$ and $P(g)$ in the independent contexts, these three probabilities are assumed to follow a zero-one inflated beta distribution, displayed in Equations [32]–[37], together with the prior distributions of the parameters.

Marginal Probability (P_1) :

$P_1 = P(b) \sim \text{ZOIB}(\pi_1, \gamma_1, \alpha_1, \beta_1)$, therefore:

$$c_1 \sim \text{Bernoulli}(\pi_1) \quad [32]$$

$$c_1 = 1 \Rightarrow P(b) = 0 \vee P(b) = 1;\ c_1 = 0 \Rightarrow P(b) \in (0, 1)$$

$$P_1 = P(b) \sim \begin{cases} \text{Bernoulli}(\gamma_1) & \text{if } c_1 = 1 \\ \text{Beta}(\alpha_1, \beta_1) & \text{if } c_1 = 0 \end{cases} \quad [33]$$

Conditional Probabilities $(P_2$ and $P_3)$:

$$P_2 = P(g \mid b) \sim \begin{cases} \text{ZOIB}(\pi_k, \gamma_k, \alpha_k, \beta_k) & \text{if } P(b) \neq 0 \\ \text{undefined} & \text{else} \end{cases} \quad [34]$$

$$P_3 = (g \mid \neg b) \sim \begin{cases} \text{ZOIB}(\pi_k, \gamma_k, \alpha_k, \beta_k) & \text{if } P(b) \neq 1 \\ \text{undefined} & \text{else} \end{cases} \quad [35]$$

Prior distributions:

$$\pi_k, \gamma_k \sim \text{Uniform}(0, 1) \quad [36]$$

$$\alpha_k, \beta_k \sim \text{Exponential}(0.25) \quad [37]$$

with $k \in [1, 2, 3]$

For the two conditional probabilities, there is a particularity worthwhile mentioning. When the marginal probability of the blue block to fall is 0 or 1, $P(g \mid b)$, respectively $P(g \mid \neg b)$, is undefined (see Equations [34] and [35]). This is, however, not problematic since in that case, the marginal probability and the respectively other (defined) conditional probability are sufficient to fully define the joint probability; for an example, see Equation [38] where $P(b) = 0$ (first

	$P(g \mid b)$				$P(g \mid \neg b)$				$P(b)$			
context	π	γ	α	β	π	γ	α	β	π	γ	α	β
if$_1$:LI	0.23	0.79	4.85	2.93	0.48	0.21	1.51	1.36	0.39	0.05	1.52	2.08
if$_1$:UI	0.36	0.91	3.74	1.78	0.53	0.16	1.49	1.53	0.12	0.91	4.01	2.42
if$_1$:U¯I	0.30	0.93	3.41	1.69	0.55	0.20	1.17	1.27	0.07	0.73	2.77	1.95
if$_1$:HI	0.22	0.95	4.12	1.67	0.27	0.38	2.36	2.07	0.42	0.95	5.63	2.04
if$_2$:LL	0.31	0.75	2.45	2.05	0.48	0.07	1.03	1.62	0.12	0.5	2.01	1.98
if$_2$:U¯L	0.22	0.5	2.03	1.48	0.47	0.08	3.17	4.00	0.22	0.71	3.13	1.77
if$_2$:UL	0.22	0.67	2.17	1.67	0.45	0.18	2.89	3.4	0.23	0.96	2.42	1.24
if$_2$:HL	0.24	0.65	1.57	1.34	0.41	0.15	3.31	3.70	0.32	0.97	6.51	2.12

Table 14: Expected values (rounded to 2 digits) of parameters of ZOIB-distributions fitted to $P(g \mid b), P(g \mid \neg b), P(b)$ from the PE-data of the dependent trials.

row) and only the conditional probability $P(g \mid \neg b)$ is required to derive the distribution $P(B, G)$.

$$\begin{aligned}\langle P(b,g) &= 0, & P(b,\neg g) &= 0,\\ P(\neg b, g) &= (1 - P(b)) \cdot P(g \mid \neg b), & P(\neg b,\neg g) &= (1-P(b)) \cdot (1 - P(g \mid \neg b))\rangle\end{aligned} \quad [38]$$

Taken together, we fit 3 ZOIB-distributions for each dependent context with 4 parameters each, $\alpha_k, \beta_k, \gamma_k$ and π_k ($k \in [1, 2, 3]$) to the respectively observed data from the PE-task.[6] Again, the expected values of each fitted parameter, that is, the mean posterior values, are used to parameterize the likelihood function $P(t \mid C_i)$ for dependent contexts C_i, displayed in Table 14. As for the independent contexts, these parameters were estimated in one swoop with a custom model written in WebPPL.

<u>SECTION SUMMARY</u> In this subsection, I introduced a new data-generating model for the PE-task data, considering the independent and dependent contexts separately. In the former, probability tables are assumed to be derived based on two ZOIB-distributed marginal probabilities ($P(b), P(g)$) and some noise so that the joint probability distribution $P(B, G)$ is noise-disturbed with respect to the probability distribution that it would take on if variables B and G were probabilistically independent. For the dependent contexts, probability tables are assumed to be derived based on three ZOIB-distributed probabilities, $P(b), P(g \mid b)$ and $P(g \mid \neg b)$. These distributions are

6 The posterior distributions are computed with WebPPL (using method 'incrementalMH') and are approximated by 5000 MCMC-samples drawn for 4 chains with a lag of 10 after a burn-in period of 50,000 samples. $\hat{R}$ is smaller than the commonly used threshold of 1.1 for all parameters. Values below 1.1 would indicate problems with the convergence of the MCMC-chains.

fitted to the original, *un*smoothed slider ratings observed in the PE-task. We then use the mean posterior values of the fitted parameters to define the likelihood functions for each context.

7.1.2.2 *Weights* $P(s \mid C_i)$

In the previous section, we introduced an alternative data-generating model for participants' slider ratings from the PE-task. For a comparison between the new ZOIB-model and the old Dirichlet-regression model, consider Figure 32 which displays the posterior predictive distributions for context if_1:UI, given the ZOIB-model (left panel) and the Dirichlet model (right panel), together with the observed slider ratings in that context. For this context — one of the most relevant contexts for us as it is among those for which we particularly expect participants to choose conditionals — the Dirichlet regression model does not capture the structure of the data very well, in particular not for w_{bg}. The ZOIB-model is not perfect either, but it does seem to be more adequate.[7]

The motivation for spelling out the data generating model was the need to get an appropriate probabilistic representation of each context. By means of this representation we can decide which of the RSA-model states shall be considered to what extent for the computation of the model's prediction for a context C_i, defined in Equation [23], repeated here from above.

$$P_S(u \mid C_i) = \sum_{s \in S} P(s \mid C_i) \cdot P_S(u \mid s) \qquad [23 \text{ revisited}]$$

The probabilistic representation of contexts are defined by the assumed likelihood functions, parameterized with the mean posterior values (see Tables 13 and 14), after having fitted them to the empirically observed data, the slider ratings from the PE-task.

EXAMPLE. For a state $s' = \langle r, t = \langle w_{bg} = 0.49, w_b = 0.01, w_b = 0.01, w_\varnothing = 0.49 \rangle\rangle$ to represent context if_1:UI, we need to compute

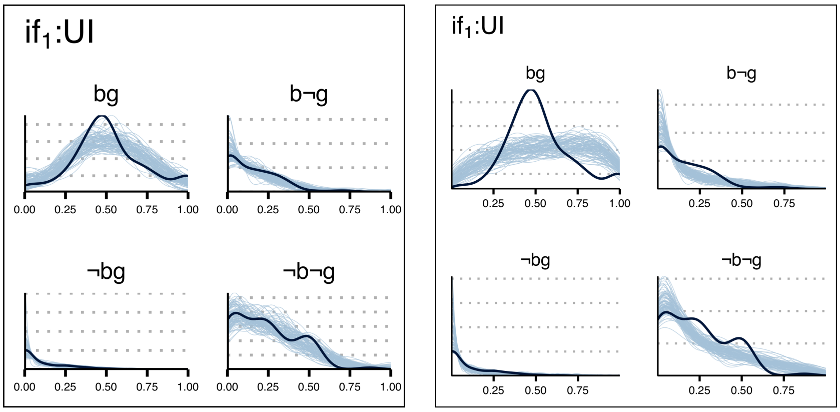
$$P^{(s')}(b) = 0.5 \sim \text{ZOIB}(\pi = 0.12, \gamma = 0.91, \alpha = 4.01, \beta = 2.42)$$

$$P^{(s')}(g \mid b) = 0.98 \sim \text{ZOIB}(\pi = 0.36, \gamma = 0.91, \alpha = 3.74, \beta = 1.78)$$

$$P^{(s')}(g \mid \neg b) = 0.02 \sim \text{ZOIB}(\pi = 0.53, \gamma = 0.16, \alpha = 1.49, \beta = 1.53)$$

Summing up the respective log-values gives us the log likelihood $P(s' \mid C_{if_1:UI})$, that is, the weight how much s' is taken into account for the model's prediction for context if_1:UI.

7 For the posterior predictive plots of the other contexts, see Figure 48–49 in Appendix A.



(a) ZOIB-model (b) Dirichlet regression model

Figure 32: Posterior predictive distributions for context if$_1$:UI: $P(\tilde{X} \mid D^{PE}_{if1:UI}) = \int P(\tilde{x} = \langle w_{bg}, w_b, w_g, w_\varnothing \rangle \mid \theta, C_{if_1:UI}) \cdot P(\theta \mid D^{PE}_{if1:UI}) d\theta$; θ denotes the set of model parameters.

7.1.3 *Baseline Models*

Before we come to the predictions of our RSA-speaker model in the next section, we will first introduce two baseline speaker models. These are useful sanity-checks to see whether the proposed model is better than these simpler models, thus justifying the additional complexity of the proposed model. The utterance-choice predictions for a context C_i are computed as before (see Equation [23 revisited]), except that the RSA-speaker, $P_S(u \mid s)$, is replaced by the respective baseline speaker.

7.1.3.1 *Random speaker model* $P_{S_{rnd}}$

The first baseline speaker model is the easiest and least complex, a random speaker who chooses each utterance uniform at random from the set of alternative utterances, defined in Equation [39].

$$P_{S_{rnd}}(u \mid s) = 1/20 \qquad [39]$$

$$P_{S_{rnd}}(u \mid C_i) = \sum_{s \in S} P(s \mid C_i) \cdot 1/20 = 1/20 \cdot 1 = 1/20 \qquad [40]$$

That is, even though the model states for which model predictions are likely taken into account differ across contexts ($P(s \mid C_i)$), this speaker makes the same predictions for all contexts (see Equation [40]) since its predictions do not depend on a given state, they are simply always the same.

7.1.3.2 *Literal speaker model* $P_{S_{lit}}$

As second baseline model, we consider a literal speaker who chooses randomly among true utterances, defined in Equation [41] where |

$U_{(s)}$ | denotes the size of the set of utterances that truthfully describe state s according to the semantics given in Chapter 3 and repeated above in Table 10.

$$P_{S_{lit}}(u \mid s) = {}^{\delta_{s\in[\![u]\!]}}/_{|U_{(s)}|} = {}^{\delta_{s\in[\![u]\!]}}/_{\sum_{u\in U}\delta_{s\in[\![u]\!]}} \tag{41}$$

$$P_{S_{lit}}(u \mid C_i) = \sum_{s\in S} P(s \mid C_i) \cdot P_{S_{lit}}(u \mid s) \tag{42}$$

If, for example, $s' = \langle r, t\rangle$, $t = \langle w_{bg} = 0.23, w_b = 0.75, w_g = 0.01, w_{\varnothing} = 0.01\rangle$ and assertability threshold $\theta = 0.9$, the literal speaker would choose each utterance that truthfully describes this state, with equal probability. For s' there are seven such utterances ($U^{(s')} = \{$'blue/green might fall', 'blue/green might not fall', 'if green falls, blue falls', 'if green does not fall, blue does not fall', 'blue falls'$\}$). Thus, $P_{S_{lit}}(u \mid s') = {}^1/_7$ for $u \in U^{(s')}$, for all other utterances, $P_{S_{lit}}(u \mid s') = 0$.

7.2 MODEL FITTING

The free parameters of the model are the rationality parameter α, the assertability (literal meaning) threshold θ (θ_{might} is fixed to 0) and utterance cost. We fit α and θ to the empirical data from the UC-task, keeping utterance cost fixed at 0 for now. That is, α and θ are tuned so that the overall log likelihood (summed across contexts) of the observed average selection probability of each utterance is maximized under the speaker's predictions $P_S(u \mid C_i, D^{UC}, D^{PE})$. Put differently, for each context C_i we consider the log likelihood of the data vector, containing the number of observed selections of each utterance in context C_i, assuming the speaker's predictions for this context, $P_S(u \mid C_i, D^{PE}, D^{UC})$ (a vector of probabilities of size 1×20).

The prior distributions for α and θ are given in Equations [43] and [44].

$$\log \alpha \sim \text{Gaussian}(\mu = 1.5, \sigma = 1) \tag{43}$$

$$\theta \sim \text{Beta}(\alpha = 4, \beta = 2) \tag{44}$$

Since for some utterances, in particular conjunctions, we observed quite low probability estimates in the PE-task for the event that participants subsequently described in the UC-task for the same shown scene, we chose a quite broad beta distribution for the assertability threshold θ so that low θ values are not *a priori* excluded.

7.3 MODEL RESULTS & DISCUSSION

The comparison between model predictions and empirical data is, on the one hand, used as a proof of concept to see whether the model is able to explain the observed data. On the other hand, we aim to see whether there are any systematic, insightful divergences between the model's predictions and the empirical data.

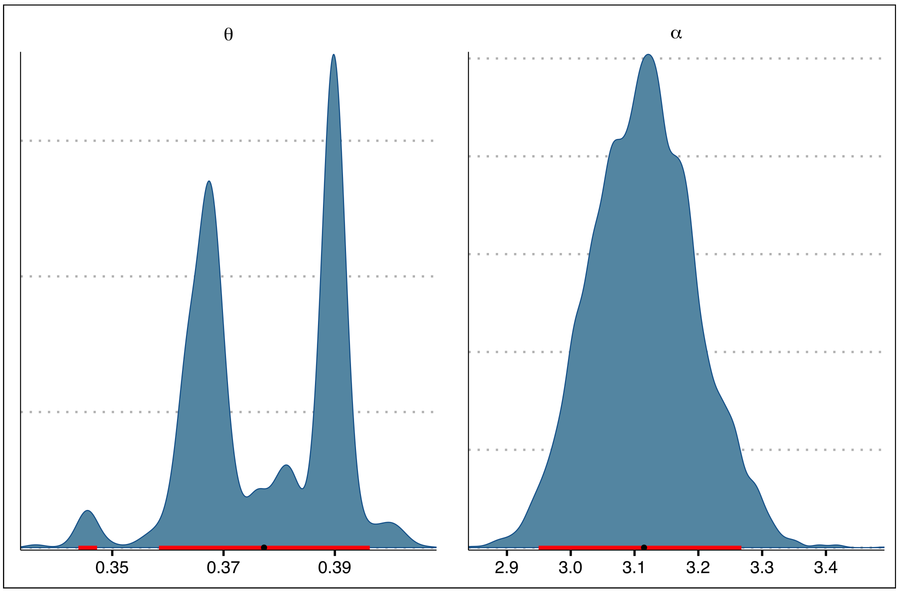


Figure 33: Approximated Posterior $P(\alpha, \theta \mid P_S, D^{UC}, D^{PE})$ with 95% highest density intervals, 5000 MCMC-samples (lag=10) after a burn-in period of 10,000 samples.

7.3.1 *Pragmatic speaker*

fitted parameters. Figure 33 shows the posterior distributions for the two fitted parameters α and θ for the pragmatic speaker model, computed based on 5000 MCMC-samples drawn with a lag of 10 for 4 chains after a burn-in period of 10,000 samples.[8] While for the empirical production data, the inferred best values for α are reasonable, the values for θ that explain the data best are very low (mean posterior 0.377, MAP 0.39) which is conceptually not very reasonable and strongly suggests that the model is missing something fundamental about the observed data. Yet, from a technical point of view, this result is reasonable since the data revealed that participants tended to select conjunctions — which were the most frequent utterances — to describe a shown scene, although they gave rather low estimates in the pe-task for the probability of the corresponding event. Thus, for the model to be able to select conjunctions in these cases at all, the assertability threshold must be correspondingly low.

comparison pragmatic speaker model vs. data. Figure 34 shows the predictions of the pragmatic speaker model with 95% high-

8 For θ the posterior looks bi-modal, indicating problems with the MCMC-samples. The value for $\hat{R}$ is not suspicious being ≈ 1.001, that is smaller than 1.01, the value above which it is likely that the MCMC-chains have not mixed. A diagnostic pairs-plot is shown in Figure 51 in Appendix A. I will not discuss this issue further since we can still draw inferences about the model's performance compared to the performance of the baseline models, even though the posterior samples cannot reliably be considered to correspond to the true values.

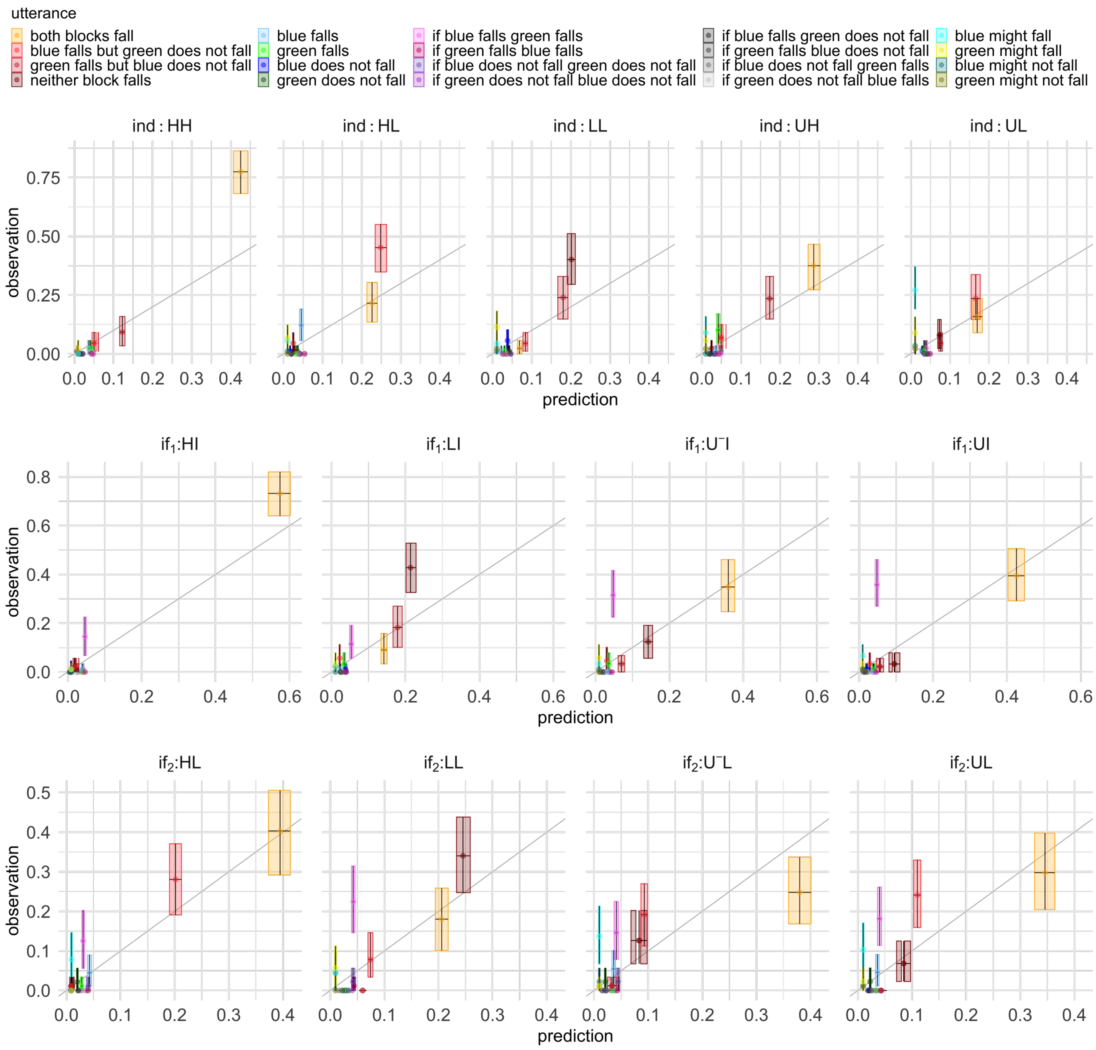


Figure 34: Prediction of the pragmatic speaker model (x-axis) with 95% highest density intervals, plotted against the empirically observed average selection frequencies (y-axis) for each utterance (color coded), with 95% bootstrapped confidence intervals.

est density intervals (HDIs) plotted against observed utterance selection frequencies with 95% bootstrapped confidence intervals. Overall, the model seems to capture at least parts of the observed data reasonably well, as most rectangles in Figure 34 cross the diagonal, which represents a perfect fit between model and data.

Let us turn to those cases where model predictions deviate from the empirically observed frequencies. Across all dependent contexts, the model particularly underestimates the use of the conditional 'if the antecedent-block falls the consequent-block falls' (henceforth $A \rightarrow C$).[9] Further, the use of the conjunction 'neither block falls' is underestimated in contexts if_1:LI and (less so) in if_2:LL, and slightly overestimated in context if_1:UI. The conjunction 'both blocks fall' is underestimated in contexts if_1:HI and overestimated in context if_2:U^{-}I. In contexts with relation if_2, except for if_2:LL, the model also underestimates the use of the conjunction 'the antecedent-block falls but the consequent-block does not fall' ('the blue block falls but the green block does not fall'). Lastly, we observe an underestimation of the two utterances 'the blue/green block might fall' more or less for all if_2 contexts. In the if_1 contexts, we also observe an underestimation of these utterances, although less strong and only for contexts if_1:UI and if_1:U^{-}I.

In summary we can say that for the dependent contexts there are three main observations: the underestimation of the conditional $A \rightarrow C$ as well as the over- and (depending on the context) also under- estimation of some conjunctions and the underestimation of utterances with 'might'.

Concerning the independent contexts, the picture is very similar. The model underestimates the use of the conjunctions 'both blocks fall', 'the blue block falls but the green block does not fall' and 'neither block falls' respectively in contexts ind:HH, ind:HL and ind:LL. Further, the two positive utterances with 'might' are also underestimated more or less in all independent contexts, except for ind:HH. Although for independent contexts, the predicted probabilities for each of the eight available conditionals are low (between 0.01 and 0.05), in sum the model however overestimates them — participants hardly selected conditionals at all in independent contexts.

7.3.2 *Baseline models*

literal speaker. Fitting the literal speaker model to the empirical data results in posterior θ values that are still smaller than they already were when fitting the pragmatic speaker model. The

9 Remember that for the purpose of a cleaner presentation, in all shown dependent contexts, the antecedent-block (the block on the upper platform) corresponds to the blue and the consequent-block to the green block. In the experiment, colors were randomly assigned for each participant and trial to the antecedent- and consequent-block.

Figure 35: Prediction of the literal speaker model (x-axis), with 95% highest density intervals (for some utterances, HDI-interval split into several non-overlapping intervals), plotted against the empirically observed average selection frequencies (y-axis) for each utterance (color coded), with 95% bootstrapped confidence intervals.

95% highest density interval (HDI) for θ is given by $(0.007, 0.012) \cup (0.034, 0.064)$ with a mean estimate of 0.045.[10] Figure 35 shows the posterior predictive distributions for the literal speaker plotted against the observed data with 95% bootstrapped confidence intervals. Unmistakably, this model does not align well with the observed data.

RANDOM SPEAKER. For the random speaker model it does not make sense to fit parameters α and θ since the predicted utterance choice probabilities will always be identical for all utterances, independently of the values for α and θ.

With an equal predicted probability of 1/20 for all 20 utterances, the random speaker model overestimates nearly all utterances — however only slightly. This is due to the fact that the utterances that were mostly selected by participants comprise only a handful of the 20 available utterances. However, these few utterances are clearly very strongly underestimated by the random speaker model, including the conditional $A \rightarrow C$, all four conjunctions and the utterance 'the blue block might fall'. The observed selection frequencies for the other utterances are thus very close to 0 and thereby not too far from the random speaker prediction of 0.05. When taking a closer look at participants' utterance selections we observe that the number of utterances that are on average selected by more than 5% of participants (in each context) is reduced to a maximum of 6 (context if$_2$:U$\bar{~}$L, minimal frequency is 0.06, maximal frequency is 0.25) and a minimum of only two in context ind:HH, namely 'the blue block does not fall but the green block falls' with a selection frequency of 0.09 and 'both blocks fall' with a selection frequency of 0.77.

7.3.3 *Model comparisons*

The posterior predictive distributions suggested that the pragmatic speaker model explains the observed production data better than the two baseline models. This result is confirmed by the values of the log likelihood of the data under the respective speaker model, shown separately for each context in Figure 36 (models with names ending with '.gamma' are discussed in next section). Except for one context (ind:UL), the pragmatic speaker model results in the largest, thus best values.

In context ind:UL, the pragmatic speaker predicts the utterance 'the blue block might fall' to be hardly selected at all (mean posterior value ≈ 0.01), whereas participants selected this utterance very often in this context (observed frequency of $24/89 \approx 0.27$). For the log likelihood of the observed data — given the Multinomial speaker model — it is particularly bad when a category that is frequently observed

10 See Figure 50 in Appendix A. $\hat{R} \approx 1.001$, thus it is smaller than the commonly used threshold 1.01.

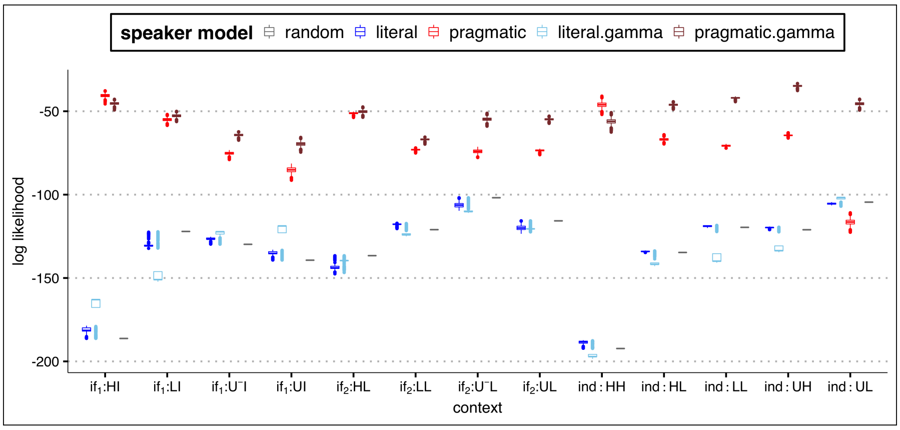


Figure 36: Boxplots of the log likelihood values of the observed utterance selection frequencies in each context for 5000 MCMC-samples from the posterior over free model parameters (α, θ and, for the extended model additional parameter γ (see Section 7.4)).

in the data, is predicted to be very unlikely. This is what happens in context ind:UL. Compared to the random and literal speaker models, the pragmatic speaker model reduces its prediction for the two utterances 'the blue/green block might fall' as this utterance is highly uninformative, which is why, whenever possible, the pragmatic speaker would choose a different utterance. However, compared to other contexts, in which the pragmatic speaker predicts these two utterances to be similarly unlikely, participants select them considerably more often in context ind:UL, in particular the utterance 'the blue block might fall', which has a selection frequency of 0.27 in context ind:UL. This is twice as much as the same utterance is, for instance, selected in context if_2:U¯L.

Further, we observe very similar results for the two baseline speaker models. Since participants often selected utterances although they had given quite low probability estimates in the PE-task for the corresponding outcome, the values for the assertability threshold are pushed towards very low values (for the literal speaker model) so that the model is able to select these utterances in those states contributing to the model's prediction for the respective context at all. A low assertability threshold, in turn, reduces the differences between the informativeness of utterances. When $\theta = 0$, that is, the speaker is not restricted at all in her utterance choice for any state, all utterances are equally informative or better to say uninformative — which in fact is a random speaker. The larger θ, the fewer states can truthfully be described by the most specific utterances, making these highly informative. For the literal speaker model, the *maximum a posteriori* (MAP) value for θ — the value that maximizes the log likelihood of the overall data, across all contexts — is as low as 0.01. Overall, the literal speaker model does not make any prediction much beyond a value of 0.05, its largest prediction over all posterior samples is as low as

0.058, which is very close to the prediction of the random speaker model.

7.3.4 *Discussion*

We proposed an RSA speaker-model for the use of conditionals as compared to non-conditional utterances and compared its predictions, as well as the predictions of two baseline speaker models, to the empirical production data that we collected in a behavioral experiment. The RSA speaker model seems to capture parts of the observed data but clearly does not capture it in its entirety.

Throughout all dependent contexts, we observed some systematic divergences between model predictions and the data. The model strongly underestimates the use of the conditional $A \rightarrow C$, where A refers to the antecedent-block (the upper block, in the stimuli shown here, this is the blue block) and C to the consequent-block (lower, here green block). Contrary to that, the conditionals $C \rightarrow A$, $\neg C \rightarrow \neg A$ and $\neg A \rightarrow \neg C$ are overestimated by the model. In the current setup of the model, there is effectively not much of a difference between these conditionals and the underestimated conditional, $A \rightarrow C$. If, for instance, the latter is assertable the conditional with swapped antecedent and consequent will most likely be assertable as well and thus, the model will predict both to be approximately equally likely. Participants, however, show a strong preference for the conditional $A \rightarrow C$ mentioning the antecedent-block in the antecedent and the consequent-block in the consequent, reflecting the time course as shown in the scenes; among the 157 conditionals selected in the UC-task, these amount to 155 ($\approx$ 99%), with only 6 of the 157 conditionals selected in independent contexts. This preference could be integrated into the model, for example by using a salience prior for utterances, so that some utterances are *per se* preferred over others. It is also conceivable to make it dependent on the causal relation of the speaker's target state. When $r = B \overset{++}{\rightsquigarrow} G$, the utterance $A \rightarrow C$ could, for instance, be modeled as more likely than the conditional with swapped antecedent and consequent whereas the opposite could hold for states with $r = G \overset{++}{\rightsquigarrow} B$.

There is, however, another, more severe, aspect concerning the use of conditionals as predicted by the model. While participants hardly selected conditionals at all in independent contexts, the model does not account for this difference in utterance choices depending on the relation of the respective context. The predicted probability for conditionals is roughly the same independently of the context. A possible reason contributing to this behavior might be the estimated low values for the assertability threshold θ (mean estimate ≈ 0.338). While large θ values render some utterances very informative (e.g., conjunctions) as they are only assertable in few states, smaller θ values decrease the difference in the informativeness between utterances. This is particularly important for the predicted probabilities in independent contexts. In these contexts, an assertable conditional always comes along with an assertable literal since $P^{(s)}(x \mid y) = P^{(s)}(x)$. Thus, the more informative conjunctions and literals are — for instance due to larger θ values — the smaller the predicted probabilities for conditionals should be in independent contexts.

The small *a posteriori* values for θ also need to be addressed further since they are conceptually unreasonable and thus suggest that the model is missing something fundamental in the structure of the observed data. They can be traced back to the quite prominently observed divergences between participants' slider ratings and their selected utterance for one and the same context. Participants showed a tendency to make a certain claim (e.g., 'both blocks fall') while providing rather low probability estimates for the described outcome. We make this observation particularly for conjunctions, for all other utterances that were selected by more than 1 or 2 participants, the corresponding probability estimates were reasonably high (see Figure 27). Therefore, it does not seem to be the case that participants simply failed to map their beliefs onto the sliders. In a sense participants were more conservative with respect to the beliefs they indicated through the slider ratings than they were when selecting utterances to describe the scenes. We speculated that their incentive to choose only utterances that they are really convinced of being true might not have been high enough so that they could have adopted the strategy to simply decide for one of the four possible and mutually exclusive outcomes and choose the respective conjunction straight away. The fact that it was *not* possible to claim something about one block while expressing uncertainty about the other unless using conditionals (e.g. by an utterance like 'the blue block might fall but the green bock does not'), may have further reinforced this. To test whether integrating the tendency to use this strategy — to describe the scene by claiming one of the four possible outcomes with the respective conjunction — into our model yields more reasonable θ values, we extended the model as we will explain in the next section.

7.4 MODEL EXTENSION: WORLD-SAMPLING

7.4.1 *Model Definition*

In the behavioral data from the PE-task, we observed the tendency that even though participants selected a conjunction (e.g., 'both blocks fall') to describe a shown scene, they often gave rather low estimates for the probability of the corresponding outcome. In order to account for this observation, without the need to fall back on unreasonably low assertability threshold values, we expand the model by a second component that we refer to as the *world-sampling* part. The updated prediction of the model for a state s is referred to as $P^*_S(u \mid s)$, defined in Equation [45] where O denotes the set of possible and mutually exclusive outcomes, $O = \{bg, b\neg g, \neg bg, \neg b\neg g\}$ (both blocks fall, only the blue/green block falls, neither block falls).

$$P^*_S(u \mid s) = \gamma \cdot P_S(u \mid s) + (1-\gamma) \cdot \sum_{o \in O} P(o \mid s) \cdot P_S(u \mid s^{(o)}) \quad [45]$$

That is, $P(o \mid s)$ denotes the speaker's beliefs that o is the actual outcome (world). For example, if $s = \langle r, t = \langle w_{bg} = 0.1, w_b = 0.2, w_g = 0.3, w_\varnothing = 0.4\rangle\rangle$, $P(o = b\neg g \mid s) = 0.2$. $s^{(o)}$ refers to the model state s that assigns probability 1 to outcome o, for example, $s^{(o=bg)}$ corresponds to a state where probability table $t = \langle 1, 0, 0, 0\rangle$.[11] γ is a new free parameter of the model that regulates how influential the world-sampling part is for the model's prediction ($\gamma \in (0, 1)$). Parameter combinations where $\gamma = 1$ correspond to the non-extended version of our model without the world-sampling part.

7.4.2 *Model Fitting*

To fit the extended pragmatic speaker model, $P^*_S(u \mid C_i)$, to the observed production data, we use the same prior distributions for α and θ as before (see Equation [43] and [44]). For the new additional free parameter, γ, which represents the extent to which the speaker describes her true beliefs instead of describing one of the four outcomes straight away, we do not have any theoretical knowledge beyond that it lies in the interval between 0 and 1. Therefore, we choose a broad beta distribution as prior distribution for γ, shown in Equation [46].

$$\gamma \sim \text{Beta}(2, 2) \quad [46]$$

7.4.3 *Results & Discussion*

Figure 37 shows the posterior densities of the free model parameters, α, θ and γ for the extended pragmatic speaker model. The new parameter, γ, is estimated to be relatively low (mean estimate 0.298, 95% HDI [0.266, 0.329]) — much smaller than 1, which is the value that corresponds to the non-extended version of our model. The parameter for the assertability threshold, θ, is now estimated to be quite high (mean estimate 0.898, 95% HDI [0.893, 0.899] ∪ [0.901, 0.903]).[12]

---

11 More precisely, since we use beta distributions — with support values $\in (0, 1)$ — to generate the RSA-model states, we define $s^{(o)}$ as the four model states for which the respective probability is closest to 1. Note that here only the probability table, not relation r, is important since model predictions do not (at least not directly) depend on r; $P_S(u \mid s = \langle r, t\rangle) = P_S(u \mid s = \langle r', t\rangle)$ for any r, r' (in particular when $r \neq r'$).

12 For θ the posterior samples look multi-modal again. Like for the basic pragmatic speaker model, $\hat{R}$ is not suspicious, $\hat{R} \approx 1.002$. The pairs plot is shown in Figure 52 in Appendix A.

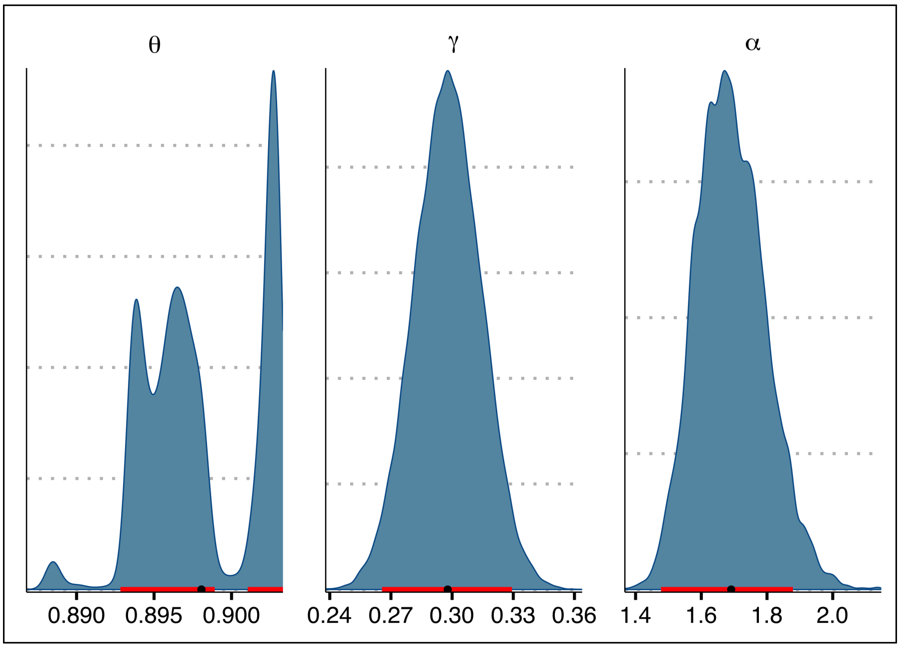


Figure 37: Approximated Posterior $P(\alpha, \theta, \gamma \mid P_S^*, D^{UC}, D^{PE})$ with 95% highest density intervals, 5000 MCMC-samples (lag=10) after a burn-in period of 10,000 samples.

Concerning model performance, the extended model explains the observed data equally well or better than the non-extended pragmatic speaker model, as shown by the mostly increased log likelihood values displayed in Figure 36.

We particularly expected that larger θ values would decrease the predicted probability for conditionals in independent contexts. Indeed, the extended model — for which θ is estimated to be reasonably high — predicts conditionals to be much less likely in independent contexts than predicted by the non-extended pragmatic speaker model, shown in Figure 38. Due to the increased probabilities for conjunctions that naturally comes along with the extension of the model, the predicted probability for conditionals in dependent contexts is, however, also decreased, although much less.

7.5 GENERAL DISCUSSION

We have proposed an RSA speaker-model for the use of conditionals in which the speaker aims to communicate her probabilistic beliefs about two events. To put the model to the test, we conducted a behavioral online experiment in which participants saw scenes of block arrangements. First they were asked in the PE-task to indicate their beliefs about whether or not a blue and a green block would respectively fall by adjusting four sliders, each representing one of the four mutually exclusive possibilities ($bg/\neg b\neg g/\neg bg/b\neg g$). In the second

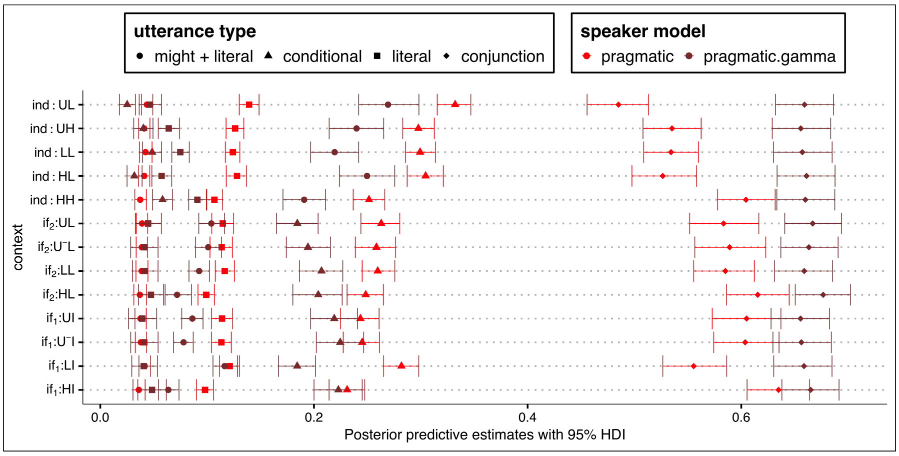


Figure 38: Mean estimates of the posterior predictive distributions for the basic (pragmatic) and extended (pragmatic.gamma) RSA speaker models with 95% HDIs for the use of the four types of utterances (coded by shape) in each context.

task, the UC-task, they were then asked to describe the same scene by creating utterances from a set of given words.

Based on the slider-ratings from the PE-task, we inferred which of the world states in our RSA-model best reflect participants' probabilistic beliefs for each shown scene (context). The resulting probability distributions, $P(s \mid C_i)$, determine how much the model's prediction for a particular model state s contributes to the overall model prediction for context C_i.

We then fitted the free parameters of the model — the assertability threshold θ and the rationality parameter α — to the empirical production data from the UC-task, once for the pragmatic speaker model and once for the the literal speaker baseline model. Due to the observed slider ratings from the PE-task, which often revealed low estimates for the events that participants later described in the UC-task, θ was estimated to be conceptually unreasonably low for both models. Therefore, we introduced an extended version of the model, assuming that to some extent, modeled by an additional free parameter (γ) participants decide for one of the four mutually exclusive outcomes — although they may not be certain about it — and describe the shown scene with the corresponding conjunction straight away. When fitting the parameters of the extended model (α, θ and γ) the assertability threshold θ is estimated to be much larger and thus conceptually reasonable. This further influences the model's predictions positively in that conditionals are now predicted to be relatively unlikely in independent contexts and particularly less likely in independent than in dependent contexts, like we had observed in the empirical data. Similarly, the extended model increases its predictions for the use of utterances with 'might' (e.g., 'the blue/green block might fall'); see Figure 38. With larger θ, there are simply more states for which only utterances with 'might' are assertable, in particular in the independent contexts.

Utterances with 'might' also showcase the influence that participants' exact slider ratings have on the model's predictions. Since this is the least informative type of utterance, the pragmatic speaker would only select them if none of the other alternative utterances was assertable in the respective state to be described. Consequently, it is essential which states we consider to represent our experimental conditions best, which is, in turn, determined based on participants' slider ratings. Purportedly small differences in participants' probability estimates thus influence the likelihood functions, that is, the weights $P(s \mid C_i)$, which, in turn, may lead to different sets of possibly assertable utterances for the model in context C_i. Therefore, it would be desirable to have some more certainty about the reliability of participants' beliefs provided in the PE-task, for example by showing the same contexts several times to each participant. This would, however, also require to reduce the complexity of the experiment so that it does not become too long, thereby introducing a different type of complexity that should be avoided.

Another possibility might be to simplify the stimuli on the one hand and, on the other hand, reduce the possible beliefs that participants can communicate with the four sliders or possibly without using sliders at all. We might, for instance, let participants decide for example between 'certain that X will happen' (e.g., X = both blocks fall), 'undecided between X and Y' (e.g., X = 'both might fall', Y = 'neither will fall'), 'quite uncertain, but exclude that X' (e.g., 'not the case that neither will fall'), 'completely undecided' (would not be surprised by any of the four possible outcomes) which translate to much simpler probability distributions than those that we observed (e.g., $\langle 1, 0, 0, 0 \rangle, \langle 0.5, 0, 0, 0.5 \rangle, \langle 1/3, 1/3, 1/3, 0 \rangle, \dots \rangle$). One might argue that in our experiment in which it was possible to give more precise estimates than this, participants did make use of this option. However, this might be an artifact induced from the experimental situation in the sense that participants may be urged to be as precise as possible in their responses to do the experiment properly. It seems rather unlikely that, if we had asked participants to describe their beliefs in words instead of using continuous sliders, they would have given responses of the same precision. Utterances like those suggested above do seem more plausible as possible responses.

Another aspect that we have left aside so far and that may be worth looking at, are utterance cost. The analysis of our experimental data revealed that participants' choices were likely influenced by utterance cost, defined in terms of the number of clicks that were required to create an utterance. As to be expected, we observed more short utterances than long utterances. Utterance cost are commonly used in RSA-models and could be integrated relatively easily into our model.

In that case, it would, however, be advisable to distinguish between the exact utterances that participants created. Remember that in the current analysis we collapsed all different realizations for the same meaning into a set of 20 standardized utterances. Therefore, we do not make a difference, for example, between the utterances 'both blocks fall' and 'the green block falls and the blue block falls'.

We have taken a population-level approach here, modeling the overall utterance selection frequencies across all participants, based on an overall probabilistic representation of the shown scenes that we inferred from participants' slider ratings. That is, we assume the same rationality parameter α as well as the same value for γ for all participants. To account for individual-level beliefs, it seems, however, reasonable to consider a hierarchical model, allowing for different values of α and γ across participants.

Part III

CONDITIONAL PERFECTION

This chapter concerns a phenomenon observed within the communication with conditionals, called *conditional perfection*, referring to the interpretation of an indicative conditional 'if p, q' as biconditional 'if and only if p, q'. The content of this chapter was published in slightly different form (Grusdt, Liu, et al., 2022) in the Proceedings of the conference *Experiments in Linguistic Meaning 2*, which was held at the University of Pennsylvania from May 18-20, 2022.

8 TESTING THE INFLUENCE OF QUDS ON CONDITIONAL PERFECTION

In natural language conversations, speakers often communicate 'if and only if' when they say 'if'. The reasons why in some circumstances, yet not all, conditionals receive a biconditional interpretation remain under investigation. In Section 2.1.2.2 I shortly introduced the phenomenon *conditional perfection* (CP) in the context of pragmatic explanations of CP proposed in the literature. Here, we will test an account from von Fintel (2001) who predicts the interpretation of a conditional ("if p, then q") as simple conditional or as biconditional ("if and only if p, then q") to depend on the focus of the conversation which may either lie on the conditions that make the consequent, q, true or on the consequences following when the antecedent, p, is true.[1] We present two novel behavioral experiments with stimuli that are not text-based but take advantage of participants' intuitive understanding of physics. We find some supporting evidence for the tested account that is not conclusive but suggests that other aspects, like the nature of potential alternative causes for the consequent to become true (e.g., with or w/o the influence of an external variable), also play a role for the interpretation of the conditional.

Before I describe our experimental setup and results, I will introduce von Fintel's (2001) proposed account in a bit more detail in the next section and consider other experimental studies on CP, which are in one way or another relevant for our own study.

8.1 BACKGROUND

While Geis and Zwicky, who (re)initiated the debate about CP among linguists,[2] argue that conditionals are commonly attributed a CP-reading, this regularity has been questioned by others providing various counterexamples (e.g., de Cornulier, 1983; Lilje, 1972). von Fintel (2001) goes one step further by making a proposal of when exactly a conditional is interpreted as biconditional and when it is not, which had not been precisely formulated in previous accounts. Similar to de Cornulier (1983), von Fintel refers to exhaustivity: he argues that a biconditional interpretation arises when the antecedent of a conditional is interpreted as exhaustive list of *conditions for the consequent*

1 In this chapter, I stick to the notation we used in our publication (Grusdt, Liu, et al., 2022), that is, I will use 'p' and 'q' to refer to the antecedent and the consequent instead of using A and C.

2 As noted by van der Auwera (1997), CP had already been discussed before Geis and Zwicky (1971), e.g. by Ducrot (1969).

whereas it is not triggered when the conditional is interpreted as exhaustive list of *consequences of the antecedent*. In other words, when the speaker is required to provide an exhaustive list of conditions for the consequent (q) and only mentions a single condition (p) the listener will infer that p is a sufficient and necessary condition for q, corresponding to a CP-reading. On the other hand, mentioning p as single condition in the antecedent does not trigger a CP-reading when the speaker is asked to provide an (exhaustive) list of consequences of p. According to von Fintel (2001), a (possibly implicit) question under discussion (QUD) determines whether the conditional targets the conditions of the consequent (e.g., under which conditions q?) or the consequences of the antecedent (e.g., what follows from p?). To illustrate the hypothesized effect of the QUD, consider the conditional in 44, inspired by an example from Lilje (1972):

(44) If a ball bounces off the table, it is a foul.

In a situation where a person A explains the rules of the game pool to a person B who has no experience with this game and where B asks A ...

(i) what happens if a ball bounces off the table (QUD: *if-p*)

(ii) which actions count as foul / whether there are actions that count as foul (QUD: *when-q*)

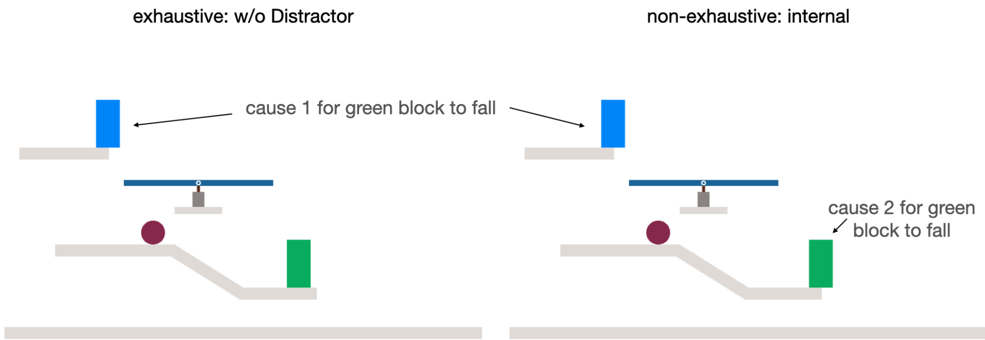
the same answer — the conditional in 44 — seems to be interpreted as biconditional only when the conversation is guided by the question in (ii). Given the context provided by the QUD *if-p* (i), the speaker is not expected to mention all possible actions that are considered a foul and, thus, CP does not arise in this case.

We are not the first to test the QUD-effect on the occurrence of CP as proposed by von Fintel (2001), his account has been tested in previous studies, though yielding conflicting results (Cariani & Rips, 2016, 2023; Farr, 2011). Farr's (2011) results provide quite strong evidence for the hypothesized effect of the QUD, whereas only a minute effect, if any at all, was found by Cariani and Rips (2016). In the experiment from Farr, participants read short vignettes and were asked whether a given conditional (e.g., "If you sell an eel, you get 2.50 euros") is a sufficient answer to a question, encoding the QUD (e.g., "What happens if I sell an eel?" vs." When do I get 2.50 euros?"). The framing of the question about the sufficiency of the conditional answer may be considered problematic (e.g., see Cariani & Rips, 2016; López Astorga, 2014); since the vignettes describe two possibilities to achieve the consequent (e.g., both, an eel and a pike cost 2.50), a no-answer to the question "Did Sahra answer Kerstin's question sufficiently?", with Kerstin's question being "When do I get 2.50 euros?" (QUD: *will-q*) and Sahra's answer "If you sell an eel, you get 2.50 euros", does not necessarily imply — even though it may strongly suggest

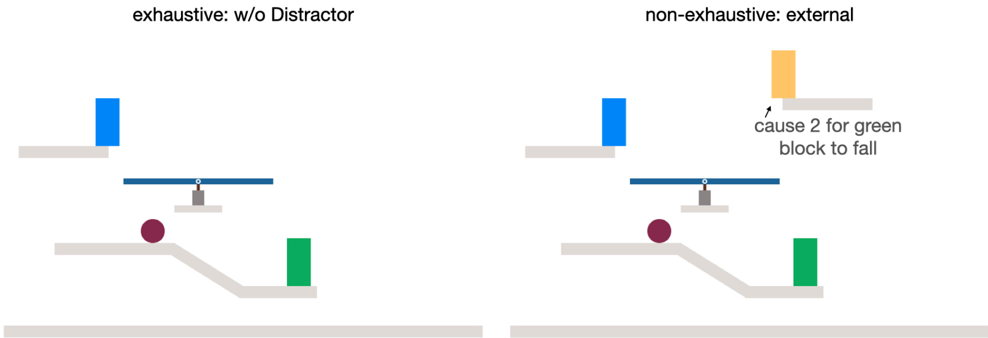
— that participants interpreted the conditional as biconditional. Cariani and Rips (2016) avoid this problem by measuring participants' endorsement rates of the four classical conditional reasoning inferences (MP, MT, AC, DA; see Section 2.2.1) to investigate the degree to which participants' interpreted a conditional as biconditional. However, the experimental stimuli from Cariani and Rips (2016) — again short vignettes — come along with world knowledge that is hard to control for. In one of their trials participants, for instance, learned that 'John has taken a test on Chapters 4-6 that has not been graded yet'. They were then asked whether the conditional (that they were told was true) 'If John understood Chapter 5, then John did well on the test' implies that 'John understood Chapter 5' when they also know that 'John did well on the test' (testing AC). As Cariani and Rips note themselves, participants might assume that the conditional simply does not state all conceivable conditions for the consequent; in this example, it is hard *not* to think of other reasons why John could do well on the test without having understood Chapter 5 (e.g. by cheating), which may have influenced participants' responses. We hope to benefit in this regard from our stimuli which are not text-based, making it easier to control participants' elicited beliefs about the situations at hand.

8.2 Introduction to our experiment

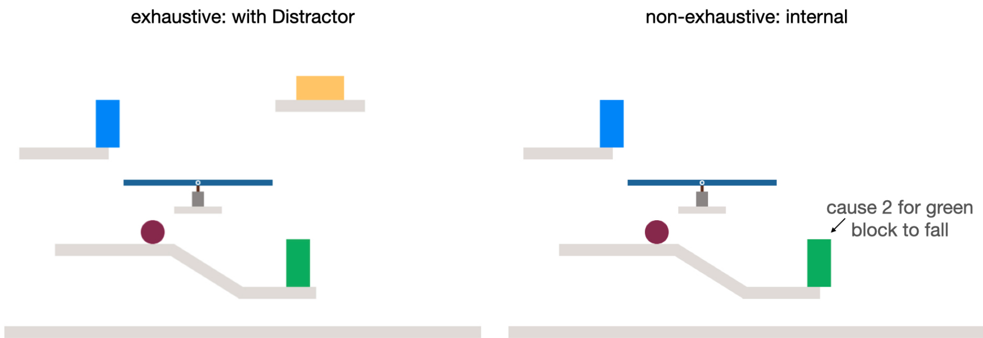
Here we present data from two novel behavioral experiments that we designed to investigate the influence of QUDs on participants' interpretation of conditionals as biconditionals. More precisely, we aim to test whether a QUD that puts the focus of the conversation on the antecedent of the conditional, by asking about the conditions for the consequent (QUD: *will-q*), positively influences a biconditional interpretation of the conditional, as compared to a QUD that puts the focus of the conversation on the consequent by asking about the consequences of the antecedent (QUD: *if-p*). In both experiments, participants are shown scenes of toy blocks together with a dialog between two characters that consists of a question, the QUD, and a conditional answer. Participants' task is to select the scene(s) that they believe to be best described by the conditional. The set of scenes among which participants have to choose contains at least two scenes, an *exhaustive*, and a *non-exhaustive* situation, as we call them. Both situations respectively contain (possibly among others) a blue and a green block, one in the upper left, the other in the lower right of the scene, where the falling of the upper block causes the lower block to fall as well; see Figure 39 for the critical situations from Experiment 2. Since the conditional answer is always the same — "If the upper left block falls, the lower block will fall" where 'upper left' and 'lower' are replaced by the respective color (green or blue) — we refer to the upper left block



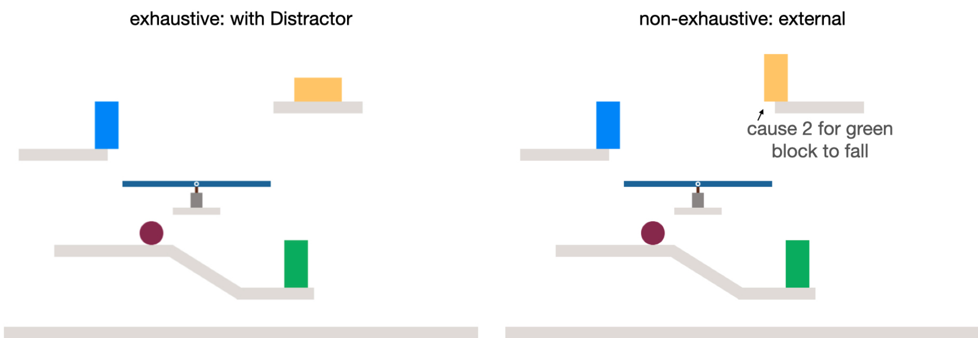
(a) Stimulus A



(b) Stimulus B



(c) Stimulus C



(d) Stimulus D

Figure 39: Exhaustive and non-exhaustive situations (with annotations) of critical trials in Experiment 2. The scenes in Experiment 1 were identical except for the yellow distractor block in stimuli C and D, which was also centered but standing upright. For simplicity, in all pictures shown here, the ant-block is blue and thus, the cons-block green; in the experiments, the color of the ant- and the cons-block was randomly chosen for each participant and trial.

as the 'ant-block', mentioned in the antecedent, and to the lower block as the 'cons-block', mentioned in the consequent. The two situations are manipulated with respect to the number of conceivable causes for the cons-block to fall. While, in both situations, the cons-block will fall if the ant-block falls, only in the non-exhaustive situation there is a second possible reason for the cons-block to fall: Either because of its own position on the edge of the platform (condition *internal*, Figure 39 stimulus A/C) or because of the falling of an additional block (condition *external*, Figure 39 stimulus B/D). The idea is that when participants interpret the conditional "If the ant-block falls, the cons-block will fall" as biconditional, they should prefer to select the exhaustive situation as better described by the conditional. The difference between the two experiments concerns the concrete setup as we will explain below.

As mentioned above, an advantage of our stimuli is that they allow to control participants' elicited beliefs about the situations at hand much better than text-based stimuli do. By showing participants animations of how the blocks behave, we hope to reduce the influence of additional world knowledge further. Also, our measurement for how the conditional is interpreted does not involve a direct question about the sufficiency of the conditional as answer to the QUD; we make participants select the situation in which they believe the conditional to be more appropriate.

To anticipate our results, we find some evidence for an influence of the QUD in the predicted way, yet the QUD cannot fully explain the observed data. The results are nonetheless interesting as they suggest other aspects to play a role for the occurrence of CP, like the nature of the potential alternative causes (external vs. internal), leading to different sets of alternative utterances as well as different possible types of interpretations (causal vs. epistemic).

8.3 EXPERIMENT 1

We preregistered the experiment based on a pilot study, in which we collected and analyzed data from 25 subjects. The code and the preregistration report are available on OSF.[3]

PARTICIPANTS A total of 300 participants were recruited via the online platform Prolific, including the 25 participants from our pilot study.[4] All of them were self-reported native English speakers, at least

3 See online resources for the preregistration: https://osf.io/47w85 and the code repository: https://tinyurl.com/255yaztv.

4 Note that, since two prolific ids were erroneously recorded twice, we eventually recorded 302 instead of the initially planned 300 participants such that all 300 data sets are ensured to come from 300 distinct participants. From the two data sets that were associated with the same prolific id, the one with the later time stamp was excluded.

18 years old and had an approval rate on Prolific of at least 80%. The cleaned data (see below) comprises 282 participants (103 male, 175 female, 3 other, 1 not specified) with a mean age of 32.8 years (range 18 – 84). For their participation, each participant received £1.25.

SETUP & MATERIALS The experiment consisted of a training phase with 7 trials and a testing phase with 18 trials. 12 of the test trials were critical trials, the remaining 6 were control trials including 3 attention-checks. In the training phase, participants saw animations of block arrangements that were created with the rigid body physics engine 'matter.js'.[5] The pictures shown in the test phase are screenshots (800 × 500 pixel) of the corresponding animations right before they would start.

MANIPULATIONS. The manipulated variables comprise the QUD as encoded in Ann's question and the shown pair of situations (exhaustive / non-exhaustive). The QUD has the following three levels: *neutral*: "Which blocks do you think will fall?", *if-p*: "What happens if the ant-block falls?" and *will-q*: "Will the cons-block fall?". The exhaustive and the non-exhaustive situation have two levels each: the former either contains or does not contain a yellow distractor block in the upper right (*exhaustive*: with distractor, w/o distractor) and in the latter, the second cause why the cons-block might fall is either due to its own position (*non-exhaustive*: internal) or due to the falling of a third block (*non-exhaustive*: external), as shown in Figure 39.[6]

TRAINING PHASE The main purpose of the training phase was to familiarize participants with the physical properties of the blocks. To induce a maximal degree of uncertainty in the critical test trials about whether the ant-block will fall, the blocks in the training trials are all positioned such that it should be quite easy to judge whether they will fall, in particular after having seen a few examples.[7] Contrary to that, in the critical trials of the test phase the center of the ant-block lies exactly on the edge of the platform. The order of the training trials was randomized within participants, but always alternated between trials where some blocks fall and trials where nothing happens. In each training trial participants were first asked to select all blocks that they believed to fall, by clicking on buttons with the respective block icons (or saying 'none'). Only then, they were able to click on

5 https://brm.io/matter-js/
6 In the exhaustive situations, the distractor block never moves and has no influence on the falling of the other blocks. In the non-exhaustive situations, participants learn in the training phase that the cons-block falls if the antecedent- or the yellow block falls or if both fall.
7 80% of the participants responded correctly in their respectively last trial of the training phase, whereas only about 50% gave the correct answer in the first training trial.

'RUN' to start the animation to see which blocks actually fall and whether their selection was correct. We explicitly asked participants to look at all blocks shown in the scene to encourage them to consider the potential influence of each block on any other block.

TEST PHASE In the test phase participants read a dialog between two characters, consisting of a question from Ann and an answer to that question from Bob. After reading the dialog, participants were shown two pictures of block arrangements and were asked to select the one that they rated as more likely described by Bob.

8.3.1 *Results*

DATA EXCLUSION We excluded all data from participants who did not give the correct response in: (i) all three attention-check trials or in (ii) more than one of the control trials or in (iii) the example test trial in the end of the training phase. Additionally, we excluded the data from two participants whose comments in the end of the experiment indicated that they did not do the experiment properly. Overall, the data of 282 participants remained to be included in the analysis.

BEHAVIORAL DATA Figure 40 shows the proportion of participants who selected the exhaustive situation as the situation that is more likely described by Bob's conditional answer to Ann's question. Participants' choices seem to depend strongly on the stimulus (color coded); while similar results are observed for stimuli B and D on the one hand and stimuli A and C on the other hand, there is a striking difference between the results for stimuli B and D as compared to the results for A and C. The difference between stimuli A/C and B/D lies in the second cause for the cons-block to fall in the non-exhaustive situation: for stimulus B and D, it is the potential falling of an additional block whereas for stimulus A and C, it is the position of the cons-block itself that may cause it to fall. In the former two stimuli, participants show a strong preference for the exhaustive situation and in the latter two stimuli, participants seem to prefer the non-exhaustive situation (selection rates consistently below 0.5).

The difference in responses observed between QUDs is much less striking. By eyeballing the data, we observe the expected tendency for stimulus A and C towards higher selection rates of the exhaustive situation when QUD=*will-q* as compared to QUD=*if-p*; for stimulus B and D, the same tendency is observed, but much less pronounced.

STATISTICAL MODEL We run a Bayesian logistic regression model, using the R-package brms (Bürkner, 2017), with the QUD, the stimulus (pair of two situations) and their interaction as predictors. As random effects, we include varying intercepts and slopes per partic-

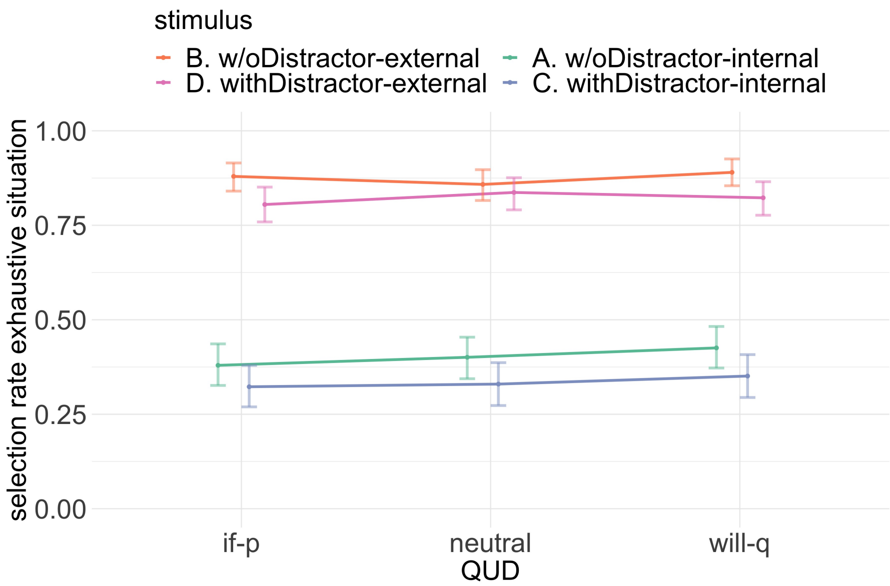


Figure 40: Relative frequency of the selection of the exhaustive situation in Experiment 1 as better described by the conditional "If the ant-block falls, the cons-block will fall", for each QUD, color coded by stimulus. Errorbars are 95% bootstrap confidence intervals.

ipant and use brms default priors for all parameters. Only for stimulus A, there is good reason to belief that the selection rate of the exhaustive situation will be larger when QUD=*will-q* as compared to QUD=*if-p*, reaching a posterior probability of approximately 96% ($P(\beta_{\text{qudif-p}} + \beta_{\text{stimulusC:qudif-p}} < \beta_{\text{qudwill-q}} + \beta_{\text{stimulusC:qudwill-q}}) = 0.959$, 95% CI: [-1.08, -0.03]). For the remaining three stimuli the posterior probabilities and 95% credible intervals for the respective comparison of parameters are 0.71 ([-1.01, 0.48], stimulus B), 0.833 ([-0.89, 0.23], stimulus C) and 0.712 ([-0.80, 0.38], stimulus D). The estimated posterior probability for the selection rate of the exhaustive picture to be larger when QUD = *will-q* compared to when QUD = *if-p* across all four stimuli amounts to 0.953.

8.3.2 *Discussion*

We found supporting evidence for the postulated effect of the QUD only for stimulus A, where the posterior probability for the exhaustive situation being selected more often when QUD=*will-q* as compared to QUD=*if-p* is reasonable large. However, the results for stimulus B,C and D also show a tendency towards this effect.

These results may be related to the unexpectedly strong difference in participants' responses between the four stimuli. Especially salient is the systematic preference for the exhaustive situation in stimuli B and D (non-exhaustive: *external*) and for the non-exhaustive situation in stimuli A and C (non-exhaustive: *internal*). Put differently, the biconditional interpretation is overall not very prominent for stimulus A and C, but it is for stimulus B and D. One aspect that might have influenced this difference is the set of salient alternative utterances: for stimulus B and D (external cause), there is a salient alternative to describe the *dispreferred* non-exhaustive situation, namely "If the *yellow or* the ant-block falls, the cons-block will fall", which may help explain the selection rates of the exhaustive situation in stimulus B and D close to ceiling. For stimulus A and C, for which the selection rate of the exhaustive situation is much lower throughout all QUDs, there is no similarly salient alternative for the non-exhaustive situation where the cons-block may fall without the influence of any other block.

Further, for stimulus B and C, the observed preferences (exhaustive for B, non-exhaustive for C) may be strengthened by the presence of the yellow distractor block in only one of the two shown situations; participants may have favored the situation *without* the yellow block — corresponding to the respectively preferred situations — just because it is not mentioned in the conditional. This is not per se problematic to test for an effect of the QUD, at least as long as the potential effect is not superposed completely by selection rates close to ceiling which we do observe for stimulus B.

Another possibility that may have influenced the results, in particular the preference for the non-exhaustive situation in stimuli A and C, is an epistemic instead of a purely causal interpretation of the conditional in these cases.[8] Assuming an epistemic interpretation, the conditional is particularly true in the non-exhaustive situation — which is overall preferred in stimuli A and C.

In order to circumvent the possibility that a putative effect is not found due to participants' strong tendency to prefer either the exhaustive or the non-exhaustive situation depending on the concrete scenes, we conducted a follow-up experiment where we do not force participants to choose one among two situations but allow them to select both. When QUD=*if-p*, the conditional $p \rightarrow q$, should in fact be accepted as description for both situations since other possible reasons for the consequent are simply expected to be irrelevant and thus, the conditional is an appropriate answer in the exhaustive as well as in the non-exhaustive situation. Indeed, several participants in Experiment 1 mentioned in their comments in the end of the study that for *some* trials, both pictures were possible.

8 Epistemic interpretation in the sense that if the "nature of" the ant-block is such that it falls, the cons-block should fall as well since it is of the same "nature" as the ant-block: both blocks are identically centered on the edge of their platforms.

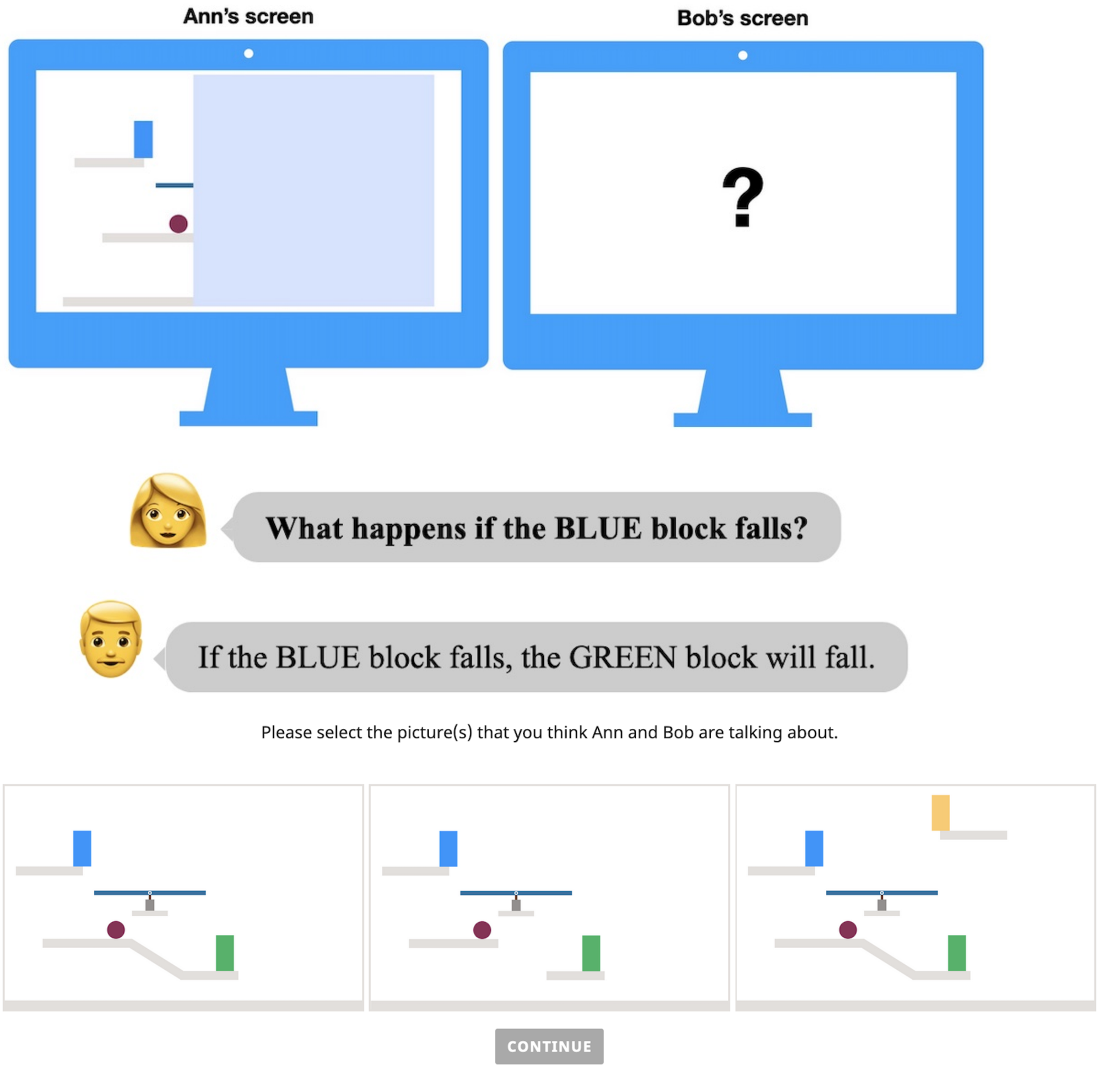


Figure 41: Critical test trial from Experiment 2 where QUD=*if-p*, non-exhaustive (right) *external* and exhaustive (left) *w/o distractor*. The picture in the middle is the control scene.

8.4 EXPERIMENT 2

The code and the preregistration report for Experiment 2 are publicly available on OSF.[9]

PARTICIPANTS 315 participants were recruited via the online Platform Prolific, using the same eligibility criteria as for Experiment 1.[10] The cleaned data (see below) comprises data from 181 participants

9 For the preregistration see https://tinyurl.com/pphck52h, for the code repository see https://osf.io/sjdax/.

10 We increased the number of recorded participants by 100 after having cleaned the data of the originally planned 215 participants since we had to exclude many more participants than expected who did not fulfill our predefined criteria. All data was analyzed only after all 315 participants were recorded.

(76 male, 103 female, 2 other) with a mean age of 37.12 years (range 18 – 68). For their participation, each participant received £1.67; on average they finished the experiment in approximately 14 minutes (range 4.5 – 46).

SETUP & MATERIALS The training phase consisted of 8 trials and was followed by the test phase consisting of 17 trials split into 3 blocks, a practice block with 4 trials followed by two test blocks with 7 and 6 trials respectively. The trials of the two test blocks alternated between filler and critical trials and included an attention check trial after the first test block. The order of trials within filler, critical and practice trials was randomized for each participant. After each block, participants had the possibility to take a break before proceeding with the next block. In the end, we further asked participants to answer a set of questions about the experiment to (i) verify that they did not ignore Ann's question and (ii) to get an idea of how certain participants had to be such that they would select only one scene.

PROCEDURE The most important difference in the procedure of Experiment 2 compared to Experiment 1 is that in Experiment 2, participants were not forced to select a single picture. They read the same dialog as in Experiment 1 but were shown three instead of two scenes to choose from. The third picture is referred to as the *control scene* since it shows a situation that contradicts Bob's conditional answer and should, thus, never be selected.

By telling participants that Ann sees part of the scene that Bob describes, Ann's question was meant to be more purposeful: she seeks to get more information about a scene that she only has partial access to, see Figure 41 for an example trial. While the partial scene was immediately visible in each trial, Ann's question, Bob's response and the three scenes had to be revealed one by one. With this setup we hoped to enforce participants to process both, the QUD and the conditional, before making a selection.

Further, Experiment 2 was built up as a game where participants can earn points, with the aim to incentivize that participants do not always select a single situation (or always both): when selecting two scenes, they get 50 points if the correct one is among them, otherwise, they loose 100 points. Selecting only the correct scene is awarded with 100 points, but when a single scene is selected that is *not* the correct one, participants lose 100 points. Therefore, in the long run, participants are better off to select two scenes when they are undecided.

MANIPULATIONS As in Experiment 1, we manipulate the QUD encoded in Ann's question, but without using the *neutral* condition ("Which blocks do you think will fall?") in the critical test trials. While the four critical stimuli (pairs of an exhaustive and a non-exhaustive situation) are the same as in Experiment 1, there are only 6 critical trials in Experiment 2: since Ann's question, the QUD, relates to the block that is shown in the partial scene (when QUD=*if-p*, the partial scene shows the ant-block, when QUD=*will-q*, it shows the cons-block), it should be at the same position as the respective block in the three pictures of scenes among which participants make their selection; only then none of the three situations can be excluded just because it does not match the part of the scene that Ann sees. Thus, stimuli A / C are not combined with QUD *will-q*.

TRAINING PHASE The animations in the training phase were the same as in Experiment 1 plus one additional trial which showed the situation that contradicts the conditional "If the ant-block falls, the cons-block will fall" (see Figure 41, middle), which is the control scene in the critical conditions of the test phase.

TEST PHASE The test phase consists of three blocks, a practice and two test blocks. The practice block consists of 4 trials in which participants got feedback about the correct picture and the number of points they received with their selection. The main purpose of the practice trials was to demonstrate that Bob's responses are informative and to make participants learn how their choices impact the amount of points they get. The procedure in the the trials of the two test blocks was the same as in the practice block, except that participants did not receive feedback anymore. In order to keep the character of the game up without influencing participants' choices, they were told that they would get their final score in the end of the experiment. The filler trials in the test blocks were designed such that there is a similar number of trials where QUD=*if-p* (6) and QUD=*will-q* (5).

8.4.1 *Results*

DATA EXCLUSION We excluded all data from participants who fulfilled at least one of the following criteria: (i) they did not select the correct scene in the attention-check trial, (ii) they selected the control scene at least once in the test phase (excluding the practice block), (iii) they affirmed either that they only read Bob's answer, but not Ann's question, or that Ann's question was always the same or (iv) they responded within less than 6 seconds in at least 2 of the critical trials. In total, we had to exclude the data from 134 participants so that in total the data from 181 participants was included in the analysis.[11]

11 Number of participants excluded for each combination of criteria they are excluded for. In total, 111 participants were excluded because of 1 criteria they didn't fulfill; 83 selected control scene, 24 ignored or didn't realize different QUDs, 4 failed in the attention-check trial. Among the 83 participants whose data was excluded because they selected the control scene, 32 selected it in several trials, 51 just once.

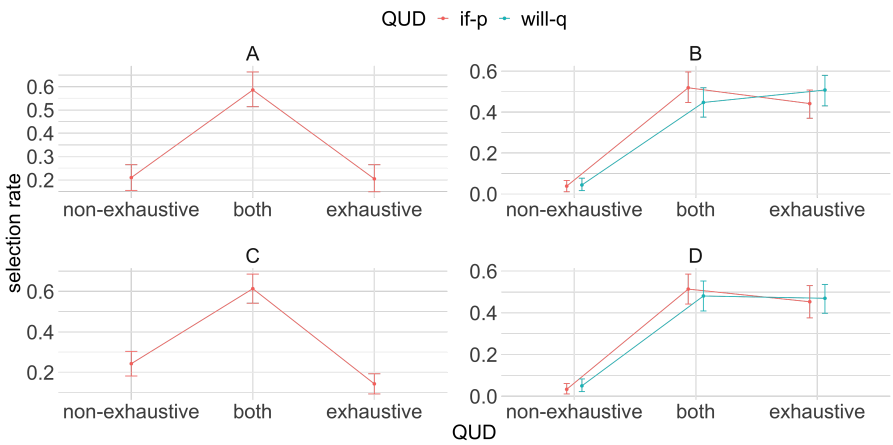


Figure 42: Relative frequency of the scene(s) participants selected in Experiment 2 based on the conditional "If the ant-block falls, the cons-block will fall"; color code represents the two QUDs, errorbars are 95% bootstrapped confidence intervals.

BEHAVIORAL DATA Figure 42 shows participants' average responses in Experiment 2 for each of the four critical stimuli. Similarly to Experiment 1, we observe a preference away from an exhaustive interpretation in stimuli A and C, where selecting both situations is much more likely than selecting only one situation; in stimulus C, selecting only the non-exhaustive situation is even more likely than selecting only the exhaustive situation. Contrary to that, stimuli B and D again show a preference towards an exhaustive interpretation: the selection rate of only the exhaustive situation is much higher in these stimuli than it is in stimuli A and C. Selecting both situations is, however, almost equally likely for stimuli B and D compared to A and C.

By eyeballing the results in Figure 42, the QUD *will-q* shows the predicted effect for stimulus B: we observe an increase in the selection of the exhaustive situation when QUD=*will-q* compared to QUD=*if-p* and a decrease in the selection of both situations. For stimulus D, selecting both situations is more likely than selecting only the exhaustive situation when QUD=*if-p*, but we do not observe the predicted effect of the QUD on the selection rate of the exhaustive situation,

	control	qud	attention	rt	comment	n	n_criteria
1	TRUE	FALSE	FALSE	FALSE	FALSE	83	1.00
2	FALSE	TRUE	FALSE	FALSE	FALSE	24	1.00
3	TRUE	TRUE	FALSE	FALSE	FALSE	9	2.00
4	TRUE	FALSE	TRUE	FALSE	FALSE	8	2.00
5	FALSE	FALSE	TRUE	FALSE	FALSE	4	1.00
6	TRUE	FALSE	FALSE	FALSE	TRUE	2	2.00
7	TRUE	FALSE	FALSE	TRUE	FALSE	2	2.00
8	TRUE	TRUE	FALSE	FALSE	TRUE	1	3.00
9	TRUE	TRUE	TRUE	FALSE	FALSE	1	3.00

which is in fact on average more likely when QUD = *will-q* but the increase seems to be marginal.

STATISTICAL MODEL We run an ordinal regression model with brms (Bürkner, 2017; Bürkner & Vuorre, 2019) with the QUD, the non-exhaustive and exhaustive situation and the interaction between the QUD and the exhaustive situation and between the exhaustive and the non-exhaustive situation as predictors, including by-participants random intercepts and slopes for each predictor except for interactions. We chose an ordinal regression model as the three response categories reflect the degree of how exhaustive the conditional is interpreted: the selection of only the non-exhaustive situation corresponds to a maximally non-exhaustive interpretation and the selection of only the exhaustive situation to a maximally exhaustive interpretation, selecting both situations corresponds to an interpretation in between both extremes.

For stimulus B, the posterior probability that participants interpret the conditional more exhaustively when QUD=*will-q* as compared to QUD=*if-p* amounts to 93%($P(\beta_{\text{qudwill-q}} + \beta_{\text{exhwoD}} + \beta_{\text{qudwill-q:exhwoD}} > \beta_{\text{exhwoD}}) = 0.93$, 95% CI: [-0.04, 0.58]), pointing towards the postulated effect of the QUD. As Figure 42 suggested, for stimulus D, the posterior probability is much lower ($P(\beta_{\text{qudwill-q}} > 0) = 0.56$, 95% CI: [-0.26, 0.32]) and does not provide evidence for our hypothesis.

We further speculated (a) that when QUD=*will-q*, the selection rate of only the exhaustive situation will be larger than the selection rate of both situations and (b) when QUD=*if-p*, the selection rate of both situations will be larger than of only the exhaustive situation. Our data only provides evidence for (b) in stimuli A and C, where the posterior probability for $P(\text{both} \mid \text{QUD}=\textit{if-p}) > P(\text{exhaustive} \mid \text{QUD}=\textit{if-p})$ is 1.

8.4.2 *Discussion*

Only the data for stimulus B provides a reason to believe in a more exhaustive interpretation of the conditional when QUD=*will-q* as compared to QUD=*if-p*, for stimulus D, the QUD does not seem to have the same alleged effect.

When we look at the exact conditions in which participants responses differ between stimuli B and D, we find the strongest difference in the selection rate of the exhaustive situation for stimuli B and D when QUD=*will-q*, with a posterior probability for $P(E \mid \text{QUD} = \textit{will-q}, \text{stimulus} = B) > P(E \mid \text{QUD} = \textit{will-q}, \text{stimulus} = D)$ of 0.92. When QUD=*if-p*, the posterior probability that $P(E \mid \text{QUD} = \textit{if-p}, \text{stimulus} = B) > P(E \mid \text{QUD} = \textit{if-p}, \text{stimulus} = D)$ is 0.52. Assuming that potential alternative causes are irrelevant for the choice that participants make when QUD = *if-p*, it seems reasonable that the observed difference between participants' responses in stimuli B and D can mainly be ascribed to the condition where QUD=*will-q* since when QUD=*if-p*, the focus of the conversation lies on the consequences of the antecedent and so, participants are not expected to consider other potential causes for the consequent. Further, it may have been the case that the presence of the distractor block in the exhaustive situation of stimulus D (that is absent in B) made participants more hesitant to decide for a single scene and thus they tend to choose both situations more often in this stimulus when alternative causes are considered, i.e., when QUD=*will-q*. Participants were indeed encouraged to select a single situation only when they were very confident, which 86% of participants confirmed in the questions in the end of the experiment.

Concerning the results for stimuli A and C, we observe a tendency towards a non-exhaustive interpretation of the conditional (selection of the non-exhaustive situation or both), similarly to what we saw in Experiment 1. In fact, for stimulus C, the posterior probability of the probability to select only the non-exhaustive situation to be larger than the probability to select only the exhaustive situation amounts to almost 0.98 (P(non-exh. | QUD=*if-p*, stimulus = C) > P(exh. | QUD=*if-p*, stimulus = C)). The fact that in Experiment 2 we find that if participants choose a single scene in stimulus C, they select the non-exhaustive situation significantly more often than the exhaustive situation — which is not the case for stimulus A — might help to explain why, in Experiment 1, we did not find the predicted QUD-effect for stimulus C which we found for stimulus A; a general tendency towards the non-exhaustive situation may have interfered with a putative QUD-effect which may thereby become harder to find, especially when we assume that this preference would only become stronger when the alternative causes are assumed to be particularly considered, that is, when QUD=*will-q*.

8.5 CONCLUSION

Overall, we find some evidence supporting the hypothesis that a QUD that focuses on the conditions bringing about the consequent yields a more exhaustive interpretation of an indicative conditional than does a QUD that focuses on the consequences of the antecedent. Our results are far from being conclusive, yet they show that the interpretation of conditionals as biconditionals is likely to be the result of an interplay of various factors.

Especially Experiment 1 showed that when forced to choose either an exhaustive or a non-exhaustive situation, participants showed substantially different preferences depending on the nature of the second conceivable cause for the cons-block to fall in the non-exhaustive situation. It either fell because of a third block (stimuli B/D) or because of its own position on the edge of a platform (stimuli A/C). Stimuli B/D yield a strong preference of the exhaustive-situation whereas in stimuli A/C we observed a (less strong) preference of the non-exhaustive situation — across all QUDs including a neutral question. The second cause for the cons-block to fall as it is is realized in stimuli A and C, namely because of its own position on the edge, comes along with another possible interpretation of the conditional that does not apply to stimuli B and D: the conditional may receive an epistemic instead of a causal interpretation. Consider the following conditional as an example for a conditional describing a situation similar to those in stimuli A and C, receiving an epistemic interpretation: "If that guy solved the puzzle, she will solve it [too/all the more]". It does not seem to suggest that 'if that guy does not solve the puzzle, she will not solve it', rather, it suggests that 'she might solve it, while he might not, but if he does, she will as well'. And this seems to be the case even if the conditional is an answer to the question "Will she solve the puzzle?". In other words, under an epistemic interpretation of the conditional we should not expect to see a difference between the QUDs whereas we do expect a difference when the conditional receives a non-epistemic, causal interpretation. Therefore, disentangling a causal versus an epistemic interpretation of the conditional may help to get a cleaner picture of what is going on here.

Further, the observed tight connection between conditionals and causality generally suggests that it may be worth to look at the production of conditionals in comparison to the use of causal language (e.g., "X may make Y fall" or "Y may fall because of X") to learn more about how participants use (and interpret) conditionals.

Part IV

GENERAL DISCUSSION & CONCLUSION

9 GENERAL DISCUSSION

9.1 SUMMARY OF MAIN CONTRIBUTIONS

In this thesis I have investigated how natural language conditionals ('if ..., then ...') are used by speakers and interpreted by listeners. While formal models have successfully been applied to many pragmatic phenomena, so far no such model has been proposed for conditionals. In this thesis, I set out to fill this gap by the development of a computational model that aims to explain the diverse interpretations observed within the communication with conditionals. It makes quantitative predictions about (i) the situations when a speaker likely chooses to utter a conditional as compared to a non-conditional utterance and (ii) the inferences that a listener draws with respect to the speaker's beliefs when interpreting the speaker's utterance of a conditional.

With the developed *Rational-Speech-Act* (RSA) model that I presented in Chapter 3, we demonstrate that given an appropriate representation of the relevant aspects of the world, in particular the causal structure of the modeled variables, we can explain a set of different common observations in the communication with conditionals on pragmatic grounds (Chapter 4). These include the dependency relation between the antecedent and the consequent, the infelicity of missing-link conditionals whose constituents bear no relation whatsoever, the listener's inference about the speaker's epistemic uncertainty with respect to antecedent and consequent and the tendency to interpret 'if' as 'if and only if'. The flexibility of our model further allows to represent more concrete utterance contexts that lead to specific context-dependent interpretations of the conditional. In Chapter 5, we showed, for example, that the model is able to reproduce intuitive interpretations for *Douven's puzzles*, three conditionals, assumed to be uttered in specific utterance contexts, that are all interpreted differently with respect to the listener's belief in the antecedent, which either remains unchanged, increases or decreases upon the listener's uptake of the conditional.

A strength of our model — and of RSA-models in general — are its quantitative predictions that allow for a direct comparison with empirical data. In Chapter 7, we compared the model's predictions with empirical data collected in the behavioral experiment reported in Chapter 6. Although not all aspects of the observed data were captured by our pragmatic speaker model (e.g., participants showed a strong preference for the conditional $A \rightarrow C$ compared to the conditional C → A which are predicted to be approximately equally likely by the model), the model was largely able to explain the data. Importantly, it performed better than our literal baseline speaker, modeled to select all assertable utterances with equal probability and, thus, without considering the influence of the speaker's utterance choice on the listener's interpretation.

Overall, the proposed model is a first step towards a concrete formalization of the pragmatics of conditionals. Through the integration of the utterance context in terms of rich prior beliefs and the reasoning processes about the mental states/intentions of the interlocutors, the model may have the potential to be a starting point for a unified account of how seemingly different interpretations of conditionals might be explained with a single model when the respectively relevant aspects of the world are appropriately represented.

To be sure, our model operates on the computational level of analysis according to Marr (1982). Models at this most abstract level in Marr's (1982) hierarchy follow a reverse-engineering strategy (see Zednik & Jäkel, 2016): they aim to reproduce the observed phenomenon in order to learn how a certain input may lead to the observed output. That is, they describe the problem under investigation without making claims, neither about the exact algorithm that the system (here the human brain) is using, and much less about its exact implementation (here on a neuronal level).

In light of recent developments of language models like ChatGPT, that go far beyond classical language models that predict the most likely next word given a sequence of words, a question that we have to ask ourselves is if and how our formal model makes a contribution that these powerful new language models do not cover.[1] While ChatGPT seems to be able to use and interpret conditionals pragmatically — in accordance with how human language users would interpret them intuitively — it does *not* give answers to the question how these interpretations arise. However, with our model we aim to make a step toward a better *understanding* of how meaning of natural language arises. So far models like ChatGPT reproduce — in a clearly impressive manner — the massive data that they have been trained on, but do not provide answers to questions like the one that we have been asking in this thesis about how the diverse interpretations and uses of conditionals in natural language can be *explained*. Therefore, computational models like ours that provide quantitative answers which can be compared to empirical data remain a useful tool to gain insights about the mechanisms behind human abilities like the interpretation and use of natural language.

Besides the theoretical considerations in form of the model presented in the first part of the thesis, I ran several behavioral experi-

1 https://chat.openai.com/

ments to investigate the pragmatics of conditionals further. The first experiment presented in Chapter 6, and mentioned above, aimed at testing the proposed model, while two more experiments presented in Chapter 8 concerned a specific phenomenon — the frequently observed interpretation of 'if' as 'if and only if' — called *conditional perfection* (CP). In the two experiments on CP, we tested an account from von Fintel (2001), who predicts an influence of the QUD ('what if p?' vs. 'when q?') on the interpretation of a conditional 'if p, q' as 'iff p, q'. Our results point at an influence of the QUD in specific circumstances, yet they were inconclusive and further research is needed to investigate this influence further.

In all of the three experiments presented in this thesis, we exploit peoples' intuitive understanding of physics: we showed participants scenes of block arrangements and asked them either to describe the shown scene by choosing an utterance among a set of different available utterances (Experiment 1, UC-task) or we showed them an utterance (e.g., a conditional) and asked them to choose among a set of different shown scenes (CP-Experiments). The advantage of this approach is that we can restrict participants' prior beliefs much more through the shown scenes, for example with respect to the number of possible causes for a certain effect, than it is possible with text-based stimuli.

In the reminder of this chapter I would like to discuss our pragmatic model for the use and interpretation of conditionals in light of some arguments recently considered in the literature, which explicitly challenge pragmatic solutions, in particular to the commonly inferred link between the antecedent and consequent.[2] I will address these in Section 9.2. In Section 9.3 I then consider our model with respect to recent experimental data on how listeners update their beliefs upon the uptake of a conditional. Lastly, in Section 9.4, I will point out limitations of the proposed model.

9.2 THE INFERRED LINK BETWEEN A AND C: A CONVERSATIONAL IMPLICATURE?

One aspect of conditionals has been particularly debated recently is the link between antecedent and consequent that is commonly inferred upon a listener's uptake of a conditional. We have shown in Chapter 4 that the combination of pragmatic reasoning with richly structured prior knowledge about the modeled variables is sufficient for the listener to infer a dependency between antecedent and consequent — assuming a simple literal semantics for conditionals based only on the relevant conditional probability. In this section, I would

2 Following Skovgaard-Olsen et al. (2017) I will also use the less verbose expression *reason relation* here to refer to the link between antecedent and consequent.

like to discuss our approach in light of arguments that have been put forward against a pragmatic account of the reason relation.

9.2.1 *The informativeness of utterances and the modulation of the selected utterances by the speaker's intentions*

Krzyżanowska et al. (2021) present data that they interpret as evidence against a Gricean explanation for the inferred link between antecedent and consequent. In a series of experiments, they asked their participants to rate the assertability of what they call TT-conditionals, referring to conditionals whose antecedent and consequent are both known to be true. On the other hand, their participants were asked to rate the assertability of the corresponding conjunction of the antecedent and the consequent. The variables that they manipulate are sentence type (conditional vs. conjunction), clause content (type vs. token, e.g., roses in general vs. a specific instance of a rose) and the inferential connection between antecedent and consequent (e.g., roses are plants and roses have thorns (unconnected) vs. roses are plants and roses need water (connected)).

The authors argue that if the inferred link was to be explained on Gricean grounds, more precisely by means of a Gricean conversational implicature, TT-conditionals should be rated — independently of the inferential connection — as *less* natural than the corresponding conjunctions. This is because the conjunction is more informative than the conditional, rendering the former the preferred choice of a Gricean pragmatic speaker. This is, however, not what Krzyżanowska et al. observed. Whether TT-conditionals were considered assertable differed between the inferential connection condition: when there was no connection between antecedent and consequent, only conjunctions received high assertability ratings. But when antecedent and consequent were connected, the conditional and the conjunction both received high assertability ratings which Krzyżanowska et al. ascribe to the existing inferential connection between antecedent and consequent. Although they do not exclude the possibility that the connection can be be modeled pragmatically and consistent with their data, they conclude that in light of their data and other empirical results "an explanation on merely pragmatic grounds becomes less and less plausible" (Krzyżanowska et al., 2021, p.22).

Their argument against a Gricean account of the reason relation is based on the assumption that the conjunction is the stronger utterance and should thus be rated as more assertable than the weaker utterance, the conditional. We have been arguing on the same grounds: if our pragmatic speaker is in the position to use a conjunction instead of a conditional, the former will be the model's preferred utterance choice since it is indeed more informative. However, contrary to Krzyżanowska et al. (2021), we use a concrete definition for the informativeness of an utterance in our model, assuming that the speaker's intention is to communicate her probabilistic beliefs. If that's the speaker's first and foremost goal, a conjunction 'A and C' is expected to be preferred over the conditional $A \rightarrow C$ because it diminishes the set of states that come into question as the speaker's target state more than the conditional would — in other words, the conjunction is more informative. The speaker may yet follow a different goal, for example, the communication of the relation between both propositions. In case that the speaker's intention changes, for example from communicating her beliefs about the probabilities of two events to communicating her beliefs about the relation between them, there will also be a change in the informativeness of utterances, which is directly linked to the speaker's intentions. Since in the experiment from Krzyżanowska et al., no such intentions were provided, participants might have silently come up with reasonable possibilities, possibly leading to similarly high assertable ratings for conjunctions and conditionals. This seems particularly reasonable in their Experiment 1, in which the utterance to be rated was given within a classroom context in which participants read a conjunction that they were told to be a student's summary of what they learned in class. Then, depending on the sentence type condition, participants either read the same conjunction again or they read the corresponding conditional, which they were asked to rate according to how natural it would be for the teacher to say. In a classroom context it is, however, easy to imagine that the speaker (the teacher) aims to emphasize an existing connection, which would rationalize the choice of the conditional, given that contrary to the conditional, the conjunction does not communicate a particular connection between the two conjuncts.

In order to reconcile their account with a Gricean view of communication, Krzyżanowska et al. (2021) propose to consider the inferred link to be due to a conventional implicature. This means that it would not result from reasoning about the speaker's production protocol, but would be inherently connected with the meaning of the word 'if', much like 'but' is associated to signal a contrast, while otherwise being indistinguishable from the word 'and' (Grice & White, 1961). The downside of this proposal is that instances of conditionals that are felicitous although antecedent and consequent lack a (causal) relation, are *per se* excluded; the authors remain silent in this regard and rather limit their theory to common 'non-special' indicative conditionals.

Krzyżanowska et al. argue that with the inferred link being conventionally implicated the link could be the additional information that renders the conditional equally informative as the conjunction, thereby explaining the observed similarly high assertability ratings for conditionals and conjunctions. This seems very plausible, but, as I have argued above, such high assertability ratings may also be achieved differently, namely by an appropriate specification of the speaker's intentions. In terms of our model, assuming the speaker's goal is to communicate an existing relation between antecedent and consequent, our speaker would be better off to say $A \rightarrow C$ than 'A and C' given that both propositions are (very likely) true: with this intention, the speaker would preferably choose utterances that increase the probability for a literal listener to infer the target relation (e.g., $A \overset{++}{\rightsquigarrow} C$). And since among all model states, the majority of those for which $A \rightarrow C$ is assertable (i.e., $P^{(s)}(c \mid a) \geqslant \theta$) are states with a dependent relation, the conditional $A \rightarrow C$ becomes more informative than the conjunction 'A and C' which is assertable for an approximately equal number of independent and dependent states. Contrary to that, when the speaker's goal is to communicate her probabilistic beliefs, instead of the relation, our modeled speaker strongly prefers the conjunction 'A and C' over TT-conditionals (states where $P^{(s)}(a, c) \geqslant \theta$) *independently* of the relation between A and C.

Taken together, it does not seem necessary to accept the disadvantages that come along with the integration of the reason relation into the core meaning of the conditional, an appropriate representation of the speaker's intentions will do.

Another example from Krzyżanowska et al. (2014) that can similarly be explained with the speaker's intentions is shown in (45).[3]

(45) If the UK is ruled by a king, it is a monarchy.

(45) is a reasonable utterance for a speaker who does not know whether the UK has a king or a queen or a different form of government altogether. Under the assumption that the speaker wants to communicate her beliefs about the probabilities of the two events and assuming it common knowledge that the UK is ruled by King Charles, (45) is rather surprising as it conveys uncertainty about the antecedent (the UK is ruled by a king) on part of the speaker. In other circumstances, for example when (45) is a student's answer to the teacher's question "What form of government does the UK have?", the same conditional seems to be an appropriate utterance assuming that the speaker's intention is, for example, to provide as much relevant information as possible. With the conditional, the student signals not to know who is ruling the UK, but also signals to know the relation between a king/queen and a monarchy, which presumably is relevant information in a school context.

3 To account for (45) (among other examples) Krzyżanowska et al. (2014) propose a relatively complex condition with several subconditions: *A speaker S's utterance of "If p, q" is true iff (i) q is a consequence — be it deductive, abductive, inductive, or mixed — of p in conjunction with S's background knowledge, (ii) q is not a consequence — whether deductive, abductive, inductive, or mixed — of S's background knowledge alone but not of p on its own, and (iii) p is deductively consistent with S's background knowledge or q is a consequence (in the broad sense) of p alone.* (Krzyżanowska et al., 2014, Definition 1, p.776)

9.2.2 *Cancellability and reinforceablity*

Contrary to Gricean conventional implicatures (e.g., the contrast communicated with 'but'), conversational implicatures are characterized by being context-dependent. This comes along with the possibility of the speaker to cancel or, on the contrary, emphasize, what was *conversationally* implicated with the selected utterance. For conventional implicatures this is, however, not possible, making these two features (reinforceability and cancellability) good test cases to decide whether a certain inference may rather be explained by a conversational or a conventional implicature (e.g., see Birner, 2013; Sadock, 2006).

Krzyżanowska (2019) thus argue that if the relation, that listeners commonly infer to exist between antecedent and consequent when interpreting a conditional, can be traced back to a conversational implicature, it should be cancellable as well as reinforceable. The authors discuss several examples that they take to show that neither is true, which would disqualify a conversational implicature as explanation for the inferred relation.

For an example of the reinforceability and cancellability of the common implicature from the utterance 'some' to the interpretation 'some but not all' consider (47), and respectively (46), taken from Krzyżanowska (2019). Note that there are several ways how, and to what extent, the implicature may be cancelled. It may be cancelled entirely, as in (46c), or only in part by leaving open the possibility that the more specific case holds, which is realized in (46a) and (46b) alike. While (46a) is expressed by double negation (I *did not say* that it is *not* the case that ...) (46b) is expressed by directly stating that the more specific case remains a possibility.

(46) Some of my students passed the exam.

a. Oh, I didn't mean to imply that some of them didn't, I just haven't checked all the exams yet. [cancel 'not all']

b. In fact, it is possible that all students passed, I just haven't checked all exams yet. [cancel 'not all']

c. In fact all of them did. [cancel 'not all']

(47) Some of my students passed the exam ... but not all of them did. [reinforce 'not all']

CANCELLABILITY. The examples that Krzyżanowska (2019) gives to test for the cancellability of the inferred link between antecedent and consequent are given in (48)-(49).

(48) # If dolphins have fins, then they can't breathe under water. Oh, I didn't mean to imply that fins have anything to do with the ability to breathe under water. (Krzyżanowska, 2019, Ex. (5))

(49) # If Bobby is fond of drawing, he will be good at mathematics when he goes to school. [...] Oh, I didn't mean to imply that it is because he draws a lot he will be good at mathematics. These abilities are independent. (Krzyżanowska, 2019, Ex. (6))

The unsuccessful intents to withdraw the suggested connection between antecedent and consequent in (48) and (49) are admittedly not very convincing, they sound rather odd. However, we need to be careful about how exactly we assume the inference of the reason relation to arise. We argue that the alternative utterance that is more informative than the conditional $A \rightarrow C$ is for example 'A and C' or 'C'. Thus we should be able to cancel the inference that *the speaker does not know whether 'A' or 'C' hold* like in (50) or in the false-link example from Krzyżanowska (2019) shown in (51).[4]

(50) If he isn't sick, he will vote today. In fact, he isn't sick and will vote today.

(51) If dolphins have fins, then they can't breathe under water. Oh, I didn't mean to imply that I do not know whether dolphins have fins and whether or not they can breathe under water. I know that they have fins and can't breathe under water.

Similarly to Krzyżanowska (2019), Skovgaard-Olsen et al. (2019) also consider participants ratings for the cancellation of the reason relation and argue against it being due to a conversational implicature. In both cases, the reason relation is attempted to be canceled by claiming that antecedent and consequent are not related. However, since we argue that the speaker choose to utter the conditional because of her epistemic uncertainty with respect to A and C, we do not consider an utterance like 'A and C are independent' or, more generally, that 'C has something to do with A', as alternative stronger utterances, but refer to a different utterance, for instance the conjunction 'A and C', the negation of which *is* cancellable (see (50) and (51)).

The fact that the speaker did not choose such a more informative utterance then allows the listener to infer that most likely there is a

4 With *false-link* conditionals they refer to conditionals whose antecedent and consequent are related content-wise but there is no causal link between them. I see the difference between false-link and missing-link conditionals in that the former might be assertable for speakers who are not aware of the correct causal link (e.g., children) whereas it is hard to imagine a situation in which an ordinary missing-link conditional would be felicitous. For our argumentation here, it does not matter whether the conditional is a false-link or an ordinary indicative conditional, assuming that both are assertable and consistent with the speaker's background knowledge. In case that one considers the false-link conditional unassertable, i.e. in particular inconsistent with the speaker's background knowledge, we treat it like a missing-link conditional: there is no reason for the speaker to assert the conditional as its assertion would result in a failure on part of the listener to see a connection between antecedent and consequent that would rationalize the speaker's utterance choice.

relation between antecedent and consequent: only under the assumption that A and C are dependent, the conditional is the speaker's best (most informative) utterance choice. Clearly, $A \rightarrow C$ can also be truthfully assertable when A and C are independent — however, in that case there would be a more informative utterance available to the speaker (that was not selected); if $A \rightarrow C$ is assertable, $P(c \mid a) \geq \theta$, and due to the independence $P(c \mid a) = P(c)$, therefore $P(c) \geq \theta$, so 'C' would be assertable as well.

Thus, the partition of the speaker's belief states into states in which there is a relation between both variables and states in which there is not is crucial here. The speaker's utterance of the conditional first leads to the inference that the speaker is not in a position to make a more informative assertion. But assuming that the speaker's target state is a state in which there is *no* relation between antecedent and consequent, the speaker could have said something more informative which is not necessarily the case if we assume that there *is* a relation between antecedent and consequent. Therefore, the implicature that the speaker does not know whether A and C hold comes along with the inference that most likely there is a relation between antecedent and consequent.

We, thus, may consider the relation an *indirect* inference of the pragmatic reasoning since the more informative utterance — which the speaker did not utter — only indirectly concerns the relation between antecedent and consequent. Lowering the probability of those states for which the more informative utterance applies, happens to increase the probability of states with a dependent relation.

One might be tempted to argue that the reason relation should then be cancellable after having cancelled the implicature about the speaker's beliefs concerning antecedent and consequent. Let us consider this by means of the example shown in (52). The implicature that the speaker does not know whether Alex, nor whether Riley comes to the party can be cancelled with any of the assertions in (52a)–(52d). The conjunction $a \wedge \neg r$ is not listed as option since it would contradict the selected conditional. Further, the claim that the speaker knows that Alex is coming but does not know whether Riley is coming ($a \wedge ?r$) seems odd because with the selected conditional, the speaker is committed to infer from Alex's coming that Riley is coming as well. Similarly, it seems odd, or at least less natural, to cancel the implicature by only saying that Riley is not coming without saying that Alex is not coming, since Alex's not coming is implied if one knows that Riley is not coming and that $a \rightarrow r$.

(52) If Alex comes to the party, Riley will come. ["$a \rightarrow r$"]
Actually, I know that . . .

a. Alex does not come, but Riley does. [$\neg a \wedge r$]
b. neither Alex nor Riley come. [$\neg a \wedge \neg r$]

c. Alex and Riley come. $[a \wedge r]$

d. Riley comes, but I don't know about Alex. $[a? \wedge r]$

Continuing any of (52a)–(52d) with the claim that whether Alex and Riley respectively come is completely independent of one another, seems indeed odd. However, this seems reasonable when we consider the indirect reasoning again. The listener does not only infer that the speaker was most likely not in a position to claim something more informative but further knows that if there was no connection between antecedent and consequent, the speaker would most likely have chosen a different utterance (e.g., 'Alex and Riley might come'). In other words, the implicated uncertainty is incompatible with A and C being independent, in that case, $A \to C$ would not be assertable in the first place (assuming a high conditional probability $P(c \mid a)$ as assertability condition). And, importantly, this still holds even if the speaker takes back the implicature about the uncertainty with respect to A and C.

However, it *is* possible to cancel the more specific inference that Riley comes *because of* Alex. Consider for example the following continuation of (52b): "But I don't mean to say that Riley would come because of Alex. They don't like each other and just have to take the same train, which was cancelled though."

REINFORCEABILITY. Similarly to the cancellability Example, we do *not* reinforce the inferred link directly which is what Krzyżanowska (2019) try to do (see (53)). The inference that we assume to be directly related to the speaker's selected utterance is again the inference that the speaker does not know whether antecedent and consequent are true or false. This inference can be reinforced as shown in (54) and (55).

(53) # If Dolphins have no gills, then they can't breathe under water. Actually, dolphins inability to breathe under water is related to their having no gills.

(54) If Dolphins have no gills, then they can't breathe under water. But I neither know whether Dolphins have gills nor whether they can breathe under water.

(55) If he isn't sick, he will vote today. But I do not know whether he is sick.

Again, we assume that it is the implicature about the speaker's epistemic uncertainty with respect to A and C, triggered by the fact that the speaker did not select a more informative utterance, that leads to the inference that antecedent and consequent are most likely connected. The implicature is naturally reinforceable (see first sentence of (56b)). So is the inference about the relation between A and C when formulated as in (56a) or, although a bit more cumbersome, in (56b). Only reinforcing the *concrete* relation between A and C (e.g., C because of A) without first reinforcing the implicature like in (56c) does not seem infelicitous either.

(56) If Dolphins have no gills, then they can't breathe under water.

a. If I knew that dolphins have gills, I would know that they can breathe underwater, because it is the gills that allow them to breathe underwater.

b. But I neither know whether Dolphins have gills nor whether they can breathe under water and since it is the gills that allow underwater animals to breathe underwater, learning that Dolphins do or do not have gills would allow me to infer that they can or cannot breathe underwater.

c. You know, it's the gills that allow underwater animals to breathe underwater.

Note the difference between how we formulated the reinforcement in (56) and how Krzyżanowska did in (53). They do not only claim that there is a relation between having no gills and the inability to breathe under water, they also state that dolphins *are* unable to breathe under water. In other words, they try to reinforce the relation between A and C while, at the same time, canceling the implicature about the speaker's uncertainty with respect to A and C. Under the assumption that the dependency relation is inferred based on the speaker's implicated uncertainty, it is not surprising that it is infelicitous to reinforce the former while canceling the latter.

Further, it seems more natural to reinforce the dependency relation by explicitly using causal verbs like 'because' or 'allow' instead of claiming that A and C are related. A possible reason for this observation might be that what is reinforced should be non-redundant. By saying 'C because of A' we particularize the utterance $A \rightarrow C$, which is not the case when we say that A and C are simply related. Assuming a high conditional probability $P(C \mid A)$ as assertability condition for the conditional $A \rightarrow C$, a mere relation is already literally communicated by the conditional just because of the assertability condition. Even if the relation between A and C is spurious, that is non-causal, there is a *probabilistic* relation — under the assumption that the conditional $A \rightarrow C$ is truthfully assertable. Take an example mentioned by Pearl and Mackenzie (2018, p.69): there was found a strong correlation between a nation's per capita chocolate consumption and its numbers of Nobel Prize winners.[5] Looking at the corresponding graph, one might say "If a nation has a high per capita chocolate

5 This correlation may be explained by a confounder variable: more people in Western countries eat chocolate and Nobel Prize winners tend to come from wealthy Western countries (see Pearl & Mackenzie, 2018, p.69).

consumption, the number of Nobel Prize winners from that nation is high". In that context, this seems to be a truthful utterance and the inference of a stochastic relation also seems to be reinforceable: the speaker may, for example, add: "There clearly is a relation, but the per capita chocolate consumption of a nation cannot be the cause for the number of Nobel prize winners of that nation".

Taken together, we demonstrated that the conversational implicature about the speaker's uncertainty with respect to antecedent and consequent is, as expected, cancellable as well as reinforceable. Further, it is possible to cancel the inferred relation to be *causal* and reinforce it to be *causal* or *stochastic* whereas reinforcing or canceling the fact that there *is* any relation is infelicitous, assuming the assertability of the conditional is based on the conditional probability being high, which already communicates a relation even though not necessarily of a causal nature.

Another experimental study that tested the reinforceability of the communicated relation between the antecedent and the consequent of conditionals in comparison to the reinforceability of semantic entailments and conversational implicatures was done by Rostworowski et al. (2021). They also confronted their participants with a speaker's utterance of a conditional and a subsequent statement (of the same speaker) that reinforces the relation between antecedent and consequent. They systematically differentiated between abductive, deductive and causal relations communicated with the conditional; (57) is one of their examples (without the corresponding vignettes) for a deductive relation and (58) for an abductive relation. But instead of asking for the naturalness of the speaker's reinforcement of the respective relation, they asked participants to rate its redundancy.

(57) "If we also invite the Smiths, this will make more than 20 guests. Inviting Smiths entails this."

(58) "If I don't invite the Smiths, I will show myself to be ungrateful to them. Not Inviting them will demonstrate my ingratitude."
Does the second sentence uttered by SPEAKER'S NAME is redundant/unnecessary as it repeats the information from the first one?[6]

Their results show that the reinforcement of the particular relation communicated with conditionals is, like semantic entailments, perceived as highly redundant. In particular, the added information about the relation between antecedent and consequent was rated much more redundant than were typical conversational implicatures, which seems to speak against a (classical) conversational implicature account for

6 Note that this how Rostworowski et al. (2021, p.7378) present the question asking for redundancy in their paper, which seems ungrammatical but this is probably just a typo in the paper.

the relation communicated with conditionals. Yet, these results seem to be compatible with our account.[7]

For the implicature about the speaker's epistemic uncertainty we should not expect reinforcing it to be perceived as redundant, which intuitively does not seem to be the case (e.g., "But I don't know whether we should invite the Smiths" as continuation of (57) or (58)). Concerning the reinforced relation between antecedent and consequent, it seems less clear whether or not we should expect it to be perceived as redundant information, assuming that it is inferred based on the implicature of the speaker's epistemic uncertainty. In a classical conversational implicature, the utterance that triggered the implicature (e.g., 'some') does not cease to be assertable if the speaker reinforces it (e.g. , by saying 'some but not all', 'some' remains assertable). Contrary to that, reinforcing the implicature that the speaker is uncertain about A and C would make the conditional unassertable if A and C were independent; the independence of A and C in combination with the speaker's uncertainty results in a probability distribution (representing the speaker's beliefs) like $\langle w_{ac} = 1/4, w_a = 1/4, w_c = 1/4, w_\emptyset = 1/4$ where $P(c \mid a) = 0.5 << 1$. This might be a possible reason for why the reinforced relation is perceived as redundant, as observed in the data from Rostworowski et al. (2021).

9.3 EXPERIMENTAL DATA ON BELIEF UPDATE WITH CONDITIONALS

Douven's puzzles that we discussed in Chapter 5 demonstrate that different utterance contexts can lead to different interpretations of conditionals, for instance with respect to the listener's *a posteriori* belief in the antecedent. Collins et al. (2020) empirically tested in a series of experiments how participants' beliefs about the probability of the antecedent, $P(A)$, the consequent, $P(C)$, and the conditional probability, $P(C \mid A)$, change after learning the conditional $A \to C$, while manipulating participants' prior beliefs of A and C.[8]

Table 15 summarizes Collins et al.'s (2020) results, some of which we have shown in Chapter 4 to be predicted by our model. These are the null-effects on the listener's beliefs about the antecedent, respectively the consequent, upon the interpretation of the speaker's utterance $A \to C$ and the increased belief in the conditional probability $P(C = c \mid A = a)$, which follows naturally from our assertability

7 Note that Rostworowski et al. (2021) do not exclude the possibility of a pragmatic explanation of the inferred link between antecedent and consequent, but they conclude by writing *"[. . .] it is no longer easy to believe that they (their results) can be explained in a purely pragmatic way or with keeping the traditional sharp division between semantics and pragmatics."*

8 They also considered the influence of the reliability of the speaker by manipulating the number of assertions of the conditional (by one or several people) or the speaker's expertise (expert vs. non-expert), which we will leave mostly aside here.

Prior (no testimony)		Posterior (after learning "If A, C")
$P(C \mid A)$	$<$	$P_{\text{"If A, C"}}(C \mid A)$
very low $P(A)$	$<$	$P_{\text{"If A, C"}}(A)$
$P(A)$	$=$	$P_{\text{"If A, C"}}(A)$
very high $P(A)$	$\geqslant^*$	$P_{\text{"If A, C"}}(A)$
very low $P(C)$	$<$	$P_{\text{"If A, C"}}(C)$
$P(C)$	$=$	$P_{\text{"If A, C"}}(C)$
very high $P(C)$	$\geqslant^*$	$P_{\text{"If A, C"}}(C)$

Table 15: Collins et al.'s (2020) Table 7, summarizing the results of their experiments on belief update with conditionals. Asterisks in the middle column denote marginal significance of the results.

condition for conditionals defined based on the relevant conditional probability.[9] In the following, we will turn to the remaining results from Collins et al. (2020), which show that the listeners' posterior beliefs change depending on their prior beliefs about the probability of the antecedent, and respectively the consequent, being very high or very low. In particular, Collins et al.'s (2020) participants' showed an increased belief in A and C after reading the conditional when *a priori* they had a very low belief in A and C. Similarly, they showed, although less strongly, a decreased belief in A and C when *a priori* they had a very strong belief in A and C.

Collins et al. (2020) considered several models (Bayesian belief networks) that were each, however, only able to capture a subset of the observed data. Like we do, the authors take into account that the conditional was uttered by a speaker, who they further model to be possibly unreliable. They model the speaker's utterance of the conditional $A \rightarrow C$ to be dependent only on whether or not $A \rightarrow C$ holds, which they represent by setting $P(C \mid A, X) = 1$ where X denotes 'the indicative conditional 'If A, C' holds'. In some versions of their model they additionally make the speaker's utterance of the conditional $A \rightarrow C$ dependent on whether A or C or both hold. What they do not model, however, are the alternative utterances that the speaker

9 To be precise, we did not explicitly consider the pragmatic listener's point of view with respect to the inferred beliefs in antecedent and consequent, but focused on the speaker's side, whose predictions are, however, the basis for the pragmatic listener's interpretation. More precisely, we considered the situations in which the speaker's best utterance would be a conditional (corresponding to a hyperrational speaker with the RSA rationality parameter $\alpha \rightsquigarrow 1$), which were situations in which the speaker is uncertain about both, A and C (see Chapter 4, Figure 7). Under the assumption that the listener does not have any specific prior knowledge, and assuming that the listener takes over the inferred beliefs of the speaker, the conditional should, thus, not change the listener's beliefs in antecedent and consequent.

could have chosen, but didn't. This is, however, crucial if we assume that the speaker is a rational agent who does not select an utterance just because it is true, but because there is a good reason for choosing it. Moreover, Collins et al. do not seem to explicitly differentiate between the speaker's beliefs and the listener's beliefs. Both are aspects that we argue to be crucial for the listener's interpretation and belief update.

Since we generally expect conditionals to be uttered by speakers who are uncertain about antecedent and consequent, the interlocutors' beliefs should be modeled as different in this regard when the listener is assumed to be highly convinced of the truth or falsity of C or A. Therefore, the speaker's beliefs should not be identical to the listener's, that is, participants' beliefs: if that was the case, a cooperative speaker would have said something else, namely something more informative. In the examples from Collins et al., two of which are shown below in (59) and (60), the speaker's and listener's *a priori* belief in the antecedent may indeed both be very low / high, however, the fact that the speaker did utter the conditional suggests that she has some other knowledge that we, as listeners, do not have.

(59) Imagine you are visiting a Liberal Arts College.
Sue tells you, 'If Lisa, a student, is majoring in astrophysics, then she's working late in the library'.

(60) Adam is at a large car dealership which specializes in mid-range cars.
He tells you, 'If a car is a Rolls Royce, then it's black'.

In (59), Sue may have the same belief as we do, namely that studying astrophysics at an Arts College is rather unlikely, yet since Sue did utter (59), she seems to know more than we do — otherwise, why would she claim (59)? She might, for example, know that it is at least possible to study astrophysics at that Liberal Arts College, which one would probably not expect without having heard Sue's utterance. Similarly, Adam's utterance in (60) suggests that he has some more knowledge about the cars at the dealership than we do; otherwise his utterance does not seem to be reasonable. He might, for instance, already have seen a black Rolls Royce. In both examples, it is reasonable for the listener to increase her belief in the antecedent, assuming that the speakers are cooperative and reliable.

Let us consider our model in light of (61), an example based on those from Collins et al.

(61) Imagine you overhear a conversation between students in the cafeteria of your university.
Student: *"If Mary, our Frisbee instructor, has an exam tomorrow, she is in the library right now."* [$A \rightarrow C$]
You know the Frisbee instructor Mary and you know that she hardly ever goes to the library.

The listener's prior beliefs about Mary being in the library to be low does not influence her interpretation of the speaker's utterance (the conditional in (61)) *per se*. The influence that the listener's prior beliefs have rather concern the consequences that the listener draws — under consideration of her own beliefs — upon receiving the new information from the speaker. Put differently, in a first step, the listener processes the speaker's utterance from which she makes inferences about the *speaker's* beliefs. In a second step, the listener then integrates her own beliefs with the newly received information from the speaker.[10] Let us consider this for the conditional in (61) from above. Our model predicts that in a default context, the interpretation of the conditional $A \rightarrow C$ comprises three main inferences:

Most likely ...

1. the speaker is uncertain about A
2. the speaker is uncertain about C
3. the speaker believes that there is a relation between A and C

We can represent the listener's beliefs about the speaker's beliefs (abbreviated as LS) in a condensed manner by reference to the expected probability distribution over variables A and C:

$$E[P^{LS}(A, C \mid u)] = \sum_{s \in S} P^{(s)}(A, C) \cdot P_{PL}(s \mid u)$$

Let us assume the simplest case, where this expected distribution, that is, the listener's inferred beliefs about the speaker's beliefs, is given by $W^{LS} = \langle w^{LS}_{ac}, w^{LS}_{a} w^{LS}_{c}, w^{LS}_{\varnothing} \rangle = \langle 0.5, 0, 0, 0.5 \rangle$ representing the three inferences with respect to the conditional utterance given above.[11] In other words, W^{LS} represents what the listener infers about the speaker's beliefs upon the speaker made the conditional utterance. Thereby the listener rationalizes the speaker's utterance: if the speaker had the same knowledge as the listener (C is very unlikely), the listener would not expect the speaker to say $A \rightarrow C$. But the speaker chose precisely this utterance, therefore the speaker's beliefs must differ from the listener's beliefs, otherwise the speaker's

10 I do not want to make a claim about the timing of the two described steps. From a sentence processing perspective, it is likely that the integration of the listener's own beliefs with the speaker's beliefs as inferred by the speaker's utterance, starts before the speaker finishes her utterance, that is before the listener has access to the complete utterance. At least, it is observed that in conversations, speakers start responding before the interlocutor's turn has ended (e.g., see Levinson & Torreira, 2015).

11 In case of a situation like in (60) or (59) in which the speaker is also assumed to have a rather low belief in the antecedent, W^{LS} would more realistically assign a lower probability to w_{ac}, e.g., $W^{LS} = \langle 0.1, 0, 0, 0.9 \rangle$, where the probability of the antecedent remains low (0.1) but larger than the listener's own beliefs.

utterance would not be rational. We may consider the listener's beliefs as being equal to $W^L = \langle w_c^L, w_\emptyset^L \rangle = \langle 0.02, 0.98 \rangle$ or when taking into account the uncertainty about Mary having an exam $W^L = \langle w_{ac}^L, w_a^L, w_c^L, w_\emptyset^L \rangle = \langle 0.02, 0.49, 0, 0.49 \rangle$.

The question now is how this belief is integrated with the inferred belief of the speaker. An exemplary situation representing W^L from above is the following: the listener frequently goes to the library but has never seen Mary there, thus the listener has a strong prior belief that $C = \neg c$, but does not know Mary well enough to exclude that she goes there from time to time, for instance when she has an exam. In that case, the listener might update her own beliefs by fully taking over the inferred beliefs of the speaker (W^{LS}) which corresponds to an increase in the listener's beliefs for $C = c$ (in the example from 0.02 to 0.5) as observed in the data from Collins et al.

We would, however, expect a different inference on part of the listener if the listener was fully convinced that Mary would never show up in the library, for example represented by a probability distribution $W^L = \langle w_{ac}^L, w_a^L, w_c^L, w_\emptyset^L \rangle = \langle 0, 0.5, 0, 0.5 \rangle$. In that case, the listener should rather conclude, from the speaker's utterance of the conditional $A \rightarrow C$, that the speaker does not know Mary very well and particularly does not know that she would *never* go to the library. In this case, the listener's own belief about the consequent would remain identical to her belief prior to hearing the conditional.

Yet another case is the following: the listener just talked to Mary on the phone saying that she is at home. Therefore, the listener knows that Mary is not in the library ($C = \neg c$) and is thus very likely not willing to increase her belief in the consequent. The listener may, however believe what the speaker said ($A \rightarrow C$) but then rather conclude that, given her knowledge that $C = \neg c$, the antecedent cannot be true. This scenario corresponds to the inferences drawn in the *Garden Party Example* discussed in Chapter 5, in which the strong belief in the consequent being false was inferred from world knowledge whereas here, it is directly given.

Therefore, to decide which interpretation is more likely does not only depend on the listener's prior beliefs but also on the listener's readiness to revise her own beliefs in one way or another. Here, we assumed the speaker to be fully reliable in the sense that we do not only expect the speaker to only say what she believes to be true but also assume that what the speaker says is usually correct. If the speaker was not fully reliable, the listener should not simply take over the inferred beliefs of the speaker completely, but might end up with a mixture of both. For instance, by taking into account both beliefs equally, the speaker's beliefs (as inferred by the listener) and the listener's own beliefs, the listener may end up with a belief like $W^{L'}$:

$$W^{L'} = \langle w_{ac}^{L'}, w_{a}^{L'}, w_{c}^{L'}, w_{\emptyset}^{L'} \rangle \text{ where} \quad [47]$$

$$w_x^{L'} = 1/2 \cdot w_x^{LS} + 1/2 \cdot w_x^{L} \; ; x \in (ac, a, c, \emptyset)$$

$$\text{For } W^{L} = \langle 0.02, 0.49, 0.0, 0.49 \rangle \text{ and } W^{LS} = \langle 0.5, 0, 0, 0.5 \rangle,$$

$$W^{L'} = \langle w_{ac}^{L'} = 0.26, w_{a}^{L'} = 0.245, w_{c}^{L'} = 0, w_{\emptyset}^{L'} = 0.495 \rangle \quad [48]$$

In this case, the conditional probability $P(C = c \mid A = a)$ also increases; from a previous value of almost 0 (≈ 0.04) to 0.51. Since the listener does not take the speaker's utterance at face value, the updated belief $P^{L'}(C = c \mid A = a)$ is still rather low, at least not sufficiently large to reasonably justify the assertion of the utterance that the speaker chose ($A \rightarrow C$). This is reasonable as the listener is skeptical about the speaker's claim, thus, there is no reason for the listener to make the same claim.

9.4 LIMITATIONS OF OUR MODEL

The model that we have presented is promising not only because it is able to explain several observed phenomena in the communication with conditionals, as demonstrated in this thesis, but further due to its flexibility extensions are conceivable that might handle commonly observed interpretations of conditionals that we have not considered in much detail here (e.g., biscuit conditionals). However, the model clearly has its limitations. Except for the concrete cases addressed in Chapter 5 we have looked at simple scenarios which comprised two variables only. It is however reasonable to assume that most inferences drawn by natural language users in everyday situations involve more than two variables. Therefore, it would be desirable to represent more complex situations as well. This is in principle possible without requiring much adaptations, but the set of states would quickly become very large. We might reduce the number of generated states per causal relation, for example to only a handful of stereotypical cases. Like we defined distributions for high and low values (beta(10,1), beta(1,10)) and values in between (beta(5,5)) from which we sampled to generate the set of probability distributions for each causal relation, we might use fixed values for each category (e.g, high $\hat{=}$ 1 , low $\hat{=}$ 0, uncertain $\hat{=}$ 0.5), adding categories on demand (e.g., quite certain high $\hat{=}$ 0.95, quite certain low $\hat{=}$ 0.05, etc.). This is a tradeoff between the number of represented variables and how they relate to each other and the complexity of the represented beliefs. When taking into account only two variables we can represent very rich beliefs as we did here, which, however, quickly becomes computational unfeasible when introducing more variables.

These considerations may create the impression that our model could be much less complex than it is now. To model the empirical data from Chapter 6, a rich representation of participants' beliefs (in form of many different probability distributions) was necessary. However, whether participants' beliefs would be sufficiently represented by less complex data, for example by making them choose between several concrete options instead of using sliders, is a different matter. We saw that when given the possibility, participants were very precise in their answers, as we observed many different slider ratings going beyond the use of stereotypical values like 0, 1 and 0.5. Whether participants' utterance choices are actually based on these rich beliefs is yet a different question. They may have quite nuanced beliefs about the probabilities of the blocks to fall which they convey in the PE-task of our experiment, yet fall back on stereotypical categories when describing the corresponding scenes in the UC-task which would need to be tested further.

Another limitation of our model, which RSA-models generally face, is the predefined finite set of alternative utterances. In natural conversations, speakers are not constraint in the utterances that they select. It is, however, plausible to assume that for concrete contexts, the set of possible utterances that the speaker chooses from is in some way limited, although not in their exact formulations. In our experiment, participants could use various different formulations for the same meaning (e.g., 'if B, G' and 'G if B' were both possible options) that we then summarized into a set of standardized expressions in the model. How to define this standardized set of expressions is however up to us as modelers. The analysis of the custom responses that participants could give in the UC-task of our experiment revealed that certain utterances would probably have been good to include in the available set of responses. An example are conjunctions using 'might' in one conjunct, like in the utterance 'the green block falls and the blue block might fall'.

In terms of the considered alternative utterances, an interesting extension of our model may investigate the relationship between conditionals, causality and causal language further by considering different alternative utterances, including explicit causal expressions that we did not address here. We explored the interpretation of conditionals under the assumption that conditionals are used to communicate stochastic information about co-occurrence, while information about the underlying causal structure is communicated implicitly in our model. As discussed in Section 9.2.1, in some circumstances, the speaker's communicative goal might however go further than this and include the communication about causal information as well, bringing into play explicit causal language.

10 CONCLUSION

Natural language conditionals are frequently used in everyday conversations. Yet, it is difficult to formally capture the meaning of the word 'if' so that it accounts for the many different interpretations associated with such 'if'-sentences. In this thesis, I combined two independently motivated and well-established formalizations, causal Bayesian networks and the Rational-Speech-Act-model, to formalize speakers' uses of conditionals and listeners' interpretations thereof. This approach has proved useful in showing that common observations within the communication with conditionals can be explained through pragmatic processes, relying on the conditional probability as simple assertability condition for conditionals. Among these observations is the reason relation commonly associated with ordinary indicative conditionals, which has recently been strongly debated since there is disagreement about how this inference arises. Due to the flexibility of our approach, in terms of the utterance contexts that can be represented, the model may further be used as a starting point for systematic investigations of other kinds of conditionals, for example subjunctive conditionals, that were not addressed in this thesis, or the factors influencing the interpretation of special types of conditionals like biscuit conditionals. Moreover, the developed model makes quantitative predictions that can be empirically tested which helps to gain further insights about peoples' use and interpretation of conditionals, in particular from the divergences between model predictions and empirical data. These allow to refine the model, and thus our understanding of the factors that contribute to the observed use and interpretations of conditionals. This is an important step toward a formal and concrete explanation of the assumed underlying pragmatic processes, moving away from more or less vague textual descriptions toward measurable and, thus, falsifiable, quantitative predictions.

Part V

APPENDIX

A SUPPLEMENTARY MATERIAL PE-UC-TASK EXPERIMENT

Table 16: Particpants' custom responses with selected responses in UC task.

trial	custom_response	response
if1_hh	both the blue and the green blocks will fall.	both blocks fall
if1_hh	both blocks fall	the blue block and the green block fall
ind_ul	the blue block does not fall and the green block falls	the green block does not fall and the blue block falls
if2_ll	both blocks fall	the blue block and the green block fall
if1_u-Lh	the green block does not fall and the blue block falls	the blue block falls and the green block does not fall
ind_ll	the green blocks falls and the blue block does not fall	the blue block does not fall and the green block falls
ind_uh	the blue block might not fall but the green block does.	both blocks fall
ind_ul	both blocks might fall.	the blue block might fall
ind_ul	both blocks might fall	both blocks fall
ind_ul	both blocks might fall	the blue block might fall
if2_u-Ll	if the blue block falls the green block falls	the blue block might fall
ind_ul	and the blue block might fall	the green block does not fall
ind_uh	and the green block falls	the blue block might fall
if2_hl	if the blue block falls the green block might fall	if the blue block falls the green block falls as well
ind_ll	the blue block does not fall but the green block might	the blue block does not fall but the green block falls
ind_ll	the green box might fall	neither block falls
if2_hl	both could fall	both blocks fall
if2_u-Ll	both blocks might fall	both blocks fall
ind_ul	both blocks might fall	the blue block might fall
ind_hl	blue block falls but green might not.	the blue block falls and the green block does not fall
if2_ul	both blocks might not fall	neither block falls
if2_hl	blue falls green might not	both blocks fall
if2_u-Ll	the blue block falls and the green block might fall	the blue block falls and the green block falls as well

continues on next page . . .

Table 16 – continued from previous page

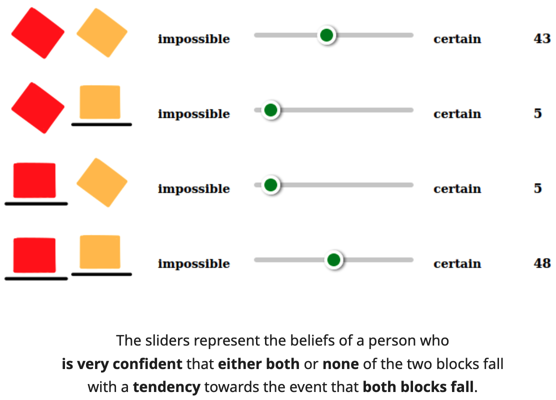
trial	custom_response	response
ind_ul	the green block and blue block might not fall	the green block might fall
if2_ll	the blue block falls and the green box might fall	the blue block falls and the green block falls as well
ind_ll	the green block falls but the blue block does not fall	the green block falls and the blue block does not fall
if2_ul	the blue block falls and the the green block might fall	the blue block falls but the green block does not fall
if2_ll	the blue block might fall and then the green block will fall	the green block might fall
if1_hh	the blue falls, the green might as well	the blue block falls and the green block falls as well
if2_ul	green falls, green might	the blue block falls but the green block does not
ind_ul	green might, green will not	neither block falls
if2_ll	both fall	both blocks fall
ind_hh	green falls, blue might	the green block falls and the blue block falls as well
if1_uh	both might fall	the blue block and the green block fall
ind_hl	blue falls, green won't	the blue block falls
if2_hl	both fall	both blocks fall
ind_uh	green might, green won't	the blue block might fall
if1_lh	neither fall	neither block falls
if2_u-Ll	blue will fall, green might	the green block might fall
ind_ll	neither falls	neither block falls
if1_hh	the blue block falls and the green block might	the green block might fall
ind_uh	the green block falls and the blue block might fall	both blocks fall
ind_ul	the blue block will probably fall and the green block will probably not fall	the blue block falls but the green block does not fall
if2_ul	the blue block might fall. if the blue block falls, the green block might fall. if the blue block does not fall, the green block does not fall.	the blue block might fall
ind_ll	the blue block does not fall and the green block might fall	neither block falls
if2_ll	if the blue block falls, the green block might fall	if the blue block falls the green block falls as well
ind_uh	the green block falls and the blue block might fall	the blue block might fall
ind_hl	the blue block falls and the green block might fall	the green block might fall

continues on next page ...

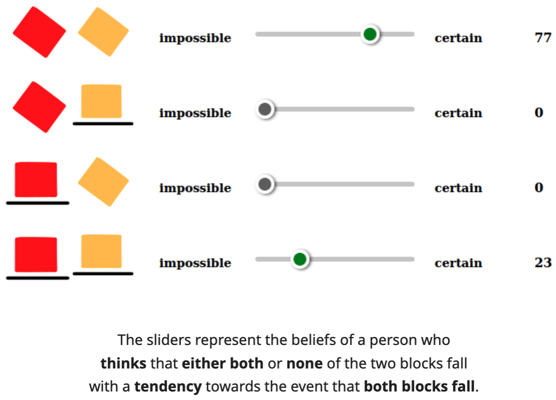
Table 16 – continued from previous page

trial	custom_response	response
if2_hl	the blue block falls and the green block might fall	the blue block falls
ind_ul	both blocks might fall	the blue block might fall
if2_u-Ll	the blue block falls and the green block might fall	the blue block falls
if2_u-Ll	the blue block does not fall but the green block might	the green block does not fall but the blue block falls
if1_uh	both blocks might fall	the blue block might fall
if2_ll	if the blue block falls it'll cause the green block to fall too	the green block falls if the blue block falls
if1_uh	if the blue block falls it'll cause the green block to fall too	the green block falls if the blue block falls
ind_ul	the green block falls while the blue block doesn't	the green block falls and the blue block does not fall
if2_ul	if the blue block falls it'll cause the green block to fall too	the green block falls if the blue block falls
ind_hl	the blue block falls but the green block doesn't fall	the blue block falls and the green block does not fall
ind_ul	the blue block falls but the green block might not	the blue block falls but the green block does not
ind_uh	the green block falls but the blue block might not	the green block falls but the blue block does not
ind_ul	the blue block might fall but the green block won't fall	the blue block might fall
if2_ll	the blue block might fall which will make the green block fall.	the blue block might fall
if1_uh	the blue block might fall which will make the green block fall.	the blue block might fall
ind_ul	the blue block might fall but the green block won't fall.	the blue block might fall
if2_u-Ll	both the green block and the blue block might not fall.	the green block might not fall
ind_uh	both blocks might fall.	the green block might not fall
if2_hl	both blocks might fall.	both blocks fall
if2_u-Ll	the blue block falls but the green block might not	the blue block falls but the green block does not
if2_ul	the blue block falls but the green block might not	the blue block falls but the green block does not fall
ind_hh	the green block falls and the blue block might not	the green block falls and the blue block does not fall
ind_ul	the blue block might fall, the green block does not fall.	the blue block might fall

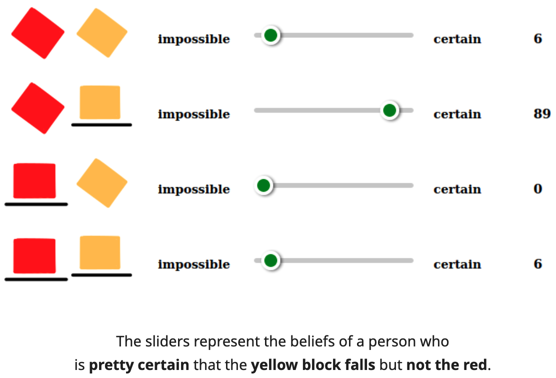
Figure 43: Slider-choice trials from the training phase for the PE-task. Participants had to answer by clicking on yes/no buttons.



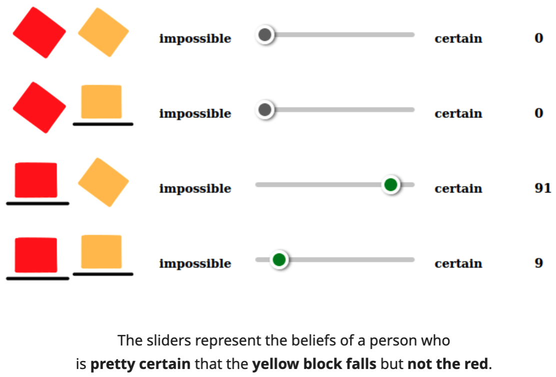
(a) Correct answer: no



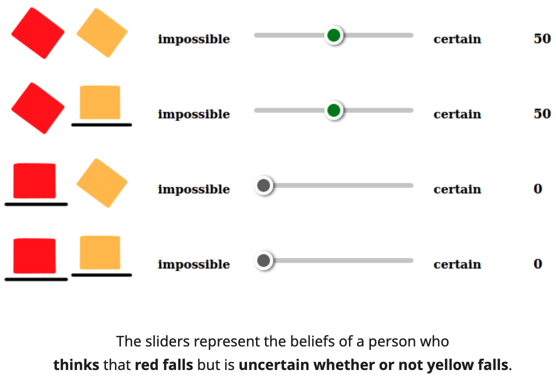
(b) Correct answer: yes



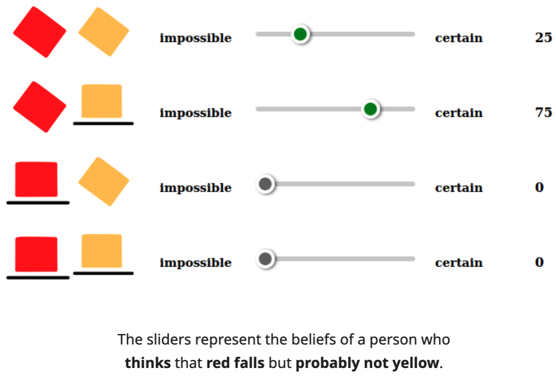
(c) Correct answer: no



(d) Correct answer: yes

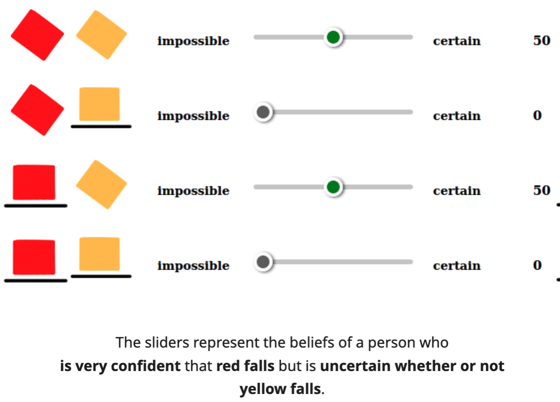


(e) Correct answer: yes

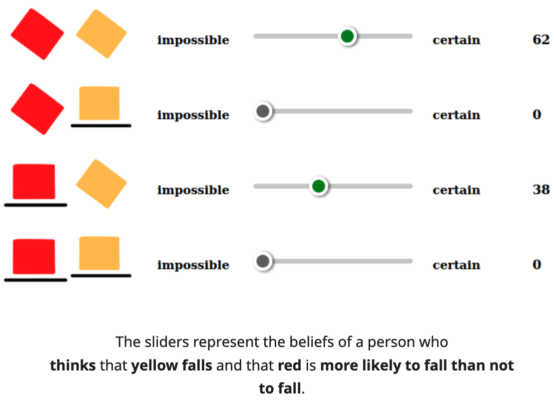


(f) Correct answer: yes

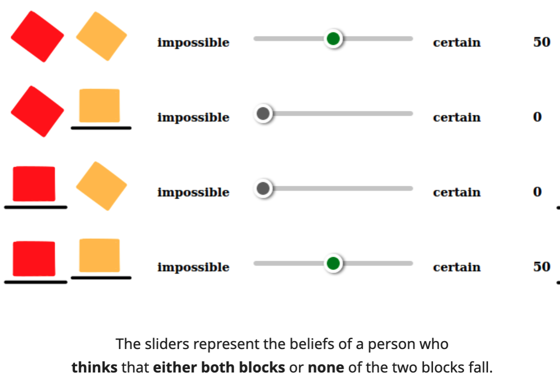
Figure 43: Slider-choice trials **– continued from previous page**



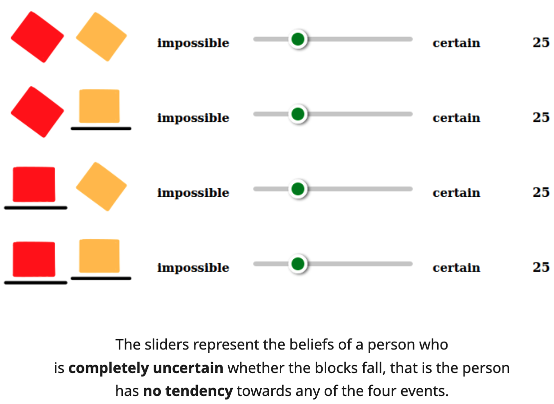
(a) Correct answer: no



(b) Correct answer: yes

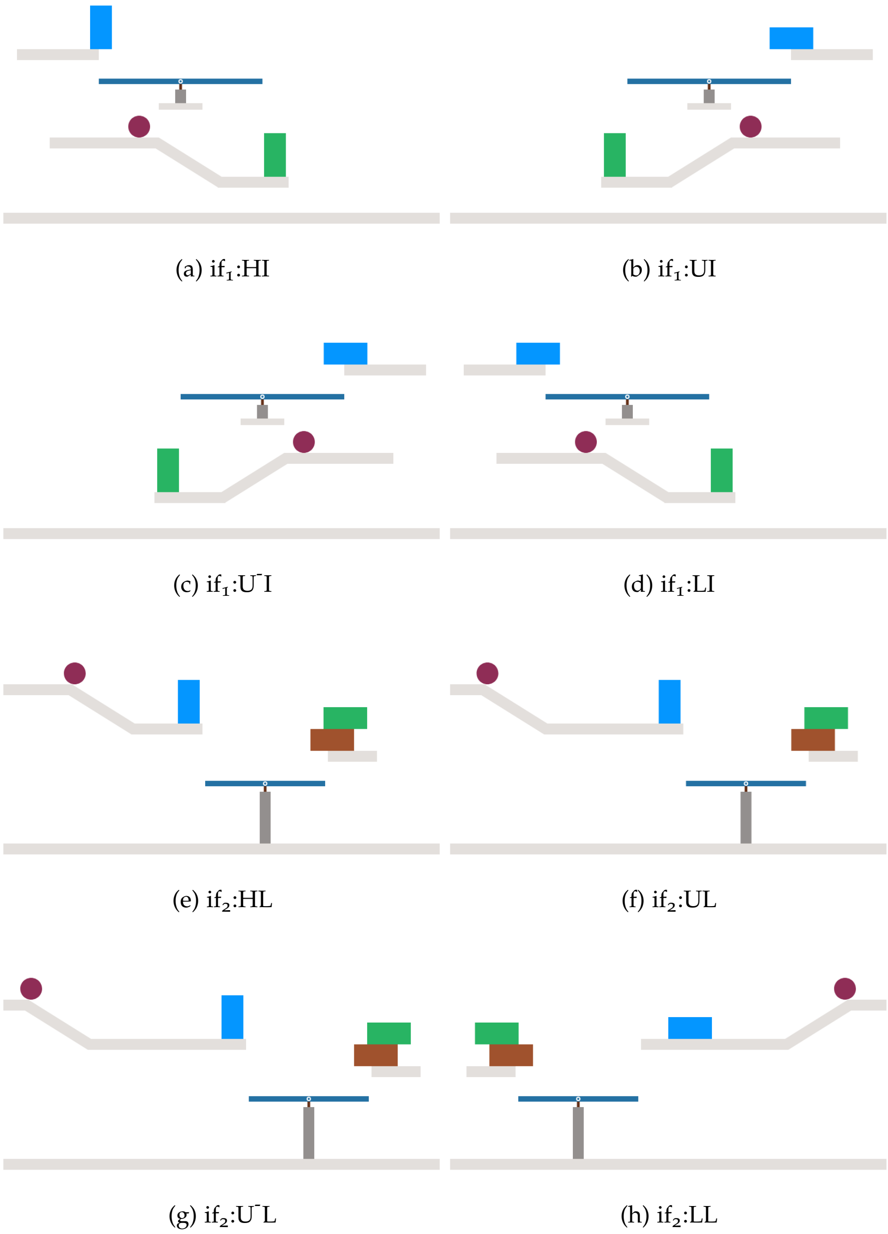


(c) Correct answer: yes



(d) Correct answer: yes

Figure 44: Visual scenes of block arrangements from the test phase, for the if_1- and if_2-conditions. Stimuli names represent the relation before the colon and the prior probability of the blue (1^{st} letter) and the green (2^{nd} letter) block to fall — respectively without the influence of the other block — after the colon. 'H' means high, 'L' low, 'I' impossible. 'U' means uncertain, and 'U¯' is a bit less uncertain than 'U'; in (c) the blue block is positioned slightly more on top of the platform than in (b), in (g) the platform on which the blue block is positioned is slightly longer than in (f).



(a) if_1:HI

(b) if_1:UI

(c) if_1:U¯I

(d) if_1:LI

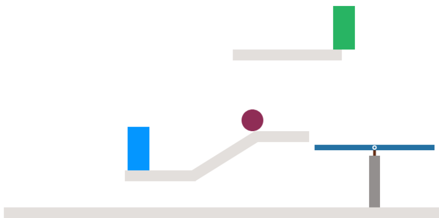
(e) if_2:HL

(f) if_2:UL

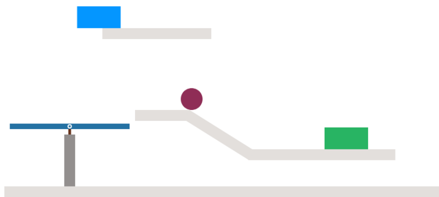
(g) if_2:U¯L

(h) if_2:LL

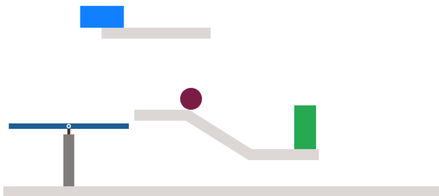
Figure 45: Visual scenes of block arrangements from the test phase, for the independent-condition (ind). Stimuli names represent the relation before the colon and the prior probability of the blue (1st letter) and the green (2nd letter) block to fall after the colon.



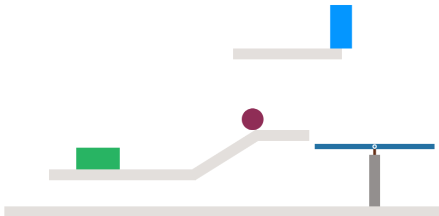
(a) ind:HH



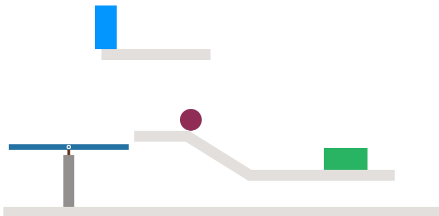
(b) ind:HL



(c) ind:UH



(d) ind:UL



(e) ind:LL

Figure 46: Posterior Predictive distributions for the Dirichlet-regression model of the PE-task data in independent conditions in which the prior probability of the blue block to fall was low (top), uncertain (middle) or high (bottom). Dark blue: observed data light blue: new data.

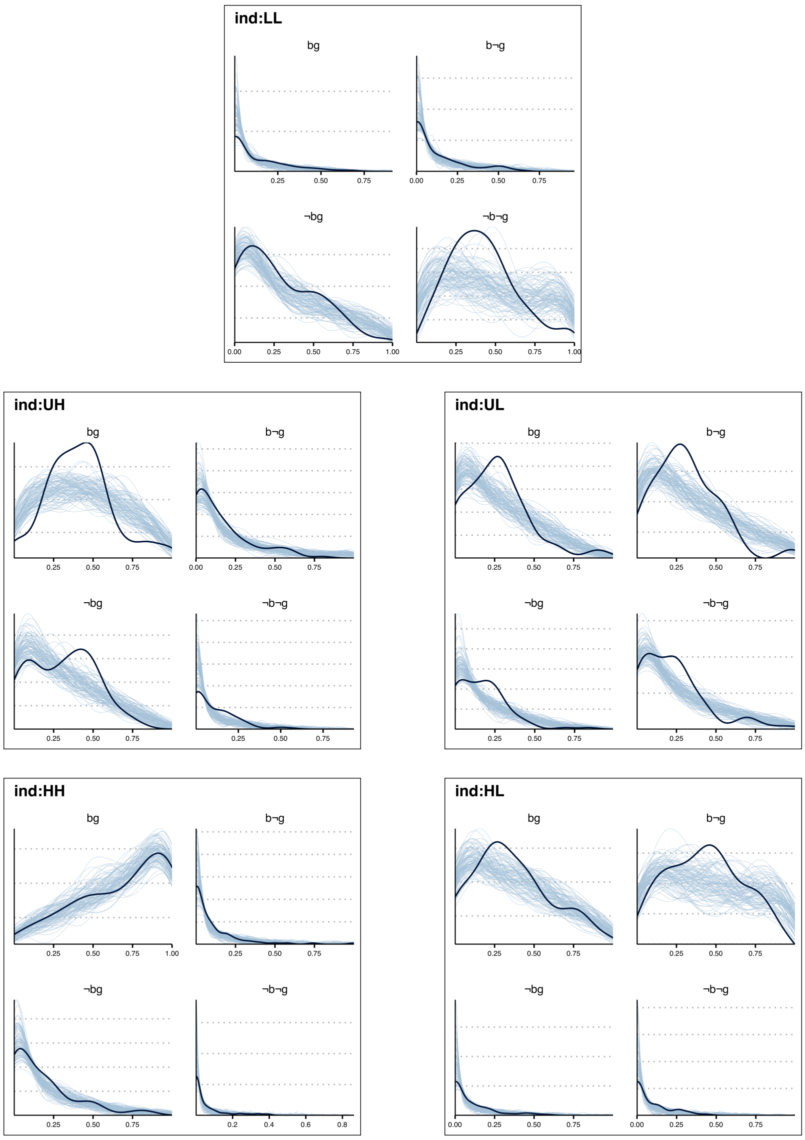


Figure 47: Posterior predictive distributions for the Dirichlet-regression model of the PE-task data in dependent conditions. Dark blue: observed data, light blue: new data.

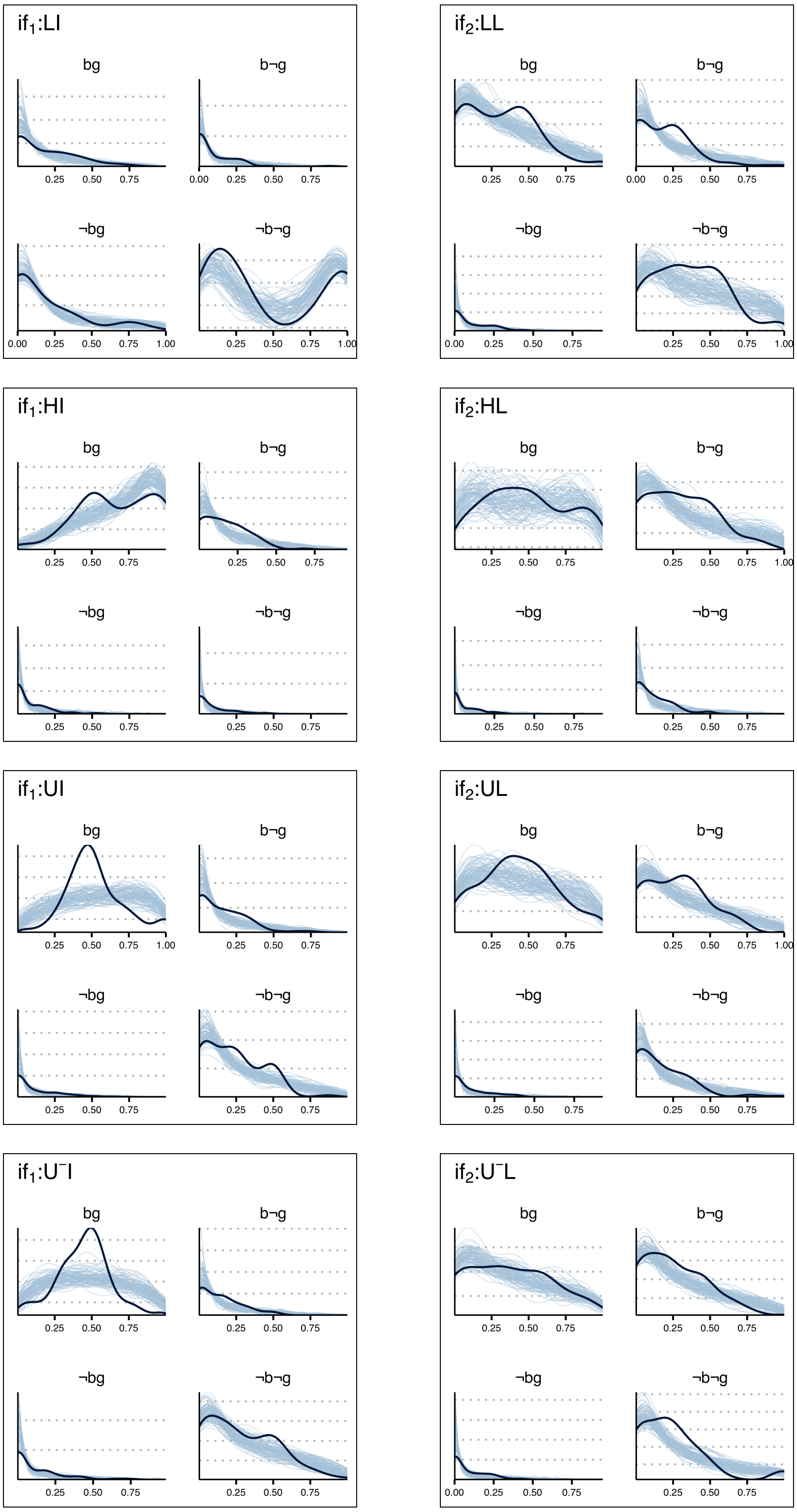


Figure 48: Posterior Predictive distributions for the ZOIB-model of the PE-task data in independent conditions in which the prior probability of the blue block to fall was low (top), uncertain (middle) or high (bottom). Dark blue: observed data light blue: new data.

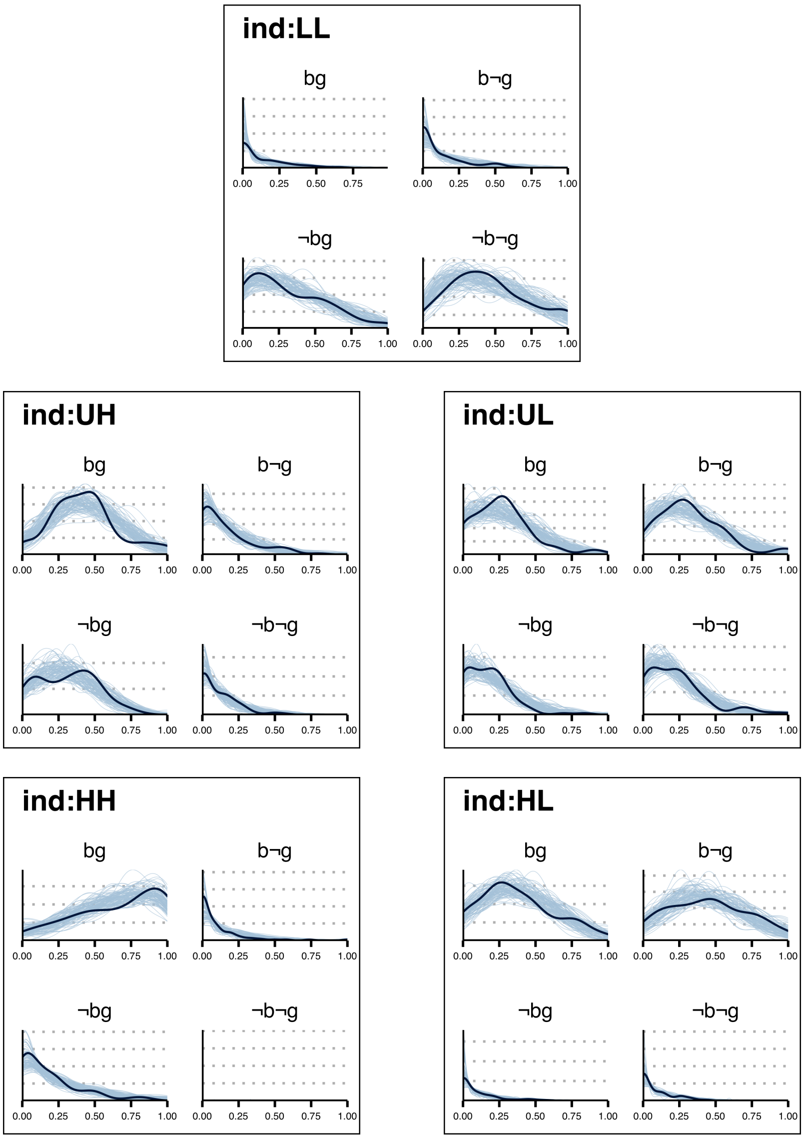


Figure 49: Posterior predictive distributions for the ZOIB-model of the PE-task data in dependent conditions. Dark blue: observed data, light blue: new data.

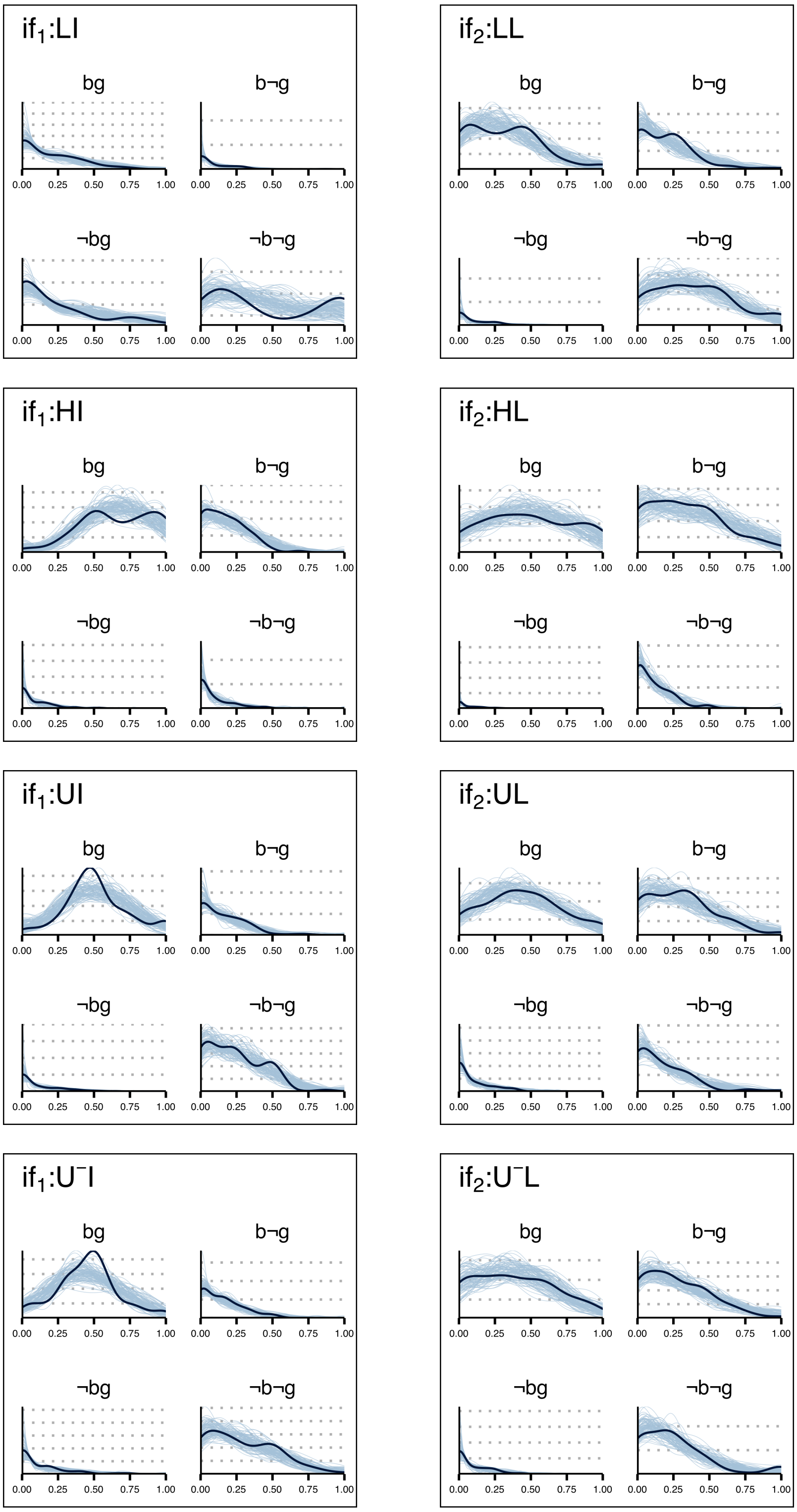


Figure 50: Approximated Posterior $P(\theta \mid P_{lit}, D^{UC}, D^{PE})$ with 95% highest density intervals, 5000 MCMC-samples (lag=10) after a burn-in period of 10,000 samples.

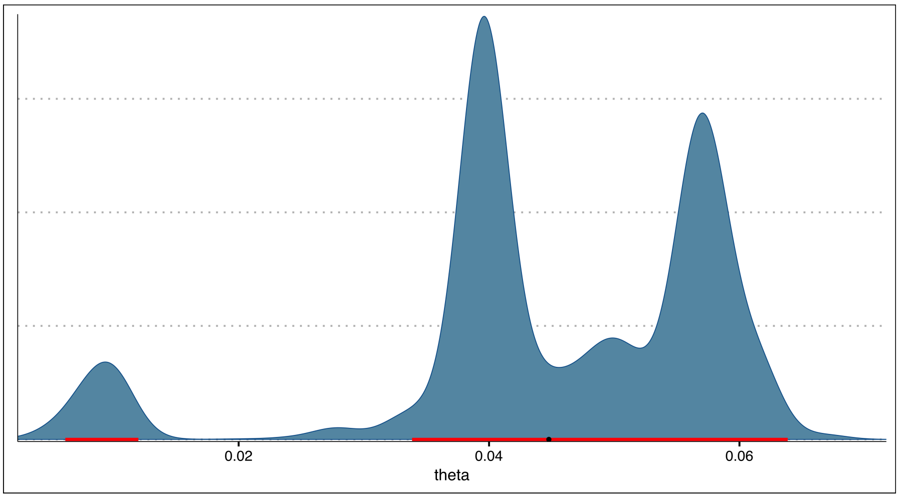


Figure 51: Pairs plot for MCMC-samples approximating the posterior $P(\alpha, \theta \mid P_S, D^{UC}, D^{PE})$.

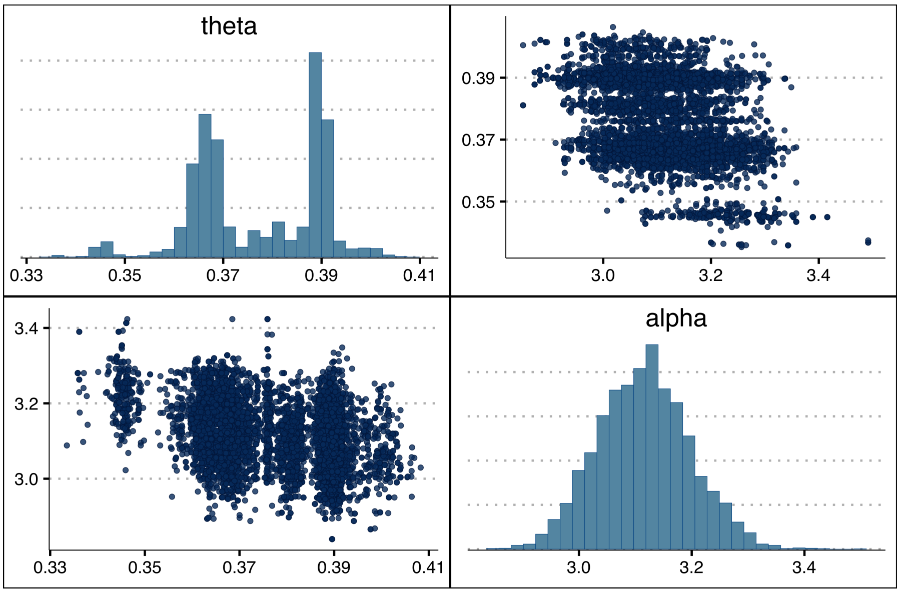
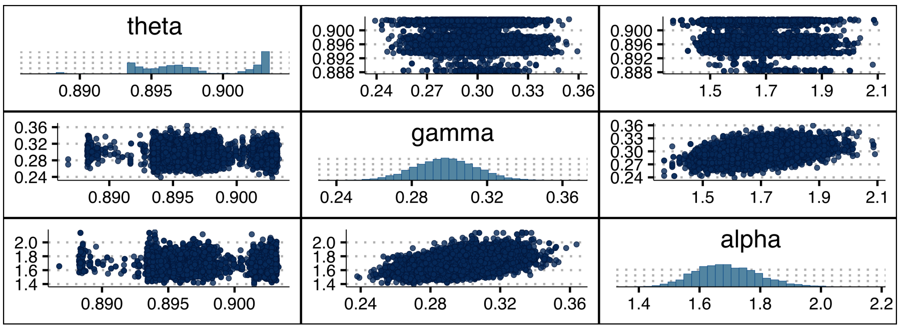


Figure 52: Pairs plot for MCMC-samples approximating the posterior $P(\alpha, \theta, \gamma \mid P_S, D^{UC}, D^{PE})$.



BIBLIOGRAPHY

Adams, E. W. (1965). On the Logic of Conditionals. *Inquiry*, *8*, 166–197.

Adams, E. W. (1975). *The logic of conditionals: An application of probability to deductive logic*. Reidel.

Aloni, M. (2007). Expressing Ignorance or Indifference (J. Carbonell & J. Siekmann, Eds.). In J. Carbonell & J. Siekmann (Eds.), *Logic, Language, and Computation*, Berlin, Heidelberg, Springer. https://doi.org/10.1007/978-3-540-75144-1_1

Anderson, A. R. (1951). A Note on Subjunctive and Counterfactual Conditionals. *Analysis*, *12*(2)jstor 3327037, 35–38. https://doi.org/10.2307/3327037

Atlas, J. D., & Levinson, S. C. (1981). It-clefts, informativeness and logical form: Radical pragmatics (revised standard version). In *Radical pragmatics* (pp. 1–62). Academic Press.

Austin, J. (1956). Ifs and cans. In *Proceedings of the British Academy* (pp. 109–132). Cambridge University Press. https://doi.org/10.2307/2964530

Barrouillet, P., Grosset, N., & Lecas, J.-F. (2000). Conditional reasoning by mental models: Chronometric and developmental evidence. *Cognition*, *75*(3), 237–266. https://doi.org/10.1016/S0010-0277(00)00066-4

Beller, A., Bennett, E., & Gerstenberg, T. (2020). The language of causation, In *Proceedings of the Cognitive Science Society*.

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

Birner, B. J. (2013). Introduction to Pragmatics, 193.

Blakemore, D., & Carston, R. (2005). The pragmatics of sentential coordination with and. *Lingua*, *115*(4), 569–589. https://doi.org/10.1016/j.lingua.2003.09.016

Briggs, R. (2012). Interventionist Counterfactuals. *Philosophical Studies*, *160*(1), 139–166. https://doi.org/10.1007/s11098-012-9908-5

Brochhagen, T., Franke, M., & van Rooij, R. (2018). Coevolution of Lexical Meaning and Pragmatic Use. *Cognitive Science*, *42*(8), 2757–2789. https://doi.org/10.1111/cogs.12681

Bürkner, P.-C. (2017). Brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C., & Vuorre, M. (2019). Ordinal Regression Models in Psychology: A Tutorial. *Advances in Methods and Practices in Psychological Science*, *2*(1), 77–101. https://doi.org/10.1177/2515245918823199

Burnett, H. (2019). Signalling games, sociolinguistic variation and the construction of style. *Linguistics and Philosophy*, *42*(5), 419–450. https://doi.org/10.1007/s10988-018-9254-y

Cariani, F., & Rips, L. J. (2016). *Experimenting with (Conditional) Perfection: Tests of the Exhaustivity Theory*.

Cariani, F., & Rips, L. J. (2023). Experimenting with (Conditional) Perfection: Tests of the Exhaustivity Theory. In S. Kaufmann, O. David, & G. Sharma (Eds.), *Conditionals: Logic, Linguistics and Psychology* (pp. 235–274). Springer International Publishing.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological review*, *104*(2), 367.

Collins, P. J., Krzyżanowska, K., Hartmann, S., Wheeler, G., & Hahn, U. (2020). Conditionals and testimony. *Cognitive Psychology*, *122*, 101329. https://doi.org/10.1016/j.cogpsych.2020.101329

Comrie, B. (1986). Conditionals: A typology. In A. T. Meulen, C. A. Ferguson, E. C. Traugott, & J. S. Reilly (Eds.), *On Conditionals* (pp. 77–100). Cambridge, Cambridge University Press. https://doi.org/10.1017/CBO9780511753466.005

Cruz, N., Baratgin, J., Oaksford, M., & Over, D. E. (2015). Bayesian reasoning with ifs and ands and ors. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00192

Cruz, N., Over, D., Oaksford, M., & Baratgin, J. (2016). Centering and the meaning of conditionals. Proceedings of the 38th Annual Conference of the Cognitive Science Society.

Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & Cognition*, *19*(3), 274–282. https://doi.org/10.3758/BF03211151

Dasgupta, I., Schulz, E., & Gershman, S. J. (2017). Where do hypotheses come from? *Cognitive Psychology*, *96*, 1–25. https://doi.org/10.1016/j.cogpsych.2017.05.001

De Finetti, B. (1995). The Logic of Probability. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 77(1)jstor 4320559, 181–190. Retrieved October 19, 2022, from https://www.jstor.org/stable/4320559

de Cornulier, B. (1983). ''If'' and the Presumption of Exhaustivity. *Journal of Pragmatics*, *7*(3), 247–249.

Degen, J., Hawkins, R. D., Graf, C., Kreiss, E., & Goodman, N. D. (2020). When Redundancy Is Useful: A Bayesian Approach to "Overinformative" Referring Expressions. *Psychological Review*. https://doi.org/10.1037/rev0000186

Dieussaert, K., Schaeken, W., & d'Ydewalle, G. (2002). The relative contribution of content and context factors on the interpretation of conditionals. *Experimental Psychology*, *49*(3), 181–195. https://doi.org/10.1026/1618-3169.49.3.181

Díez, F. J. (1993, January 1). Parameter adjustment in Bayes networks. The generalized noisy OR–gate. In D. Heckerman & A. Mamdani (Eds.), *Uncertainty in Artificial Intelligence* (pp. 99–105). Morgan Kaufmann. https://doi.org/10.1016/B978-1-4832-1451-1.50016-0

Douven, I., & Romeijn, J.-W. (2011). A New Resolution of the Judy Benjamin Problem. *Mind*, *120*(479), 637–670. https://doi.org/10.1093/mind/fzr051

Douven, I. (2008). The evidential support theory of conditionals. *Synthese*, *164*(1), 19–44. https://doi.org/10.1007/s11229-007-9214-5

Douven, I. (2012). Learning conditional information. *Mind & Language*, *27*(3), 239–263.

Douven, I. (2017). How to account for the oddness of missing-link conditionals. *Synthese*, *194*, 1541–1554. https://doi.org/10.1007/s11229-015-0756-7

Douven, I., Elqayam, S., & Krzyzanowska, K. (2022). The Experimental Philosophy of Logic and Formal Epistemology. In *The Compact Compendium of Experimental Philosophy* (p. 23). De Gruyter.

Douven, I., Elqayam, S., Singmann, H., & van Wijnbergen-Huitink, J. (2018). Conditionals and inferential connections: A hypothetical inferential theory. *Cognitive Psychology*, *101*, 50–81. https://doi.org/10.1016/j.cogpsych.2017.09.002

Douven, I., Elqayam, S., Singmann, H., & van Wijnbergen-Huitink, J. (2020). Conditionals and inferential connections: Toward a new semantics. *Thinking & Reasoning*, *26*(3), 311–351. https://doi.org/10.1080/13546783.2019.1619623

Douven, I., & Verbrugge, S. (2010). The Adams family. *Cognition*, *117*(3), 302–318. https://doi.org/10.1016/j.cognition.2010.08.015

Douven, I., & Verbrugge, S. (2012). Indicatives, concessives, and evidential support. *Thinking & Reasoning*, *18*(4), 480–499. https://doi.org/10.1080/13546783.2012.716009

Ducrot, O. (1969). Présupposés et sous-entendus. *Langue française*. https://doi.org/10.3406/lfr.1969.5456

Edgington, D. (1995). *On Conditionals*. https://doi.org/10.1017/CBO9780511753466

Edgington, D. (2003). What if? Questions about conditionals. *Mind & Language*, *18*(4), 380–401.

Edgington, D. (2020). Indicative Conditionals. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2020). Metaphysics Research Lab, Stanford University. Retrieved March 7, 2022, from https://plato.stanford.edu/archives/fall2020/entries/conditionals/

Egré, P., & Rott, H. (2021). The Logic of Conditionals. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2021). Metaphysics Research Lab, Stanford University. Retrieved October 20, 2022, from https://plato.stanford.edu/archives/win2021/entries/logic-conditionals/

Elqayam, S., & Over, D. E. (2013). New paradigm psychology of reasoning: An introduction to the special issue edited by Elqayam, Bonnefon, and Over. *Thinking & Reasoning*, *19*(3-4), 249–265. https://doi.org/10.1080/13546783.2013.841591

Eva, B., Hartmann, S., & Rad, S. R. (2020). Learning From Conditionals. *Mind*, *129*(514), 461–508. https://doi.org/10.1093/mind/fzz025

Evans, J. S. B. T., Handley, S. J., Neilens, H., & Over, D. E. (2007). Thinking about conditionals: A study of individual differences. *Memory & Cognition*, *35*(7), 1772–1784.

Evans, J. S. B. T., Handley, S. J., & Over, D. E. (2003). Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(2), 321.

Evans, J. S. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction*. Psychology Press.

Evans, J. S. B. T., & Over, D. E. (2004). *If*. New York, NY, US, Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198525134.001.0001

Farr, M.-C. (2011). Focus influences the presence of conditional perfection: Experimental evidence. *Proceedings of Sinn & Bedeutung*, *15*, 225–239.

Fernbach, P. M., & Darlow, A. (2010). Neglect of alternative causes in predictive but not diagnostic reasoning. *journals.sagepub.com*, *21*(3), 329–336. https://doi.org/10.1177/0956797610361430

Fernbach, P. M., Darlow, A., & Sloman, S. A. (2011). Asymmetries in predictive and diagnostic reasoning. *Journal of Experimental Psychology: General*, *140*(2), 168–185. https://doi.org/10.1037/a0022100

Fernbach, P. M., & Rehder, B. (2013). Cognitive shortcuts in causal inference. *Argument and Computation*, *4*(1), 64–88. https://doi.org/10.1080/19462166.2012.682655

Fillenbaum, S. (1975). If: Some uses. *Psychological Research*, *37*(3), 245–260. https://doi.org/10.1007/BF00309037

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998–998.

Franke, M. (2009). *Signal to act: Game theory in pragmatics*. Institute for Logic, Language and Computation.

Franke, M. (2011). Quantity implicatures, exhaustive interpretation, and rational conversation. *Semantics and Pragmatics*, *4*, 1–82. https://doi.org/10.3765/sp.4.1

Franke, M. (2014). Pragmatic Reasoning about unawareness. *Erkenntnis*, *79*(S4), 729–767. https://doi.org/10.1007/s10670-013-9464-1

Franke, M., & Bergen, L. (2020). Theory-driven statistical modeling for semantics and pragmatics: A case study on grammatically generated implicature readings. *Language*, *96*(2), e77–e96. https://doi.org/10.1353/lan.2020.0034

Franke, M., & de Jager, T. (2011). Now that you mention it: Awareness Dynamics in Discourse and Decisions. In A. Benz, C. Ebert, G. Jäger, & R. van Rooij (Eds.), *Language, Games, and Evolution* (pp. 60–91).

Franke, M., & Jäger, G. (2016). Probabilistic pragmatics, or why Bayes' rule is probably important for pragmatics. *Zeitschrift für sprachwissenschaft*, *35*(1), 3–44.

Fugard, A. J., Pfeifer, N., & Mayerhofer, B. (2011). Probabilistic theories of reasoning need pragmatics too: Modulating relevance in uncertain conditionals. *Journal of Pragmatics*, *43*(7), 2034–2042.

Gates, M. A., Veuthey, T. L., Tessler, M. H., Smith, K. A., & Bayet, L. (2018). Tiptoeing around it: Inference from absence in potentially offensive speech. Proceedings of the 37th Annual Meeting of the Cognitive Science Society.

Geis, M. L., & Lycan, W. G. (1993). Nonconditional conditionals. *Philosophical topics*, *21*(2), 35–56.

Geis, M. L., & Zwicky, A. M. (1971). On invited inferences. *Linguistic inquiry*, *2*(4), 561–566.

Geurts, B. (2010). *Quantity Implicatures*. Cambridge, Cambridge University Press. https://doi.org/10.1017/CBO9780511975158

Gibbard, A. (1981). Two Recent Theories of Conditionals. In W. L. Harper, R. C. Stalnaker, & G. Pearce (Eds.), *IFS: Conditionals, Belief, Decision, Chance and Time* (pp. 211–247). Dordrecht, Springer Netherlands. https://doi.org/10.1007/978-94-009-9117-0_10

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in cognitive sciences*, *20*(11), 818–829.

Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in cognitive science*, *5*(1), 173–184.

Goodman, N. D., & Stuhlmüller, A. (2014). The design and implementation of probabilistic programming languages. http://dippl.org

Grice, H. (1975). Logic and conversation. *1975*, 41–58.

Grice, H. (1989). Indicative conditionals. In *Studies in the way of words* (pp. 58–85).

Grice, H. (1991). *Studies in the way of words* (First Harvard University Press paperback edition). Cambridge, Massachusetts, Harvard University Press.

Grice, H., & White, A. R. (1961). Symposium: The causal theory of perception. *Proceedings of the Aristotelian Society, Supplementary Volumes*, *35*, 121–168.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive psychology*, *51*(4), 334–384.

Griggs, R. A., & Cox, J. R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology*, *73*(3), 407–420. https://doi.org/10.1111/j.2044-8295.1982.tb01823.x

Grove, A. J., & Halpern, J. Y. (1997). Probability update: Conditioning vs. cross-entropy, In *Proceedings of the Thirteenth conference on Uncertainty in artificial intelligence*, Morgan Kaufmann Publishers Inc.

Grusdt, B., & Franke, M. (2021). Communicating uncertain beliefs with conditionals: Probabilistic modeling and experimental data, In *Proceedings of the Annual Meeting of the Cognitive Science Society*.

Grusdt, B., Lassiter, D., & Franke, M. (2022). Probabilistic modeling of rational communication with conditionals. *Semantics and Pragmatics*, *15*. https://doi.org/10.3765/sp.15.13

Grusdt, B., Liu, M., & Franke, M. (2022). Testing the Influence of QUDs on Conditional Perfection, In *Experiments in Linguistic Meaning 2*.

Günther, M. (2018). Learning Conditional Information by Jeffrey Imaging on Stalnaker Conditionals. *Journal of Philosophical Logic*, *47*(5), 851–876. https://doi.org/10.1007/s10992-017-9452-z

Hadjichristidis, C., & Stevenson, R. J. (2001). On the Evaluation of If p then q Conditionals, In *Proceedings of the 23rd annual meeting of the cognitive science society.*

Hagmayer, Y., & Waldmann, M. R. (2007). Inferences about unobserved causes in human contingency learning. *The Quarterly Journal of Experimental Psychology*, *60*(3), 330–355. https://doi.org/10.1080/17470210601002470

Hájek, A. (1989). Probabilities of Conditionals: Revisited. *Journal of Philosophical Logic*, *18*(4)jstor 30226421, 423–428. Retrieved September 26, 2022, from https://www.jstor.org/stable/30226421

Heifetz, A., Meier, M., & Schipper, B. C. (2006). Interactive unawareness. *Journal of Economic Theory*, *130*(1), 78–94. https://doi.org/10.1016/j.jet.2005.02.007

Heim, I. R. (1988). *The semantics of definite and indefinite noun phrases*. University of Massachusetts Amherst. http://semanticsarchive.net/Archive/TkoZmYyY/dissertation.pdf

Herbstritt, M., & Franke, M. (2019). Complex probability expressions & higher-order uncertainty: Compositional semantics, probabilistic pragmatics & experimental data. *Cognition*, *186*, 50–71. https://doi.org/10.1016/j.cognition.2018.11.013

Herburger, E. (2016). Conditional perfection: The truth and the whole truth, In *Semantics and Linguistic Theory*.

Hiddleston, E. (2005). A Causal Theory of Counterfactuals. *Noûs*, *39*(4), 632–657. https://doi.org/10.1111/j.0029-4624.2005.00542.x

Horn, L. R. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. *Meaning, form, and use in context: Linguistic applications*, *11*, 42.

Horn, L. R. (1989). *A Natural History of Negation*. University of Chicago Press.

Horn, L. R. (2000). From if to iff: Conditional perfection as pragmatic strengthening. *Journal of pragmatics*, *32*, 289–326.

Horn, L. R. (2006). Implicature. In *The Handbook of Pragmatics* (pp. 2–28). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470756959.ch1

Hyttinen, A., Eberhardt, F., & Hoyer, P. O. (2011). Noisy-OR Models with Latent Confounding, In *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence*.

Icard, T. F., & Goodman, N. D. (2015). A Resource-Rational Approach to the Causal Frame Problem, In *Proceedings of the Annual Meeting of the Cognitive Science Society*.

Jackson, F. (1979). On Assertion and Indicative Conditionals. *The Philosophical Review*, *88*(4), 565–589. Retrieved September 14, 2022, from https://www.pdcnet.org/pdc/bvdb.nsf/purchase?openform&fp=phr&id=phr_1979_0088_0004_0565_0589

Jeffrey, R. C. (1983). *The logic of decision* (2nd ed.). Chicago, Ill. [u.a.], University of Chicago Press.

Johnson-Laird, P. N. (1986). *Mental models: Towards a cognitive science of language, inference, and consciousness*. USA, Harvard University Press.

Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in cognitive sciences*, *5*(10), 434–442.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ, US, Lawrence Erlbaum Associates, Inc.

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

Johnson-Laird, P. N., Khemlani, S. S., & Goodwin, G. P. (2015). Logic, probability, and human reasoning. *Trends in Cognitive Sciences*, *19*(4), 201–214. https://doi.org/10.1016/j.tics.2015.02.006

Johnson-Laird, P. N., Legrenzi, P., & Legrenzi Sonino, M. (1972). Reasoning and a sense of reality. *British Journal of Psychology*, *63*(3), 395–400. https://doi.org/10.1111/j.2044-8295.1972.tb01287.x

Kao, J. T., Bergen, L., & Goodman, N. D. (2014). Formalizing the pragmatics of metaphor understanding, In *Proceedings of the annual meeting of the Cognitive Science Society*.

Kao, J. T., Wu, J. Y., Bergen, L., & Goodman, N. D. (2014). Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences*, *111*(33), 12002–12007. https://doi.org/10.1073/pnas.1407479111

Katzir, R. (2007). Structurally-defined alternatives. *Linguist and Philos*, *30*, 669–690. https://doi.org/10.1007/s10988-008-9029-y

Kaufmann, S. (2004). Conditioning against the Grain. *Journal of Philosophical Logic*, *33*(6), 583–606. https://doi.org/10.1023/B:LOGI.0000046142.51136.bf

Kaufmann, S. (2013). Causal Premise Semantics. *Cognitive Science*, *37*(6), 1136–1170. https://doi.org/10.1111/cogs.12063

Kaufmann, S. (2023). Bernoulli Semantics and Ordinal Semantics for Conditionals. *Journal of Philosophical Logic*, *52*(1), 199–220. https://doi.org/10.1007/s10992-022-09670-8

Khoo, J. (2016). Probabilities of conditionals in context. *Linguistics and Philosophy*, *39*(1), 1–43. https://doi.org/10.1007/s10988-015-9182-z

Kratzer, A. (1986). Conditionals. *Chicago Linguistics Society*, *22*(2), 1–15.

Kratzer, A. (2008, July 14). Modality. In A. Stechow & D. Wunderlich (Eds.), *Semantik / Semantics: Ein internationales Handbuch zeitgenössischer Forschung* (pp. 639–650). De Gruyter Mouton. https://doi.org/10.1515/9783110126969.7.639

Krynski, T. R., & Tenenbaum, J. B. (2007). The Role of Causality in Judgment Under Uncertainty. *Journal of Experimental Psychology: General*, *136*(3). https://doi.org/10.1037/0096-3445.136.3.430

Krzyżanowska, K. (2019). What is wrong with false-link conditionals? *Linguistics Vanguard*, *5*(s3), 20190006. https://doi.org/10.1515/lingvan-2019-0006

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

Krzyżanowska, K., Collins, P. J., & Hahn, U. (2021). True clauses and false connections. *Journal of Memory and Language*, *121*, 104252. https://doi.org/10.1016/j.jml.2021.104252

Krzyżanowska, K., & Douven, I. (2018). Missing-link conditionals: Pragmatically infelicitous or semantically defective? *Intercultural Pragmatics*, *15*(2), 191–211. https://doi.org/10.1515/ip-2018-0004

Krzyżanowska, K., Wenmackers, S., & Douven, I. (2013). Inferential Conditionals and Evidentiality. *Journal of Logic, Language and Information*, *22*(3). https://doi.org/10.1007/s10849-013-9178-4

Krzyżanowska, K., Wenmackers, S., & Douven, I. (2014). Rethinking Gibbard's Riverboat Argument. *Springer*, *102*(4), 771–792. https://doi.org/10.1007/s11225-013-9507-2

Krzyżanowska, K., Wenmackers, S., Douven, I., & Verbrugge, S. (2013). Conditionals, Inference, and Evidentiality. *Journal of Logic, Language and Information*, *22*(3), 315–334.

Kullback, S., & Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, *22*(1)jstor 2236703, 79–86. https://doi.org/10.1214/aoms/1177729694

Lassiter, D. (2017). Probabilistic language in indicative and counterfactual conditionals, In *Proceedings of SALT 27*. Semantics and Linguistic Theory. https://doi.org/10.3765/salt.v27i0.4188

Lassiter, D. (2018a). Complex sentential operators refute unrestricted Simplification of Disjunctive Antecedents. *Semantics and Pragmatics*, *11*. https://doi.org/10.3765/sp.11.9

Lassiter, D. (2018b). Talking about (quasi-)higher-order uncertainty. *Tokens of Meaning: Papers in Honor of Lauri Karttunen*, 133–158. Retrieved October 26, 2022, from https://www.research.ed.ac.uk/en/publications/talking-about-quasi-higher-order-uncertainty

Lassiter, D. (2020). What we can learn from how trivalent conditionals avoid triviality. *Inquiry*, *63*(9-10), 1087–1114. https://doi.org/10.1080/0020174X.2019.1698457

Lassiter, D. (2022). Decomposing relevance in conditionals. *Mind & Language*, 26. https://doi.org/10.1111/mila.12418

Lassiter, D., & Goodman, N. D. (2017). Adjectival vagueness in a Bayesian model of interpretation. *Synthese*, *194*(10), 3801–3836.

Lee, M. D., & Wagenmakers, E.-J. (2014, April 3). *Bayesian Cognitive Modeling: A Practical Course* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139087759

Levinson, S. C. (1987). Minimization and conversational inference. In *The pragmatic perspective* (pp. 61–129).

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*.

Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00731

Lewis, C. I. (1912). Implication and the algebra of logic. *Mind*, *21*(84), 522–531. https://doi.org/10.1093/mind/XXI.84.522

Lewis, D. (1975). Adverbs of Quantification (E. Keenan, Ed.). *Formal Semantics: The Essential Readings*, 178–188. https://doi.org/10.1002/9780470758335.ch7

Lewis, D. (1976). Probabilities of conditionals and conditional probabilities. *The Philosophical Review*, *85*(3), 297–315. https://doi.org/10.2307/2184045

Lewis, D. (1986). Probabilities of Conditionals and Conditional Probabilities II. *The Philosophical Review*, *95*(4), 581–589. Retrieved September 26, 2022, from https://www.pdcnet.org/pdc/bvdb.nsf/purchase?openform&fp=phr&id=phr_1986_0095_0004_0581_0589

Lilje, G. W. (1972). Uninvited inferences. *Linguistic Inquiry*, 540–542.

Lin, H. (2022). Bayesian Epistemology. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Fall 2022). Metaphysics Research Lab, Stanford University. Retrieved October 12, 2022, from https://plato.stanford.edu/archives/fall2022/entries/epistemology-bayesian/

López Astorga, M. (2014). Podemos evitar la perfección del condicional enfocando el antecedente o son necesarios antecedentes alternativos? *Revista signos*, *47*(85), 267–292.

Lucas, C. G., & Kemp, C. (2015). An improved probabilistic account of counterfactual reasoning. *Psychological Review*, *122*(4), 700–734. https://doi.org/10.1037/a0039655

MacBride, F., Marion, M., Frápolli, M. J., Edgington, D., Elliott, E., Lutz, S., & Paris, J. (2020). Frank ramsey. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2020). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2020/entries/ramsey/

Manktelow, K. I., & Evans, J. S. B. T. (1979). Facilitation of reasoning by realism: Effect or non-effect? *British Journal of Psychology*, *70*(4), 477–488. https://doi.org/10.1111/j.2044-8295.1979.tb01720.x

Markovits, H. (1986). Familiarity effects in conditional reasoning. *Journal of Educational Psychology*, *78*(6), 492.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, Freeman.

Matsumoto, Y. (1995). The conversational condition on horn scales. *Linguistics and Philosophy*, *18*(1), 21–60. https://doi.org/10.1007/BF00984960

Mellor, D. H. (1993). How to Believe a Conditional. *The Journal of Philosophy*, *90*(5)jstor 2940911, 233–248. https://doi.org/10.2307/2940911

Milne, P. (1997). Bruno de Finetti and the Logic of Conditional Events. *The British Journal for the Philosophy of Science*, *48*(2)jstor 687745, 195–232. Retrieved October 26, 2022, from https://www.jstor.org/stable/687745

Moldovan, A. (2013). Denying the antecedent and conditional perfection again. *OSSA Conference Archive*. https://scholar.uwindsor.ca/ossaarchive/OSSA10/papersandcommentaries/117

Moss, S. (2015). On the semantics and pragmatics of epistemic vocabulary. *Semantics and Pragmatics*, *8*(5), 1–81. https://doi.org/10.3765/sp.8.5

Newstead, S. E. (1997). Conditional Reasoning with Realistic Material. *Thinking & Reasoning*, *3*(1), 49–76. https://doi.org/10.1080/135467897394428

Oaksford, M., & Chater, N. (2020). New Paradigms in the Psychology of Reasoning. *Annual Review of Psychology*, *71*(1), 305–330. https://doi.org/10.1146/annurev-psych-010419-051132

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.

Oberauer, K., & Wilhelm, O. (2003). The meaning(s) of conditionals: Conditional probabilities, mental models, and personal utilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(4), 680–693. https://doi.org/10.1037/0278-7393.29.4.680

Over, D. E. (2009). New paradigm psychology of reasoning. *Thinking & Reasoning*, *15*(4), 431–438. https://doi.org/10.1080/13546780903266188

Over, D. E., & Cruz, N. (2021, December 14). The suppositional theory of conditionals and rationality. In *The handbook of rationality* (p. 11).

Over, D. E., Hadjichristidis, C., Evans, J. S. B., Handley, S. J., & Sloman, S. A. (2007). The probability of causal conditionals. *Cognitive Psychology*, *54*(1), 62–97. https://doi.org/10.1016/j.cogpsych.2006.05.002

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA, USA, Morgan Kaufmann Publishers Inc.

Pearl, J. (2009). *Causality* (2nd ed.). Cambridge, UK, Cambridge University Press. https://doi.org/10.1017/CBO9780511803161

Pearl, J. (2013). Structural counterfactuals: A brief introduction. *Cognitive Science*, *37*(6), 977–985. https://doi.org/10.1111/cogs.12065

Pearl, J., & Mackenzie, D. (2018). *The Book of Why*. Basic Books.

Potts, C. (2005). *The Logic of Conventional Implicatures*. Oxford University Press.

Qing, C., & Franke, M. (2015). Variations on a bayesian theme: Comparing bayesian models of referential reasoning. In *Bayesian natural language semantics and pragmatics* (pp. 201–220). Springer.

Qing, C., Goodman, N. D., & Lassiter, D. (2016). A rational speech-act model of projective content, In *Proceedings of the thirty-eighth annual conference of the Cognitive Science Society*.

Ramsey, F. P. (1931). General propositions and causality. In R. B. Braithwaite (Ed.), *The foundations of Mathematics and other Logical Essays*. http://www.dspace.cam.ac.uk/handle/1810/194722

Rips, L. J. (2010). Two causal theories of counterfactual conditionals. *Cognitive Science*, *34*(2), 175–221. https://doi.org/10.1111/j.1551-6709.2009.01080.x

Roberts, C. (2012). Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, *5*. https://doi.org/10.3765/sp.5.6

Rostworowski, W., Pietrulewicz, N., & Bedkowski, M. (2021). Conditionals and specific links—an experimental study. *Synthese*, *199*(3-4), 7365–7399. https://doi.org/10.1007/s11229-021-03119-2

Rothschild, D. (2015). Conditionals and Propositions in Semantics. *Journal of Philosophical Logic*, *44*(6), 781–791. https://doi.org/10.1007/s10992-015-9359-5

Sadock, J. (2006). Speech Acts. In *The Handbook of Pragmatics* (pp. 53–73). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470756959.ch3

Santorio, P. (2019). Interventions in Premise Semantics. *Philosophers' Imprint*, *19*(1), 1–27. http://hdl.handle.net/2027/spo.3521354.0019.001

Schulz, K. (2015). Conditionals from a Linguistic Point of View: Two Case Studies. *Journal of Philosophical Logic*, *44*(6), 805–816. https://doi.org/10.1007/s10992-015-9361-y

Schuster, S., & Degen, J. (2019). Speaker-specific adaptation to variable use of uncertainty expressions. CogSci. https://osf.io/qnspg

Scontras, G., Tessler, M. H., & Franke, M. (2018). *Probabilistic language understanding: An introduction to the Rational Speech Act framework*. https://www.problang.org

Sheps, M. C. (1958). Shall We Count the Living or the Dead? *New England Journal of Medicine*, *259*(25), 1210–1214. https://doi.org/10.1056/nejm195812182592505

Simons, M. (2001). Disjunction and Alternativeness. *Linguistics and Philosophy*, *24*(5)jstor 25001826, 597–619. Retrieved October 4, 2022, from https://www.jstor.org/stable/25001826

Singmann, H., Klauer, K. C., & Over, D. (2014). New normative standards of conditional reasoning and the dual-source model. *Frontiers in Psychology*, *5*. Retrieved February 4, 2022, from https://www.frontiersin.org/article/10.3389/fpsyg.2014.00316

Skovgaard-Olsen, N. (2016). Motivating the Relevance Approach to Conditionals. *Mind and Language*. https://doi.org/10.1111/mila.12120

Skovgaard-Olsen, N., Collins, P. J., Krzyżanowska, K., Hahn, U., & Klauer, K. C. (2019). Cancellation, negation, and rejection. *Cognitive Psychology*, *108*, 42–71. https://doi.org/10.1016/j.cogpsych.2018.11.002

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

Skovgaard-Olsen, N., Singmann, H., & Klauer, K. C. (2017). Relevance and Reason Relations. *Cognitive Science*, *41*, 1202–1215. https://doi.org/10.1111/cogs.12462

Stalnaker, R. C. (1968). A theory of conditionals. In *Ifs* (pp. 41–55). Springer.

Stalnaker, R. C. (1970). Probability and conditionals. *Philosophy of science*, *37*(1), 64–80.

Stalnaker, R. C. (1976). Indicative conditionals. In *Ifs* (pp. 193–210). Springer.

Stalnaker, R. C., & Jeffrey, R. (1994). Conditionals as Random Variables. In E. Eells, B. Skyrms, & E. W. Adams (Eds.), *Probability and Conditionals: Belief Revision and Rational Decision* (p. 31). Cambridge University Press.

Starr, W. (2021). Counterfactuals. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2021). Metaphysics Research Lab, Stanford University. Retrieved October 20, 2022, from https://plato.stanford.edu/archives/sum2021/entries/counterfactuals/

Swanson, E. (2016). The Application of Constraint Semantics to the Language of Subjective Uncertainty. *Journal of Philosophical Logic*, *45*(2), 121–146. https://doi.org/10.1007/s10992-015-9367-5

Tenenbaum, J. B., & Griffiths, T. L. (2003). Theory-based causal inference. *Advances in neural information processing systems*, 43–50.

Trpin, B. (2020). Jeffrey conditionalization: Proceed with caution. *Philosophical Studies*, *177*(10), 2985–3012. https://doi.org/10.1007/s11098-019-01356-3

Ülkümen, G., Fox, C. R., & Malle, B. F. (2016). Two dimensions of subjective uncertainty: Clues from natural language. *Journal of Experimental Psychology. General*, *145*(10), 1280–1297. https://doi.org/10.1037/xge0000202

Van Canegem-Ardijns, I., & Van Belle, W. (2008). Conditionals and types of conditional perfection. *Journal of Pragmatics*, *40*(2), 349–376. https://doi.org/10.1016/j.pragma.2006.11.007

van der Auwera, J. (1997). Pragmatics in the last quarter century: The case of conditional perfection. *Journal of Pragmatics*, *27*(3), 261–274.

Vandenburgh, J. (2021). Conditional learning through causal models. *Synthese*, *199*(1), 2415–2437. https://doi.org/10.1007/s11229-020-02891-x

van Fraassen, B. C. (1976). Probabilities of Conditionals. In W. L. Harper & C. A. Hooker (Eds.), *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science: Volume I Foundations and Philosophy of Epistemic Applications of Probability Theory* (pp. 261–308). Dordrecht, Springer Netherlands. https://doi.org/10.1007/978-94-010-1853-1_10 van Fraassen, B. C. (1981). A Problem for Relative Information Minimizers in Probability Kinematics. *The British Journal for the Philosophy of Science*, *32*(4)jstor 687309, 375–379. Retrieved October 12, 2022, from https://www.jstor.org/stable/687309

van Rooij, R., & Schulz, K. (2019). Conditionals, Causality and Conditional Probability. *Journal of Logic, Language and Information*, *28*(1), 55–71. https://doi.org/10.1007/s10849-018-9275-5

van Rooij, R., & Schulz, K. (2020). Conditionals As Representative Inferences. *Axiomathes*, *0123456789*. https://doi.org/10.1007/s10516-020-09477-9

van Rooij, R., & Schulz, K. (2021). Why Those Biscuits Are Relevant and on the Sideboard. *Theoria*, *87*(3), 704–712. https://doi.org/10.1111/theo.12309

van Rooij, R., & Schulz, K. (2022). Causal relevance of conditionals: Semantics or pragmatics? *Linguistics Vanguard*, *8*(s4), 363–370. https://doi.org/10.1515/lingvan-2021-0030

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of MCMC. *Bayesian Analysis*, *16*(2)arxiv 1903.08008. https://doi.org/10.1214/20-BA1221

Vidal, M., & Baratgin, J. (2017). A psychological study of unconnected conditionals. *Journal of Cognitive Psychology*, *29*(6), 769–781. https://doi.org/10.1080/20445911.2017.1305388

von Fintel, K. (2001). *Conditional strengthening - A Case Study in Implicature*.

Wason, P. C., & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *The Quarterly Journal of Experimental Psychology*, *23*(1), 63–71. https://doi.org/10.1080/00335557143000068

Wason, P. C. (1966). Reasoning. In B. Foss (Ed.), *New Horizons in Psychology* (pp. 135–151). Harmondsworth: Penguin Books.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly journal of experimental psychology*, *20*(3), 273–281.

Yalcin, S. (2012). Bayesian Expressivism. *Proceedings of the Aristotelian Society*, *112*, 123–160. https://doi.org/10.1111/j.1467-9264.2012.00329.x

Yoon, E. J., Tessler, M. H., Goodman, N. D., & Frank, M. C. (2016). Talking with tact: Polite language as a balance between kindness and informativity, In *Proceedings of the 38th annual conference of the cognitive science society*, Cognitive Science Society.

Zaefferer, D. (1991). Conditionals and Unconditionals: Cross-linguistic and Logical Aspects. In *Semantic Universals and Universal Semantics* (pp. 210–236). De Gruyter Mouton. https://doi.org/10.1515/9783110870527-011

Zednik, C., & Jäkel, F. (2016). Bayesian reverse-engineering considered as a research strategy for cognitive science. *Synthese*, *193*(12), 3951–3985. https://doi.org/10.1007/s11229-016-1180-3

ERKLÄRUNG ÜBER DIE EIGENSTÄNDIGKEIT DER ERBRACHTEN WISSENSCHAFTLICHEN LEISTUNG

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.

Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise unentgeltlich geholfen.

1. Kapitel 3–5 entsprechen in etwas abgeänderter Form der Veröffentlichung *Probabilistic modeling of rational communication with conditionals*, erschienen im Journal *Semantics & Pragmatics*, Vol.15, 2022, welche von Michael Franke betreut wurde. Der letzte Teil von Abschnitt 5.4.3 über *concessive conditionals* war nicht Teil dieser Veröffentlichung.

   Mitwirkung der Autoren der Veröffentlichung Grusdt, Lassiter, et al. (2022) "Probabilistic Modeling of Rational Communication with Conditionals":

   - Britta Grusdt: Konzeptualisierung, Datenerhebung und Analyse, Schreiben.
   - Daniel Lassiter: Konzeptualisierung, Schreiben.
   - Michael Franke: Konzeptualisierung, Schreiben.

   Konzeptualisierung: Die grundlegende Idee für die Veröffentlichung stammt von Michael Franke sowie Daniel Lassiter. In regelmäßigen Treffen, zunächst mit Michael Franke und mir, später zu dritt mit Daniel Lassiter, haben wir die Arbeit gemeinsam konzipiert und diskutiert.

   Datenerhebung und Analyse: Alle Implementationen, Simulationen und Analysen wurden von mir durchgeführt.

   Schreiben: Die erste Version des Papers habe ich verfasst. Iterativ haben Michael Franke und ich das Manuskript bearbeitet, wobei Michael Franke meinen Text überarbeitet, ergänzt und kommentiert hat. Daniel Lassiter hat das Manuskript kommentiert und teilweise überarbeitet (section 1 und section 3 des Papers bzgl. conditional perfection). Außerdem hat Daniel Lassiter Teile von Kapitel 2.2 des Papers (objective vs. subjective chances) verfasst und überarbeitet.

2. Kapitel 6 basiert auf der Veröffentlichung *Communicating uncertain beliefs with conditionals*, erschienen in den Proceedings der jährlichen Konferenz der Cognitive Science Society, Vol. 43, 2021. Diese Arbeit wurde von Michael Franke betreut.

   Mitwirkung der Autoren der Veröffentlichung Grusdt and Franke (2021) "Communicating Uncertain Beliefs with Conditionals: Probabilistic Modeling and Experimental Data":

   - Britta Grusdt: Konzeptualisierung, Datenerhebung und Analyse, Schreiben.
   - Michael Franke: Konzeptualisierung, Schreiben.

   Konzeptualisierung: Die grundlegende Idee zur Art des Experiments (Stimuli, basierend auf intuitiver Physik) stammt von Michael Franke. In regelmäßigen Treffen haben Michael Franke und ich das Experiment gemeinsam ausgearbeitet und diskutiert.

   Datenerhebung und Analyse: Mit Ausnahme des Dirichlet-Modells, das Michael Franke gefitted hat (nur Teil des Papers, nicht in dieser Thesis enthalten), wurden alle Implementationen und Analysen von mir (mit Unterstützung der studentischen Hilfskraft Malin Spaniol) durchgeführt.

   Schreiben: Die erste Version des Papers habe ich verfasst, welche von Michael Franke überarbeitet und ergänzt wurde.

3. Kapitel 8 entspricht in etwas abgeänderter Form der Veröffentlichung *Testing the influence of QUDs on Conditional Perfection*, erschienen in den Proceedings der Konferenz *Experiments in Linguistic Meaning* 2,2021. Diese Arbeit wurde sowohl von Michael Franke als auch Mingya Liu betreut.

   Mitwirkung der Autoren der Veröffentlichung Grusdt, Liu, et al. (2022) "Testing the Influence of QUDs on Conditional Perfection":

   - Britta Grusdt: Konzeptualisierung, Datenerhebung und Analyse, Schreiben.
   - Mingya Liu: Konzeptualisierung, Schreiben.
   - Michael Franke: Konzeptualisierung, Schreiben.

   Konzeptualisierung: Die grundlegende Idee für das Experiment stammt von mir. In regelmäßigen Treffen mit Michael Franke und Mingya Liu haben wir das Experiment gemeinsam ausgearbeitet und diskutiert.

   Datenerhebung und Analyse: Alle Implementationen und Analysen wurden von mir (mit Unterstützung der studentischen Hilfskraft Josefine Zerbe) durchgeführt.

Schreiben: Die erste Version des Papers habe ich verfasst. Michael Franke und Mingya Liu haben Feedback zum Manuskript gegeben, welches von mir eingearbeitet wurde.

4. Alle restlichen Teile der Dissertation wurden eigenständig von mir verfasst, und inhaltlich mit meinem Betreuer bzw. meiner Betreuerin besprochen, deren Feedback zum Text von mir eingearbeitet wurde. Außerdem haben Vinicius Macuch, Tabea Kossen, Malin Spaniol, Xenia Ohmer und Pablo Prietz Feedback zur Verständlichkeit des Textes gegeben, welches von mir eingearbeitet wurde.

Weitere Personen waren an der inhaltlichen materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater oder andere 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.

....	....
(Ort, Datum)	(Unterschrift)

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

https://bitbucket.org/amiede/classicthesis/

Happy users of `classicthesis` usually send a real postcard to the author, a collection of postcards received so far is featured here:

http://postcards.miede.de/

Final Version as of October 18, 2023 (`classicthesis` version 1.0).