Possible Worlds Truth Table Task

Niels Skovgaard-Olsen

University of Freiburg

Peter Collins

University of Greenwich

Karl Christoph Klauer

University of Freiburg

Author Note

Niels Skovgaard-Olsen, Institute of Psychology, University of Freiburg, Germany.
Peter Collins, School of Human Sciences, University of Greenwich, U.K. Christoph Klauer,
Institute of Psychology, University of Freiburg.

Correspondence concerning this article should be addressed to Niels Skovgaard-Olsen
(niels.skovgaard.olsen@psychologie.uni-freiburg.de, n.s.olsen@gmail.com).

Abstract

In this paper, a novel experimental task is developed for testing the highly influential, but experimentally underexplored, possible worlds account of conditionals (Stalnaker, 1968; Lewis, 1973). In Experiment 1, this new task is used to test both indicative and subjunctive conditionals. For indicative conditionals, five competing truth tables are compared, including the previously untested, multi-dimensional possible worlds semantics of Bradley (2012). In Experiment 2, these results are replicated and it is shown that they cannot be accounted for by an alternative hypothesis proposed by our reviewers. In Experiment 3, individual variation in truth assignments of indicative conditionals is investigated via Bayesian mixture models that classify participants as following one of several competing truth tables. As a novelty of this study, it is found that a possible worlds semantics of Lewis and Stalnaker is capable of accounting for participants' aggregate truth value assignments in this task. Applied to indicative conditionals, we show across three experiments, that the theory both captures participants' truth values at the aggregate level (Experiments 1 and 2) and that it makes up the largest subgroup in the analysis of individual variation in our experimental paradigm (Experiment 3).

Keywords: Subjunctive Conditionals, Indicative Conditionals, Possible Worlds Semantics.

Introduction

Suppose that you overhear someone uttering one of the following two sentences:

(1) If Trump really won the 2020 election, then the official result was falsified.

(2) If Trump had really won the 2020 election, then the official result would have been falsified.

What would you conclude about the speaker's views? The speaker of (1) might be a conspiracist who believes that the official 2020 election result was a fraud. Or: they might be someone simply making a supposition for the sake of the argument, who has no inclination to doubt the official results, but who merely points out that their rejection would be a consequence of accepting the supposition. The linguistic form of (1) does not settle the issue, as it is in itself open to either interpretation. In contrast, it is difficult to imagine that the speaker of (2) is anything other than a conspiracist, albeit one of a peculiar kind. This conspiracist appears to take the official result to have been an inevitability: they accept after all the investigations that Trump's loss was genuine but hold that, in the event that Trump had in fact won, some process (e.g., rigged voting machines) would have prevented Trump from taking office for a second term. Despite the parallels between the sentences, then, they indicate markedly different beliefs.

Sentences (1) and (2) illustrate a classic distinction between two types of conditional: sentence (1) is an *indicative conditional*; sentence (2) is a *subjunctive conditional*. While indicatives and subjunctives clearly convey different meanings, there is no consensus theory of either indicatives or subjunctives or of the relationship between the two. These theoretical challenges are the subject of much theoretical and empirical work spanning linguistics, philosophy of language, and the psychology of reasoning. They are challenges with broad implications. Indicative conditionals have been linked to our ability to think hypothetically: to suppose that an event obtains (or obtained) and explore its possible consequences (e.g., Evans

et al., 2007). Hypothetical thinking, indeed, is the process we followed in the first paragraph, when we supposed that Trump did win the election and intuitively evaluated the likelihood of electoral fraud in that case. Subjunctive conditionals have been linked to our ability to think counterfactually: to suppose, contrary to known facts, that some event had happened and consider its consequences – how the world might have been different (Byrne, 2016). Counterfactual thinking is, in turn, linked to our abilities to explain the past, reason causally, formulate plans, and make moral judgements. Moreover, counterfactual thoughts generate positive emotions such as relief and negative emotions such as regret (Byrne, 2016). We gain insight into a wide range of cognitive abilities, then, by studying these types of conditionals.

The standard approach is to model the meaning of conditionals in terms of truth conditions: states of affairs that make the conditional true or false. This approach is by no means universal, with some arguing that only subjunctive conditionals have truth conditions (Bennett, 2003), and others that neither indicatives nor subjunctives have truth conditions (Edgington, 1995, 2008). But the approach is adopted by several key theories of conditionals, which argue for the truth conditions summarized in Table 1 for indicative conditionals (reviewed, e.g., in Evans & Over, 2004; Edgington, 2006; Over & Baratgin, 2017). For a conditional "If A, C", the "A" clause is the antecedent, and the "C" clause is the consequent. While the truth tables all agree in their first two rows, in which the antecedent is true, they differ markedly in the second two rows, in which the antecedent is false.

Table 1. Truth Tables for Indicative Conditionals

A	C	Material Implication	Possible Worlds	de Finetti	Jeffrey Table
T	T	T	T	T	T
T	F	F	F	F	F
F	T	T	T/F	NA	P(C|A)
F	F	T	T/F	NA	P(C|A)

Note. The material implication makes the truth value of the conditional a function of the truth values of the antecedent and the consequent. In contrast, the possible worlds account of Stalnaker (1968) permits the conditional to be either true or false, depending on whether the consequent is true in the most similar situation in which the antecedent is true. This feature of possible worlds conditionals means that their truth values are not truth-functional in contrast to the material implication.

Historically, each truth-table has had its supporters. The first of these truth tables, that of the material conditional, has lost favor in recent years and is rarely applied to counterfactuals (though see Williamson, 2020 for a recent defense). In psychology, this loss of favour is in large part due to a commonly reported conflict with experimental data. In philosophy, the problems with the material implication have long been well-known (C. I. Lewis, 1912; Bennett, 2003). In the truth table for the material conditional, a conditional is true whenever its antecedent is false. But experimental studies suggested that participants disagree with these truth values: that, when participants are asked if false-antecedent rows make the conditional true, false or are irrelevant to its truth or falsity, a substantial number of participants declare them irrelevant (e.g., Evans & Newstead, 1977; Evans et al., 2007; Johnson-Laird & Tagart, 1969).

This finding of frequent "irrelevant" responses lends support to the de Finetti truth table, on which account a conditional is neither true nor false when the antecedent is false. In such cases, the conditional can be viewed as "void" (for discussion, see Over & Cruz, 2018). That finding has, however, in turn been qualified through a recent meta-analysis (Schroyens, 2010), which reports that "irrelevant" is a minority response and only the modal response for the final row (the FF cases) - and then only with tasks which use implicit negation.

While the de Finetti Table takes the conditional to be void when its antecedent is false, the related Jeffrey Table replaces the "void" (above, "NA") values with the conditional probability of the consequent given the antecedent, P(C|A). This amendment is interpreted as capturing the fact that one can be more or less confident in a conditional even when its antecedent is false (Over & Cruz, 2018). In our experiments below, we will test both the de Finetti and Jeffrey truth tables and return to the issues they raise in the context of our experimental task.

Last, but not least, is the possible worlds account. Developed primarily for counterfactual conditionals (Lewis, 1973; Stalnaker, 1968), the account has been applied to both indicative and subjunctive conditionals (Stalnaker, 1968). We will provide a detailed summary of the possible worlds account below. Here it suffices to note that, on the account of Stalnaker (1968), when the antecedent is false, a conditional can be either true or false, depending on whether the consequent is true in the most similar situation in which the antecedent is true. In the psychology of reasoning, possible worlds semantics is sometimes discussed as an important alternative (e.g., Evans & Over, 2004), and sometimes set aside as psychologically implausible (Johnson-Laird & Ragni, 2019). But the truth conditions of possible worlds semantics are rarely tested.[1] Despite its roughly 50 years of existence, the theory has proved difficult to test empirically, which we hope to rectify in this paper.

It is a matter of much debate, then, which truth table best captures the meaning of conditionals. A related debate is whether indicative and subjunctive conditionals can be explained with a single theory, with theorists both advancing unified theories (e.g., Edgington, 2008; Stalnaker, 1968, 1975; Starr, 2014; Williamson, 2020) and opposing them (e.g., Bennett, 2003; Lewis, 1973, 1976). Within the psychology of reasoning, different camps have developed unified theories, including mental models theorists (Johnson-Laird & Byrne, 2002; Quelhas et al., 2018) and suppositional theorists (Over et al. 2007; Baratgin, Over et al., 2013; Pfeifer & Tulkki, 2017). In Over et al. (2007), this takes the form of interpreting

---

[1]   For some of the very few attempts to test other aspects of possible worlds semantics see: Rips & Marcus (1977), Cariani & Rips (2017), and Johnson-Laird & Ragni (2019). A further exception is Wijnbergen-Huitink, Elqayam et al. (2015), which applied Lewis/Stalnaker's theory to right-nested conditionals 'if p, then if q, r' in *indicative mood* in a betting context. However, Wijnbergen-Huitink et al. do not discuss the following two complications: 1) that Lewis (1973, 1976) only applied this theory to *subjunctive* conditionals, and 2) that Stalnaker (1975) only applies the theory to indicative conditionals under the assumption of the pragmatic principle explained below. A final exception is Douven et al. (2020), who apply one of Lewis's (1979) criteria for similarity (i.e. agreement with particular facts) to Stalnaker's theory of indicative conditionals.

counterfactuals as expressing an "epistemic past tense" (Adams, 1975), which is to be evaluated via conditional probabilities by considering a situation before the antecedent came to be disbelieved. Accordingly, the de Finetti and Jeffrey truth tables that have been applied to indicative conditionals (e.g., in Evans & Over, 2004; Baratgin, Over, & Politzer, 2013; Over & Baratgin, 2017) would also hold for counterfactual conditionals. Other related probabilistic approaches apply principles of causal reasoning to counterfactuals and emphasize that counterfactuals concern interventionalist probabilities when "rerunning history" in a causal model (Sloman & Lagnado, 2005; Rips, 2010; Pearl, 2013; Lassiter, 2017; Skovgaard-Olsen et al., 2021).

In deciding whether a unifying account of the two types of conditionals can be given, one question that deserves central attention is this: do indicative and subjunctive conditionals have the same type of truth conditions? In this paper, we present experiments which throw new light on this question by presenting a novel experimental test of the truth conditions of one of the key theories of conditionals: the possible worlds account (Lewis, 1973; Stalnaker, 1968). The experiments allow us to distinguish the possible worlds account from other key approaches, including the truth tables reviewed above.

Possible Worlds Semantics

Since this paper focuses in large part on testing the possible worlds account, we provide a more detailed overview of the theory here. Possible worlds semantics is one of the most widely used semantic frameworks for dealing with subjunctive conditionals in philosophy and linguistics (Stalnaker, 1968; Lewis, 1973; Bennett, 2003; Portner, 2009; Kratzer, 2012). Just like probability theory relies on a set of possible outcomes, possible worlds semantics relies on a set of possible worlds specifying alternative ways the world could be. Sets of such possible worlds represent propositions, which correspond to events in probability theory. The account states, roughly, that:

6

'If it had been the case that A, then it would have been the case that C' (A > C) is true in the possible world, *w*, iff C is true in all the A-worlds that are most similar to *w*.

For the examples below, '*w*' is the actual world in which the utterances are made. In Stalnaker (1968), the truth-conditions above are explicated via a selection function, f(A, *w*), which selects the closest world (or, alternatively: the set of closest worlds) to *w* in which A is true. The conditional, [A > C], is then true iff the selected A-world(s) is a subset of the set of worlds in which C is true, [C].[2] For a less formal illustration consider (3) as uttered in 2019:

(3) "If Hillary had won the 2016-election, the US relationship with the EU would have been better"

(3) is true roughly just in the following case. We consider the known facts in 2019 and (hypothetically) vary the actual circumstances such that Hillary won the 2016-election but keep as much else fixed as possible, such as Hillary's disposition to govern and the previous US/EU relationship. And, in that hypothetical world, the consequent is true: the US relationship with the EU is better than it was in 2019 with Trump as president.

Pragmatics, indicatives, and subjunctives

While Lewis (1973, 1976) restricted possible worlds semantics to subjunctives, Stalnaker (1975) extended the semantics to indicatives. In so doing he made use of pragmatic notions to account for the differences in meaning. To explicate the role of pragmatics, Stalnaker (1975) uses the notion of a common ground. The common ground between interlocutors can be understood as a context set of live possibilities that the interlocutors mutually presuppose. Through their assertions, interlocutors try to decide between the members of this *context set* to figure out which possible world they are really in (Roberts, 2015).

---

[2]   We here use square brackets to refer to the proposition expressed by a sentence as a way of specifying its truth-conditional content.

To illustrate this use of contextually presupposed information, suppose that some colleagues are discussing, in hushed tones, recent budget cuts at 'the Department'. For the conversation to proceed, the interlocutors must mutually presuppose much background information so that it is understood which department they are talking about (e.g., the psychology department). Moreover, the interlocutors also mutually presuppose a set of possibilities, which identify different candidates for deciding the latest budget cuts (e.g., Prof. Schmitt, the head of the department). In different possible worlds, the interlocutors might not work at the psychology department and Prof. Schmitt will not be the head of the department.

On Stalnaker's (1974, 2014) pragmatic notion of presuppositions, a sentence carrying presuppositions can only be felicitously uttered based on one of the following conditions: either its presuppositions are entailed by the common ground, or they are accepted by the hearer ("accommodated") as an update to the common ground for the purposes of the conversation. According to Stalnaker (1975), there is a pragmatic difference between indicative and subjunctive conditionals in that whereas an indicative conditional focuses only on the set of possibilities in the shared common ground, subjunctive conditionals allow interlocutors to consider remote possibilities, outside the common ground, in their assessment. For instance, in example (3) from above, all interlocutors presuppose that Trump did indeed win the 2016 election: there are no other live possibilities. But the interlocutors can consider possible worlds outside of this common ground in evaluating the counterfactual stating what would have happened if Hillary had won.

For this theory then, the semantics specifies the *form* of truth conditions of conditionals by referring to selection functions (i.e. '$A > C$' is true in w, iff C is true in the possible world/s returned by f(A, w)). *Which* proposition is expressed by the assertion of a conditional is, however, influenced by the context of utterance, given the pragmatic rule that f(A, w) outputs possible worlds *within* the context set for indicative conditionals, but is

permitted to output possible words *outside* of the context set for subjunctive conditionals. As a result, the theory holds that it is *appropriate* to use an indicative conditional only in a context that is compatible with the antecedent (Stalnaker, 1975). Since counterfactuals cannot conform to the constraint of selecting worlds inside the context set, however, they must be expressed as subjunctive conditionals. This principle captures the intuition that while the use of indicative conditionals generally is understood under preservation of as many of the mutually shared presuppositions of the conversation as possible, subjunctive conditionals are allowed to consider remote scenarios, where some of these assumptions are relaxed.

Subjunctive conditionals subtly cue their recipients through tense and mood that worlds may be selected outside the context set. As Iatridou (2000) has argued, subjunctives make use of an extra layer of past tense – e.g., the past perfect "had won" – as a "fake tense" to indicate the unlikeliness of the described situation rather than its temporal coordinates.

The Experiments

As we have outlined, while the possible worlds account has proved highly influential, it has not proved easy to test, a point on which we elaborate below. This paper offers a new method for testing the truth evaluations of both indicative and subjunctive conditionals according to possible worlds semantics. Experiment 1 introduces the novel truth task for indicatives and subjunctives at the aggregate level. Experiment 2 replicates these findings. Experiment 3 uses this task to focus on indicative conditionals and study individual variation by modeling participants' responses as coming from a mixture distribution of truth tables. Through three experiments, we contrast more than five different truth tables, including the novel account by Bradley (2012), which has not been tested previously, as far as we know.

Experiment 1: Truth Evaluations

The aim of Experiment 1 is to introduce an experimental setup for testing possible worlds semantics of conditionals (Stalnaker, 1968; Lewis, 1973). To motivate it, we start out by explaining some of the difficulties with testing possible worlds semantics.

Difficulties in Testing Possible World Semantics

Possible worlds semantics for counterfactuals is widely used in linguistics and philosophy (Portner, 2009; Kratzer, 2012), because by placing weak constraints on the similarity relation—e.g., of all worlds, *w* is most similar to itself (*strong centering*)—it allows the formulation of several counterfactual logics (Lewis, 1973). Yet, it turns out to be a very difficult task to explicate this similarity relation in a way that would permit the evaluation of ordinary counterfactuals in light of a set of strenuous counterexamples, and the inevitable context-sensitivity of such assessments (Rescher, 2007; Nickerson, 2015; Ippolito, 2016; Starr, 2019). To illustrate, a case can be made for either one of the following counterfactuals concerning the Korean War (Quine, 1960, p. 222; Spohn, 2013):

> If Caesar were in command, he would use the atomic bomb.
> If Caesar were in command, he would use catapults.

One of the central difficulties in testing possible worlds semantics is that participants are likely to consider very different possible worlds if left unconstrained due to differences in their background beliefs and in views on what counts as "similar". For this reason, the experimental task introduced below creates a situation, in which there is one salient hypothetical alternative to the actual state of affairs to better control what counts as most similar in the task context.

The Possible Worlds Truth Table Task

In our experiments, participants were instructed that they were going to see a small sample of photos taken by Jack. The participants were then presented with different pairs of photos, which were later used to formulate the indicative and subjunctive conditionals, as illustrated in Table 2 below. In the example below, the pair of pictures illustrate one of Jack's photos of a kitchen and of a railroad station. In the actual experiment, further pairs were presented.

First, participants were asked a range of control questions concerning these pictures (explained below). Later, one of the pictures would be presented on the page and participants were asked to evaluate the truth value of the presented indicative/subjunctive conditional. For ease of illustration, we will refer to the picture presented as "the Actual Picture" and the other picture of the pair, which was not shown, as "the Hypothetical Picture", but note that these labels were never presented to the participants.

We introduced this manipulation to investigate whether subjunctive conditionals with false antecedents would prompt participants to think about the counterfactual situation in which the Hypothetical Picture would be shown instead of the Actual Picture, as illustrated in the thought bubble in Table 2. The function of the previous exposure to a pair of pictures is to constrain assessments of similarity by making salient one possibility of an alternative picture that could have been presented. This way, the task introduces an extra source of information (i.e., information about what is true/false in a salient possible A-world) beyond the manipulated truth table cells in the actual world (TT, TF, FT, FF). The goal is to probe whether participants selectively make use of this extra source of information when assigning truth values to subjunctive conditionals. For instance, when presented with the picture of a kitchen and the conditional "If this had been one of Jack's pictures of a railroad station, there would have been a warning sign about standing too close to the edge in it", do they consider what is true in the picture of the railroad station?

Table 2. Example of Stimulus Material

Actual Picture	*Hypothetical Picture*

TT: If this had been one of Jack's photos of a kitchen,	if [*Actual picture*],
there would have been a fruit bowl in it.	[*Actual object*]
TF: If this had been one of Jack's photos of a kitchen,	if [*Actual picture*],
there would have been a warning sign about standing too close to the edge in it.	[*Hypothetical object*]
FT: If this had been one of Jack's photos of a railroad station,	if [*Hypothetical picture*],
there would have been a fruit bowl in it.	[*Actual object*]
FF: If this had been one of Jack's photos of a railroad station,	if [*Hypothetical picture*],
there would have been a warning sign about standing too close to the edge in it.	[*Hypothetical object*]
FF_{misplaced}: If this had been one of Jack's photos of a railroad station,	if [*Hypothetical picture*],
there would have been a pair of bedlamps in it.	[*Misplaced object*]

Note. The pictures were displayed in a size in which it was possible for the participants to discern the individual objects. *Actual Object*: bowl with fruit. *Hypothetical Object*: a warning sign about standing too close to the edge. *Misplaced Object*: a pair of bedlamps.

We further introduced a contrast between two ways of implementing the FF cell. We did this to probe whether participants were sensitive to variation in whether the objects talked about in the conditional sentences were present/absent on the Hypothetical Picture, although they would in each case be absent on the Actual Picture (thus giving rise to the FF cell, when evaluated relative to the Actual Picture). We label the first version of this truth table cell "FF" (where the object talked about is present in the Hypothetical Picture but absent in the Actual Picture). The second version where the object is also absent on the Hypothetical Picture, we label $FF_{misplaced}$. To illustrate: whereas the Hypothetical Object (e.g., a warning sign) was present on the Hypothetical Picture and absent on the Actual Picture, a third Misplaced Object (e.g., a pair of bedlamps) was present on neither. The $FF_{misplaced}$ condition concerns this third misplaced object. If participants were evaluating the conditional based on the Actual Picture, then this contrast between two versions of the FF cell (FF vs. $FF_{misplaced}$) should not make a difference to their truth evaluations. If the conditionals were evaluated based on the salient Hypothetical Picture (e.g., Jack's picture of a railroad station), however, then the contrast between the Hypothetical Object and the Misplaced Object should play a role.

Thought of in terms of standard truth tables, one can think of this manipulation as implementing two versions, {TT, TF, FT, FF} and {TT, TF, FT, $FF_{misplaced}$}, and our interest is whether participants are systematically influenced in their truth evaluations of indicative and subjunctive conditionals by these two versions. When we discuss our findings below, we abbreviate the two versions as {TT, TF, FT, FF, $FF_{misplaced}$} and investigate how the various truth tables perform across all five types of truth table cells.

Predictions of Competing Truth Tables

While Experiment 1 is intended primarily to test the possible worlds account, it also provides evidence that has a bearing on other truth tables of the conditional: namely, the material implication, the de Finetti truth table, the Jeffrey Table (see Table 1), mental model

theory, and the multi-dimensional approach of Bradley (2012). In Experiment 1, we attribute

the predictions in Table 3 to the theories tested for indicative conditionals. The basis for these

predictions is explained below.

Table 3. Predictions for Indicative Conditionals in Experiment 1

Cell	⊃	Possible Worlds	de Finetti	Bradley	Jeffrey	MMT$_{aux}$
TT	1	1	1	1	1	1
TF	0	0	0	0	0	0
FT	1	If (accom =1) 0; else 1	NA	0	0	0
FF	1	If (accom =1) 1; else 0	NA	1	1	1
FF$_{misplaced}$	1	0	NA	0	0	0
Model	$M_{Material}$	$M_{PossibleWorld}$	$M_{deFinetti}$	$M_{subjunctive}$		

Note. '⊃' = Material Implication. 'accom' = accommodate. The prediction for Bradley (2012) is determined by the task constraints of Experiment 1, since there is only one possible world that enters as a candidate for the counterworld, w$_j$, and this is the Hypothetical Picture. The truth value for the false antecedent cells (FT, FF, FF$_{misplaced}$) is therefore determined by the pair <actual picture, hypothetical picture>. The last row labels the truth tables by the multinomial processing tree models fitted in Experiment 1. The last row indicates which statistical model in Table 6 the truth table was mapped onto. The name 'M$_{subjunctive}$' was chosen for the last three tables, because the indexed truth tables make the same predictions for indicative conditionals in this specific task as the modal values of the subjunctive conditionals.

As Edgington (2006) explains, the material implication and the possible worlds

account in Stalnaker (1968) differ as follows. While the material implication holds that the

indicative conditional is true when the antecedent is false, the possible worlds account in

Stalnaker (1968) permits the conditional to be either true or false. For the possible worlds

account, the conditional is, in other words, not truth-functional in that its truth value is not a

function of the truth values of its clauses. Instead, the truth or falsity of the conditional in

these cases is determined by whether the consequent is true in the most similar situation in

which the antecedent is true (see Table 1).

The possible worlds semantics of Stalnaker (1975) applies the same truth conditions

for indicatives and counterfactuals conditionals, but accounts for differences between the two

through a pragmatic principle. If we are to fully understand the predictions of the possible

worlds account, we must briefly expand on this pragmatic principle.

As we have seen, Stalnaker (1975) invokes the notion of a common ground in his

account of indicative and counterfactual conditionals. Whereas an indicative conditional

focuses only on the set of possibilities in the shared common ground, subjunctive conditionals allow interlocutors to consider remote possibilities, outside the common ground, in their assessment. Accordingly, it is appropriate to use an indicative only when the antecedent is not known to be false; otherwise, a subjunctive must be used. In Stalnaker (2011, 2014), this theory is extended using the notion of *accommodation*. Although presuppositions present information as taken-for-granted – as uncontroversially part of the common ground – they can be used to communicate novel information as well. But when presuppositions communicate novel information, they trigger a special mechanism.

To illustrate, suppose that a speaker says, "My sister is a schoolteacher" in a context, where it cannot be presupposed that the interlocutor even knew that the speaker had a sister. The speaker then goes beyond the common ground. But the utterance need not be infelicitous. The hearer can accept the presupposition as true – can *accommodate* it – and add the information to the common ground. Typically, this will occur automatically, unless someone objects (Lewis, 1979; Stalnaker, 2014). Accommodation thus has the discourse effect of making the hearer adjust her knowledge so that it entails the content that the speaker presupposes (Potts, 2005, Roberts, 2015).

In (2011, 2014), Stalnaker extends the theory to permit accommodation of the common ground to make the context set compatible with the antecedent, if possible. The accommodation then creates a *posterior* context with respect to which the conditional is evaluated. Whereas the evaluation of counterfactuals permits the use of hypothetical assumptions that are not carried over to the main discourse, the accommodated presuppositions of this posterior context remain presuppositions in the subsequent discourse.[3]

---

[3]   A further possibility that Stalnaker (2011) discusses is to allow for truth-value gaps in case of false antecedents via a supervaluation approach for dealing with contextual ambiguity, just like the de Finetti table of Wijnbergen-Huitink et al. (2015) employs truth-value gaps. However, given our stimulus materials introduce a well-defined context without such

Accommodation is, thus, an important aspect of contemporary possible worlds semantics, and it will be important to investigate whether participants perform such accommodation when faced with indicative conditionals with false antecedents.

If participants accommodate the common ground to make the context set compatible with the antecedent of the conditional, then the boundary of which possibilities count as being outside the context set shifts. Such shifts are predicted to affect the truth evaluation of indicative conditionals. For instance, when participants are presented with a picture of a kitchen and the conditional "If this is one of Jack's pictures of a railroad station, then there is a warning sign about standing too close to the edge in it", do they accommodate the antecedent and evaluate the conditional with respect to the picture of the railroad station, or do they decline to accommodate and evaluate the conditional with respect to the presented picture of the kitchen? In our experiments, we test for the possible dependence of truth value judgments of indicative conditionals on accommodation. In Table 3 these shifts are illustrated through the "if … else …" clauses, which specify how the predicted truth values vary depending on whether participants accommodate the antecedent. In Experiment 2, we will return to the question of what the participants are doing who do not accommodate the antecedent in this way.

Bradley's (2012) account shares the non-truth functionality of possible worlds semantics but implements it differently. For Bradley (2012), the semantic content of conditionals is given by a set of ordered pairs of possible worlds: $\{<w_i, w_j>, …\}$. The first member (w_i) of such a pair is the actual world and the second world (w_j) is a counterworld in which the antecedent is true. If the consequent is true in w_j, then the conditional is true in $<w_i, w_j>$, otherwise it is false. Probabilities are assigned to conditionals by assigning probabilities

---

ambiguity, this extension of the possible worlds account is set aside for present purposes (see Stalnaker, 2019, Chap. 11 for further discussion).

to counterworlds. One of the great advantages of this technical refinement is that it allows Bradley (2012) to circumvent a problem known as the triviality results. The problem dates back to Lewis' (1976) demonstration that attempts to combine P(if A then C) = P(C|A) with classical truth conditions can only succeed for probability distributions subject to trivializing features that severely restrict their usefulness (for a review, see e.g., Bennett, 2003).

Through his constructions with truth conditions explicated via pairs of possible worlds, Bradley (2012) is able to show that he can combine the thesis, P(if A then C) = P(C|A), with truth conditions without being subject to the triviality results. This feature of Bradley's (2012) account is one of its main attractions. Via the exposure to participants of two pictures that Jack could have taken, only one of which is shown when evaluating the conditionals, our task allows us to experimentally fix the counterworld (w_j) and thus test Bradley's theory, possibly for the first time.

In contrast, the de Finetti truth table abandons classical truth conditions in a different way by holding that the indicative conditional is neither true nor false when the antecedent is false (Evans & Over, 2004; Fugard, Pfeifer et al., 2011; Baratgin, Over et al., 2013). A different,[4] but related approach, known as the Jeffrey-Table, holds that the semantic value of the conditional is given by P(C|A) in false-antecedent cases (Jeffrey, 1991; Stalnaker & Jeffrey, 1994; Over & Baratgin, 2017). As noted above, attempts have also been made to extend these truth tables to counterfactuals, so we will consider whether they can account for our data concerning subjunctive conditionals below.

---

[4]   Often these two approaches are taken to be similar, but Bradley (2012) shows that they differ on how to understand the probability of conditionals and that the semantic content attributed to indicative conditionals has a different meaning on the two accounts. Roughly, the de Finetti table avoids the triviality results by abandoning Bivalence and reinterpreting the probability of sentences as their probability of truth, *provided they are true or false*. In contrast, the Jeffrey table avoids the triviality results by dropping the independence of belief and the meaning of sentences by making the semantic values of indicative conditionals dependent on an agent's subjective degrees of belief.

Over and Cruz (2018) and Over (2020), have suggested that one can translate this third semantic value of the Jeffrey table into truth values supplied by participants via the auxiliary hypothesis that they assign the value 'true' when P(C|A) is "high" and 'false' when P(C|A) is "low". When outlining the predictions in Table 3, we rely on this empirical, auxiliary hypothesis. The resulting predictions are, in other words, specific to our experimental paradigm, and they presuppose that participants assign "low", "high", and "low" conditional probabilities for the FT, FF, and FF$_{misplaced}$ cells, respectively. In our experiments, we test this auxiliary hypothesis, and to anticipate, we obtain supporting evidence.

While mental model theory may earlier have adopted the material implication (Johnson-Laird & Byrne, 1991, 2002), the theory has since then been revised to reject such a commitment (Khemlani et al., 2018). On the revised version, indicative conditionals are viewed as conjunctive assertions about possibilities as shown in table 4.

Table 4. Mapping between indicative and counterfactuals, MMT

Row	Partition		Factual:	Counterfactual:
			If A then C	*If A had happened then C would have happened*
1	A	C	Possibility	Counterfactual possibility
2	A	Not-C	Impossibility	Impossibility
3	Not-A	C	Possibility	Counterfactual possibility
4	Not-A	Not-C	Possibility	Fact

Note. Quelhas et al. (2018) call indicative conditionals "factual conditionals".

Whereas the indicative conditional asserts that only rows 1, 3, and 4 are possible, the counterfactual asserts that row 4 is a fact and that rows 1 and 3 are counterfactual possibilities which did not materialize. Since conditionals are viewed as conjunctive assertions about the rows of Table 4 on the revised mental model theory, there is a difficulty of how to assign truth values when participants are only presented with one of the rows, as in our experimental task. Quelhas et al. (2018) point out that the false antecedent cases ($\neg A \& C$ and $\neg A \& \neg C$) are asserted to be possible for both true and false conditionals. So, these are not diagnostic for the truth or falsity of the conditional. However, Goodwin and Johnson-Laird (2018, pp. 2529-

2530) suggest a way of determining the truth of indicative conditionals when presented with only one of these cases. The idea is to make the truth of the indicative dependent on the truth of the corresponding counterfactual. In their example, if in fact Viv does not have shingles, then "If Viv has shingles, then she is in pain" is true provided that the following counterfactual is true: "If Viv had had shingles, then she would have been in pain". Accordingly, when one only possesses information about row 3 or 4 in Table 4, one can learn which of rows 1 and 2 are possible by considering a counterfactual version. Consequently, we attribute the prediction to mental model theory of lack of differences between indicative and counterfactual conditionals for the false antecedent cases as an auxiliary hypothesis, as outlined in Table 3.[5]

Method

Participants, and sampling procedure shared for all experiments

The experiment was conducted over the Internet to obtain a large and demographically diverse sample. A total of 292 people completed the experiment. The participants were sampled through the Internet platform Mechanical Turk from the USA, UK, Canada, and Australia. They were paid a small amount of money for their participation. The following *a priori* exclusion criteria were used: not having English as native language, completing the task in less than 240 seconds or in more than 3600 seconds, failing to answer two simple SAT comprehension questions correctly in a warm-up phase, and answering 'not serious at all' to the question 'how serious do you take your participation' at the beginning of the study. The

---

[5] One of our reviewers points out that there is a possible tension between the views in Quelhas et al. (2018), which could be read as pointing in the direction of a three-valued presupposition account of indicative conditionals, and the view of the revised mental model theory that indicative conditionals are conjunctions of claims about possibilities that are either true or false. In this paper, we will not attempt to resolve this issue, or other related controversies concerning mental model theory discussed in Oaksford et al. (2019), but merely test the predictions based on the auxiliary assumption in Goodwin and Johnson-Laird (2018) outlined above.

final sample after applying the *a priori* exclusion criteria consisted of 211 participants.[6] Mean age was 44.06 years, ranging from 19 to 75. 42.65% of the participants identified as male; with the exception of 3, the rest identified themselves as female. 74.88 % indicated that the highest level of education that they had completed was an undergraduate degree or higher. Applying the exclusion criteria had a minimal effect on the demographic variables.

Design

The experiment had a within-subject design with the following within-subject factors, which are explained below: Sentence (indicative vs. subjunctive) and Truth Table Cell (TT, TF, FT, FF, FF$_{misplaced}$). Since the experiment had three trial replications, participants saw a total of 30 within-subject conditions.

Procedure

To reduce the dropout rate during the experiment, participants first went through three pages stating our academic affiliations, posing two SAT comprehension questions in a warm-up phase, and presenting a seriousness check asking how careful the participants would be in their responses (Reips, 2002). Moreover, to ensure that the pictures were displayed properly if the participants completed the study on a smartphone, participants were asked to turn their smartphone in horizontal orientation, if they were using one.

The experiment was split into three blocks in random order. For each block, participants were instructed that they were going to see a small sample of photos taken by Jack. For each block, a pair of pictures was selected out of a pool of six possible pairs. The pairs were generated based on the stimulus materials used in a pilot study[7]. For instance, one pair of pictures might look as in Table 2, where the kitchen picture was selected as the Actual Picture shown.

---

[6]   In addition to the common exclusion criteria for all the experiments, two participants were excluded in Experiment 1, because the javascript was not displaying correctly.
[7]   https://osf.io/6x7fb/

For each of these pictures, participants were asked, in a random order on separate pages, to answer whether it was true/false that three objects were on them, as control questions. These questions established whether the truth-table manipulation worked, that is, whether participants could correctly identify whether objects were present in, or absent from, a given picture. On a separate page, participants were asked to evaluate conditional probabilities (of objects being in the picture under the assumption that the antecedent was true) for fitting the Jeffrey table. Next, participants were asked to evaluate indicative and subjunctive conditionals for whether they were true, false, or neither. Each picture in the pair was randomly assigned a type of conditional (indicative vs. subjunctive), so that participants would see one type for each pair. Accordingly, participants were presented with two subblocks for each pair of pictures: one subblock with indicative conditionals and one subblock with subjunctive conditionals in random order with random assignment to pictures. Each subblock consisted of five separate pages, in which participants were presented with the Actual Picture (e.g., the kitchen picture) and five conditionals (e.g., five subjunctive conditionals), implementing the truth table cells shown in Table 3 in a random order. Participants were then asked to evaluate whether the presented sentence was true (T), false (F), or neither true nor false (NN). Key words distinguishing indicative conditionals ("…is…, …is…") and subjunctives ("...had been …, …would ...") were highlighted in blue. The order of conditionals and response types was randomized.

After completing the task, participants were presented with an accommodation task. The accommodation task featured three pages with pictures (e.g., Jack's picture of a kitchen) and indicative conditionals in the FF condition from the pairs displayed earlier in the study, that asked three times about what photo they took the sentence as referring to.

If participants accommodate the antecedent, which is incompatible with the picture shown (and the demonstrative reference: "If *this*..."), then they should select the hypothetical

picture of a railroad station not shown. If, on the contrary, participants evaluate indicative conditionals with respect to the picture shown, they should select the displayed picture of a kitchen. Sample screenshots of these three accommodation items can be found on the osf project page: https://osf.io/6x7fb/.

Results

It was found across all pictures that participants correctly identified whether the Actual, Hypothetical, and Misplaced Objects were on the pictures in the initial control questions (median percentage of correct responses = 93 %, MAD = 2%). Participants' truth evaluations are displayed in Figure 1:
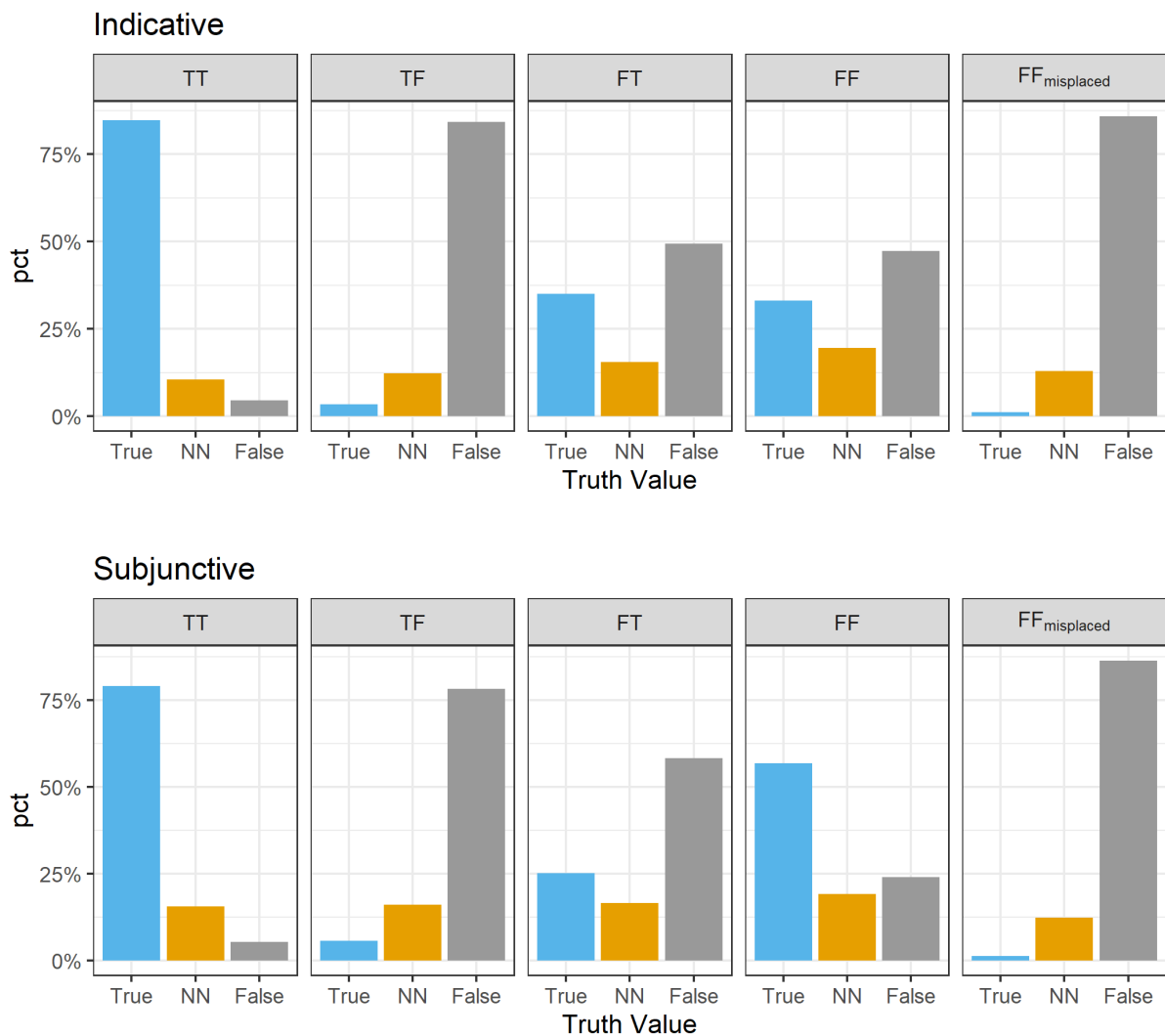
Figure 1. Truth Tables of Indicative and Subjunctive Conditionals. 'FF' = a conditional that is False False w.r.t. the Actual Picture shown but True True w.r.t. the salient Hypothetical Picture; 'FF$_{misplaced}$' = a conditional that is False False w.r.t. the Actual Picture shown but True False w.r.t. the salient Hypothetical Picture.

It was found across conditions that the average conditional probabilities were .08 (SD = .09), .93 (SD = .09), and .04 (SD = .07), for the Actual Object, Hypothetical Object, and Misplaced Object to be on the picture under the supposition that the picture displayed was the Hypothetical Picture, when performing the Ramsey test. Accordingly, our results support assigning the truth values of Table 3 to the Jeffrey table by applying the auxiliary hypothesis of Cruz and Over (2018) and Over (2020), as explained above.

Participants' truth evaluations were analyzed in two steps. First, differences between indicative and subjunctive conditionals were analyzed. Next, existing truth tables for indicative conditionals were fitted to the data. For both analyses, the observed response frequencies were analyzed with multinomial processing tree models (Riefer & Batchelder, 1988). This modeling framework is typically used to characterize the processes that underlie participants' categorical responses (see Batchelder & Riefer, 1999; Erdfelder et al., 2009). However, the framework can also be used to test hypotheses at the level of the observed response distributions through goodness of fit and model-selection statistics (e.g., Karabatsos, 2005; Klauer et al., 2015; Skovgaard-Olsen et al., 2017). For a Bayesian implementation, we followed the hierarchical extension of multinomial processing trees in Klauer (2010), which was fitted via the R package TreeBUGS (Heck et al., 2018). Further technical details on the model can be found in Appendix C.[8]

Indicatives vs. Subjunctives

---

[8] For R scripts see: https://osf.io/6x7fb/.

Using this framework, the following multinomial processing models were fitted to the data for both indicative and subjunctive conditionals to model the probabilities that a categorical response was selected:

$M_{saturated}$:   model imposing no constraints. This model fits the data perfectly using one free parameter per degree of freedom provided by the data

$M_{sentence}$:   model assuming that response probabilities are the same for indicatives and subjunctive conditionals

$M_{FF,FFmis, ind}$:   model assuming no differences between the FF and $FF_{misplaced}$ truth table cells for indicative conditionals

$M_{FT, FF, ind}$:   model assuming no differences between the FT and FF truth table cells for indicative conditionals

Table 5. Model-Comparison Results

Model	p_{T1}	p_{T2}	WAIC	Weight
$M_{FT,FF,ind}$	.37	.00	4683.8	0
$M_{saturated}$	.49	.11	4070.9	1
$M_{sentence}$	.00	.00	6705.4	0
$M_{FF,FFmis,ind}$	.00	.00	4599.7	0

Note. Note that the test statistics T_1 and T_2 represent Bayesian *p* values and are based on the posterior predictive model checks in Klauer (2010). WAIC = Watanabe-Akaike information criterion. Weight = Akaike weight of WAIC.

Model fit was assessed with the WAIC information criterion and posterior predicted *p* values based on T_1 and T_2 posterior model checks proposed in Klauer (2010). T_1 measures the adequacy of the models in capturing the mean observed truth value frequencies. T_2 measures the adequacy of the models in capturing the variability (variances and covariances) among the observed response frequencies. A small (Bayesian) *p* value for these test statistics indicates that the posterior predictive distribution of the model fails to capture an aspect of the data, because this aspect of the actual observations is unlikely to be predicted by the model. When testing several of these posterior predictive checks, it is not uncommon for a model to be

24

inadequate for some purposes but adequate for others, and the test statistics help us identify which aspects of the data are captured by a given model (Gelman et al., 2013: Ch. 6).

Based on these criteria, ($M_{FT, FF, ind}$) was the only model besides the saturated model (M_{sat}), which was able to capture the mean observed truth value frequencies (T_1). Yet, like the other non-saturated models, it accounted less well for their variances and covariances across individuals (T_2). Accordingly, the information criterion WAIC indicates that this model reduction was not justified compared to the saturated model. From the comparisons, we can infer that a) there is a difference in the truth value assignments of indicative and subjunctive conditionals, b) the difference between FF and $FF_{misplaced}$ plays a role even for indicative conditionals, but c) the difference between FT and FF need not be taken into account to account for the mean observed truth value frequencies of indicative conditionals; yet it plays a role for the variability in responses across individuals. Moreover, a glance at Figure 1 reveals that while the difference between FT and FF may not matter for indicative conditionals, it flips the modal value for subjunctive conditionals from False to True.

Truth Tables for Indicatives

To further investigate participants' truth value assignments for indicative conditionals, we fitted the truth tables from Table 3 to the data (see Table 6 below). As indicated in Table 3, the truth table for the possible worlds semantic depended on whether participants accommodated the antecedent. As a result, the model implementing this truth table was constrained to predict the truth value assignment {T, F, F, T, F} for the participants (N = 95), who accommodated the reference of the antecedent to refer to the Hypothetical Picture. In contrast, the model was constrained to predict {T, F, T, F, F} for the participants (N = 116) who did not accommodate to follow the truth evaluation of the Actual Picture vis-à-vis the strategy investigated in Experiment 1. This classification concerns participants who accommodated the reference of the false antecedent all three times when asked. Accordingly,

the prior expectation of answering "yes" all three times by a random binomial process is that only 12.5% of the participants should have fallen in this group, instead 45% were found. Appendix A displays the bimodal pattern of these two different truth evaluations of indicative conditionals separately, which are merged in the aggregate results shown in Figure 1.

Following Skovgaard-Olsen et al. (2017), the most lenient criterion was applied in assigning a stochastic interpretation to the deterministic predictions of the truth tables using the order constraints in Klauer et al. (2015), as explained in Appendix C. In each case, only a relative majority of the predicted response was required of a given model (i.e., that the predicted response occurs at least as often as each of the other responses). In contrast, an absolute majority could also have been required. But the advantage of using this very lenient criterion lies in the diagnostic power associated with its failure, as any theory that fails under these minimal constraints should be seriously questioned.

Table 6. Model-Comparison Results for Indicatives

Model	p_{T1}	p_{T2}	WAIC	Weight
$M_{PossibleWorld}$	0.13	0.00	2258.9	1
$M_{Subjunctive}$	0.00	0.00	2549.7	0
$M_{Material}$	0.00	0.00	3405.8	0
$M_{deFinetti}$	0.00	0.00	3606.7	0

Note. The test statistics T_1 and T_2 represent Bayesian *p* values and are based on the posterior predictive model checks in Klauer (2010). WAIC = Watanabe-Akaike information criterion. Weight = Akaike weight of WAIC. In Table 3, the truth tables are displayed which correspond to these models.

Model fit was assessed with the WAIC information criterion and posterior-predicted *p* values based on T_1 and T_2 posterior model checks proposed in Klauer (2010). Of all the models tested, only $M_{PossibleWorld}$ had a satisfactory fit for the aggregate truth value frequencies (T_1). But like the other models, it accounted less well for their variances and covariances across individuals (T_2). The information criterion WAIC indicates a strong preference for $M_{PossibleWorld}$ in light of the parsimony vs. fit trade-off. We discuss why $M_{PossibleWorld}$ had this edge compared to the other models further below.

Discussion

As seen from Figure 1, participants' truth evaluations of subjunctive conditionals elicit the modal pattern {T, F, F, T, F} in spite of being shown a picture that would generate the following truth table cells: TT, TF, FT, FF, FF. This pattern fits with the following hypothesis. When participants evaluate subjunctive conditionals with false antecedents, they do not consider whether the sentence is true of the Actual Picture displayed (e.g., Jack's photo of a kitchen). Instead, they consider whether it is true of the Hypothetical Picture (e.g., Jack's photo of a railroad station) that the experiment made salient. This pattern corroborates the possible worlds account of subjunctive conditionals (Stalnaker, 1968; Lewis, 1973). Given the way the task was set up, participants were already familiar with a prototype that could make the antecedent true of the subjunctive conditionals. It was thereby possible for participants to solve the task of evaluating subjunctive conditionals at the closest possible-A worlds in a relatively uniform manner. These results thereby provide some of the first direct evidence that participants' truth evaluations coincide with a possible worlds interpretation of subjunctive conditionals. Neither the material implication nor the de Finetti table can capture the modal pattern in the truth evaluations for subjunctive conditionals. Applying mental model theory to account for the same results is, however, possible but raises special theoretical issues, which we take up in the General Discussion.

The results moreover indicate that participants clearly distinguish between the truth evaluation of subjunctive and indicative conditionals. In contrast to subjunctives, the results for indicatives showed considerable individual variation. Given this variation, we see that the conjecture in Goodwin and Johnson-Laird (2018)—that indicative conditionals with false antecedents are true if the corresponding counterfactuals are true—significantly misfits the data. More specifically, the account encountered difficulties in the FF cell in which the subjunctive was evaluated as True yet the modal value for the matching indicative was False.

In general, it was found for indicative conditionals that influential truth tables, like the material implication (T, F, T, T) and the de Finetti table (T, F, NN, NN), severely misfit the data. In this, the present results are in line with the results of a recent meta-analysis (Schroyens, 2010). Some studies have recently reported stronger evidence in favor of the de Finetti truth table for indicative conditionals, however (e.g., Evans et al. 2007; Politzer et al., 2010; Wijnbergen-Huitink et al. 2015). But these studies have either relied on an experimental paradigm that asks whether a truth table cell "conforms to"/"contradicts" or "is irrelevant" to a conditional rule (for discussion, see Skovgaard-Olsen, 2021), or relied on a betting paradigm, which likewise requires participants to consider whether the evidence presented verifies a conditional rule. In contrast, the de Finetti truth table is much less well-supported when participants are asked to assign ternary truth values as here (Schroyens, 2010; Skovgaard-Olsen et al. 2017).

Proponents of the suppositional theory of conditionals (see, e.g., Over & Baratgin, 2017) have emphasized the need for replacing the "void" values of the de Finetti table with the conditional probabilities of the Jeffrey table (see Table 1). To test this idea, we measured participants' conditional probabilities. Based on the measured probabilities, we would have expected participants to give the same truth values to indicatives with false antecedents as the modal truth values for subjunctives. However, an inspection of Figure 1 shows why such a model did not fit the mean response frequencies. The main difficulty consists in accounting for the differences between indicative conditionals and subjunctives in the FT and FF cells. So, while a Jeffrey table is compatible with the truth values for subjunctive conditionals, it cannot account for the modal truth values for indicative conditionals in our experiments.

Of all the investigated alternatives, it was found that a possible worlds account was the only model which did not significantly misfit the data on the aggregate level. What permitted this account to outperform the other models was the following. The possible worlds model

adjusted its predictions based on whether participants accommodated the reference of indicative conditionals with false antecedents to refer to the Hypothetical Picture not shown (despite the demonstrative reference to the Actual Picture shown). Ca. 55% of the participants chose to accommodate in this way across three trials. It turned out that this difference in whether participants accommodated was a factor in the truth evaluation of indicative conditionals, as shown in Appendix A. But given that the predictions of the truth tables used a stochastic representation that only required that the preferred option should be the modal response, there is scope for further investigations into factors that may influence individual variation in the truth evaluation of indicative conditionals. Indeed, the failure of all models to account for the variances and covariances across individuals indicates that further research into individual variation is needed. In Experiment 3, we will return to this issue, but first Experiment 2 investigates another question left open by the results of Experiment 1.

Experiment 2

As we have seen, Stalnaker's (1975) account of the distinction between indicative and subjunctive conditionals is rooted in pragmatic principles concerning whether the closest possible world satisfying the antecedent is to be sought inside or outside the context set of mutually shared presuppositions of the interlocutors. In the 1975 paper, Stalnaker emphasizes that these pragmatic principles are not universal generalizations but that they permit exceptions. Stalnaker makes use of these exceptions to account for cases of subjunctive conditionals that do not have a counterfactual interpretation, as we shall see in the General Discussion. Yet, as one of our reviewers point out, in (2014, p. 157) Stalnaker emphasizes that if the context set does not satisfy the antecedent then the interlocutor must accommodate to include possibilities in the context set which do.

In Experiment 1, we examined whether participants resolve the referential ambiguity of being presented with an antecedent that includes a demonstrative reference ('this') and a

descriptive characterization (e.g., 'one of Jack's pictures of a railroad station') that is not satisfied by the presented picture by accommodating and taking the conditional to refer to a hypothetical picture not shown. It was found that about half the participants did this and that their previously supplied truth evaluations of indicative conditionals could be predicted by whether they did indicate that they accommodated the antecedent in this way.

The goal of Experiment 2 was to investigate two hypotheses concerning what the remaining participants do when evaluating indicative conditionals. The first hypothesis is due to one of our reviewers, and it states that participants who do not accommodate the antecedent by taking the antecedent to refer to the hypothetical picture not shown, will respond uniformly with 'False' as a reaction to the false antecedents to indicate the mismatch between the picture mentioned and the picture shown. The second hypothesis is that the participants who do not accommodate by taking the false antecedent to refer to the hypothetical picture accommodate in a different manner by imagining a way in which the actual picture shown could have been different such as to satisfy the antecedent. For these participants, the demonstrative reference fixes the designation to the actual picture shown, but they engage their imagination to find some way in which it could satisfy the antecedent (e.g., by imagining that Jack took the picture differently, or by bracketing their beliefs and imagining that the actual picture shown did satisfy the descriptive characterization). This second imaging strategy was suggested to us when interviewing participants and hearing them think-out-loud in pilot studies.

To test these two hypotheses, Experiment 2 seeks to replicate the results of Experiment 1 and to pose an additional follow-up question for participants who indicate that they do not accommodate the false antecedents of indicative conditionals by taking the conditionals to refer to the hypothetical picture not shown.

Method

Participants

30

Unless otherwise noted, the sampling procedure, experimental design, and materials of Experiment 2 followed Experiment 1. A total of 302 people completed the experiment. The final sample after applying the *a priori* exclusion criteria consisted of 222 participants.[9] Mean age was 45.47 years, ranging from 19 to 75. 48.20% of the participants identified as male; with the exception of 3, the rest identified themselves as female. 80.18 % indicated that the highest level of education that they had completed was an undergraduate degree or higher. Applying the exclusion criteria had a minimal effect on the demographic variables.

Design

Participants saw the same 30 within-subject conditions as in Experiment 1. The only difference was the addition of a follow-up question for the subset of participants who indicated that they did not accommodate the false antecedent by taking it to refer to the Hypothetical Picture not shown, for each of the three times that we posed the question.

Procedure

Experiment 1 was conducted before Covid-19, whereas Experiment 2 was conducted in 2022 after the lockdowns. As a result, we had to change one of the items. During the pandemic, participants have learned to expect disinfection dispensers in restaurants and so this item was no longer suitable as a pairing item from a veterinary treatment room that was displaced in a restaurant. Instead, it was replaced with an operating table in Experiment 2.

In addition, before providing the truth evaluations of the isolated antecedent and consequent sentences to verify that participants had understood the truth table cells shown, the instructions were extended with the information that they should only consider objects that are visible on Jack's photo to discourage participants from thinking about blind angles in the photos or hidden objects.

---

[9]   In addition to the common exclusion criteria for all the experiments, 8 participants were excluded for providing the same responses uniformly across all truth table cells (e.g., for responding 'NA' irrespectively of whether a TT, TF, FT, or a FF cell was shown).

After completing the truth evaluations of indicative and subjunctive conditionals, participants were presented with the same accommodation task as in Experiment 1. For each trial of that task, a follow-up question was added if participants indicated that they took the false antecedent of the conditional presented to refer to the picture shown (as opposed to the paired Hypothetical Picture not shown).[10] On this page, the false antecedent conditional (e.g., "If this is one of Jack's pictures of a kitchen, then there is a bowl with fruit in it") and the presented picture (e.g., of a railroad station) were presented as a reminder, and participants were asked as follows:

> When you take the if-then sentence below to be about the displayed picture [e.g., of a railroad station], which of these options best describes your line of reasoning:
>> I imagine a situation in which the displayed image depicts [e.g., a kitchen] despite the fact that it is actually [e.g., a railroad station].
>> I notice that the first part of the sentence is simply false and on that basis evaluate the entire sentence to be false.

Since the accommodation task was presented three times (one for each of the pair of pictures that the participants had been shown), participants had three opportunities to respond to this follow-up question, provided that they selected that they took the false antecedent conditional to refer to "the actual picture shown" in the first accommodation task.

Results

It was found across all pictures that participants correctly identified whether the Actual, Hypothetical, and Misplaced Objects were on the pictures in the initial control questions (median percentage of correct responses = 94 %, MAD = 8.24%).

---

[10] The stimulus materials and a screenshot of this extra follow-up question can be found on the OSF project page: https://osf.io/6x7fb/.

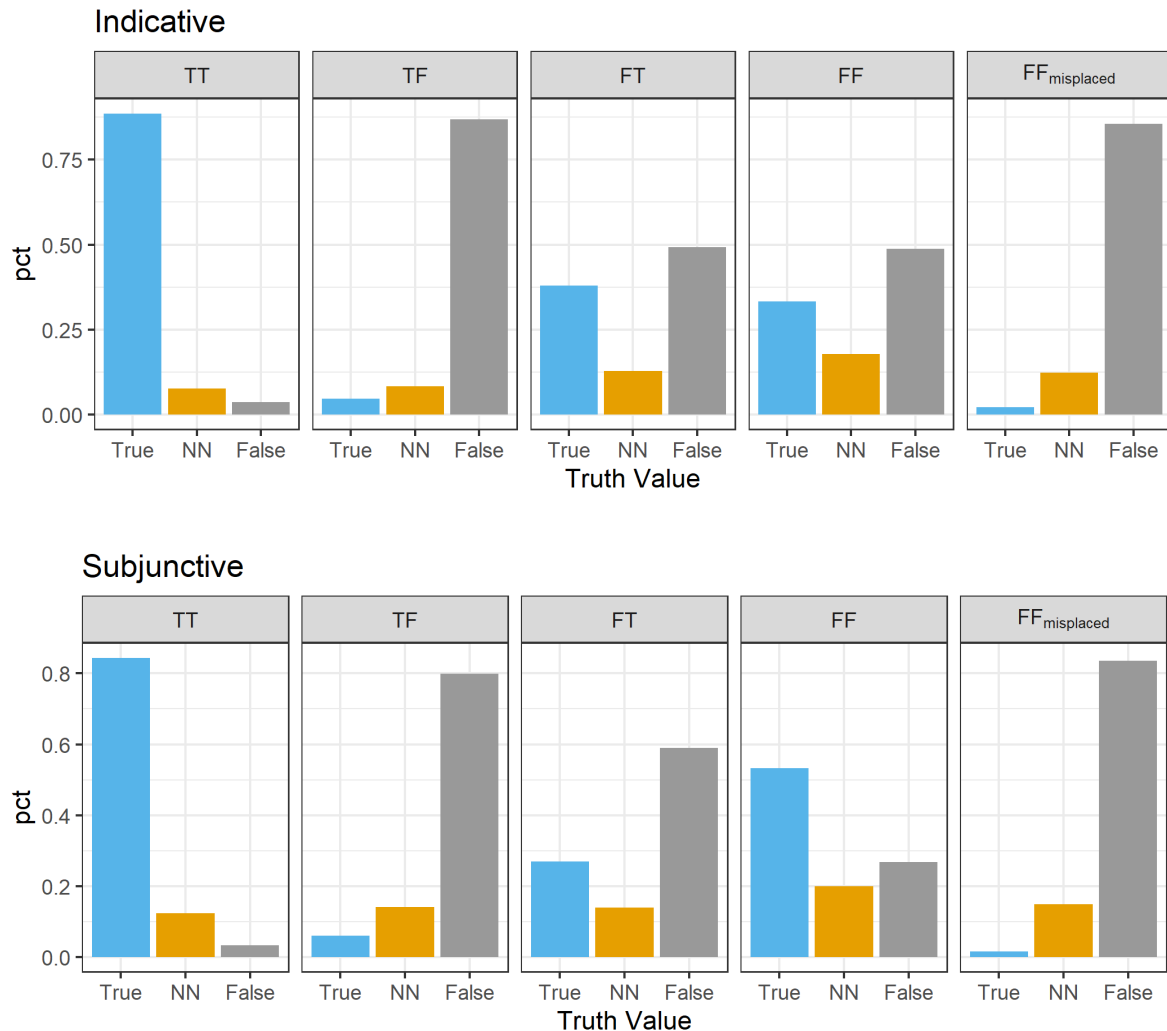Participants' truth evaluations are displayed in Figure 2:

Figure 2. Truth Tables of Indicative and Subjunctive Conditionals. 'FF' = a conditional that is False False w.r.t. the Actual Picture shown but True True w.r.t. the salient Hypothetical Picture; 'FF$_{misplaced}$' = a conditional that is False False w.r.t. the Actual Picture shown but True False w.r.t. the salient Hypothetical Picture.

As seen, the truth evaluations in Figure 2 closely replicate the truth evaluations in Figure 1 of Experiment 1. It was found across conditions that the average conditional probabilities were .08 (SD = .11), .94 (SD = .09), and .04 (SD = .09), for the Actual Object, Hypothetical Object, and Misplaced Object to be on the picture under the supposition that the picture displayed was the Hypothetical Picture, when performing the Ramsey test. Accordingly, Experiment 2 also supports assigning the truth-values of Table 3 to the Jeffrey table.

Like in Experiment 1, the truth evaluations for subjunctive conditionals show a clear pattern in modal values with no bi-modal tendencies, and the truth evaluations for indicative

conditionals show a mixed pattern for the FT and FF cells. In Appendix B, we show that we were also able to replicate the model fitting results of Experiment 1 in Experiment 2.

Accommodation

To investigate whether the mixed results for indicative conditionals are predicted by participants' subsequent responses to the accommodation task, Figure 3 partitions participants' truth evaluations of indicatives by their modal response to the first accommodation task.
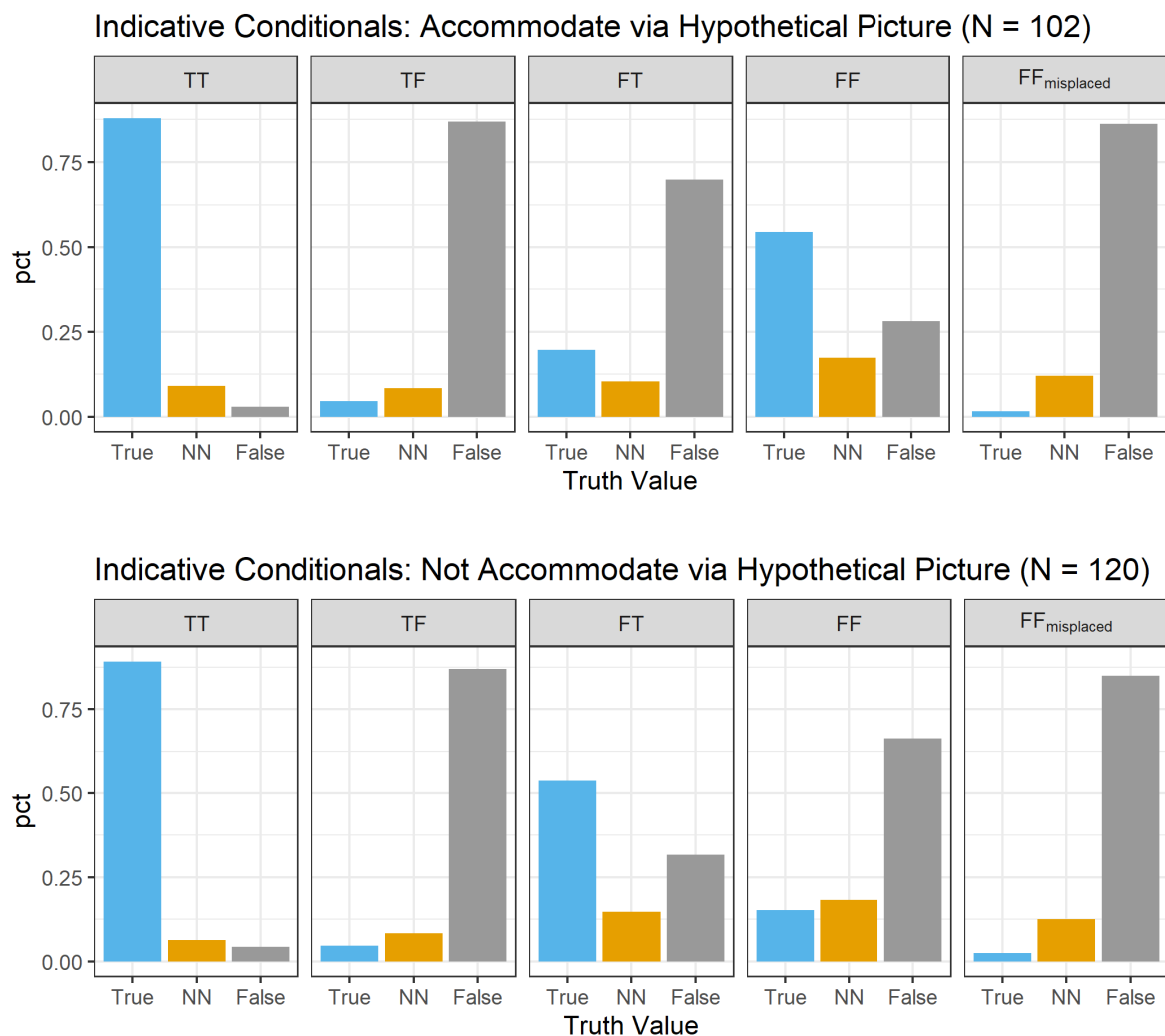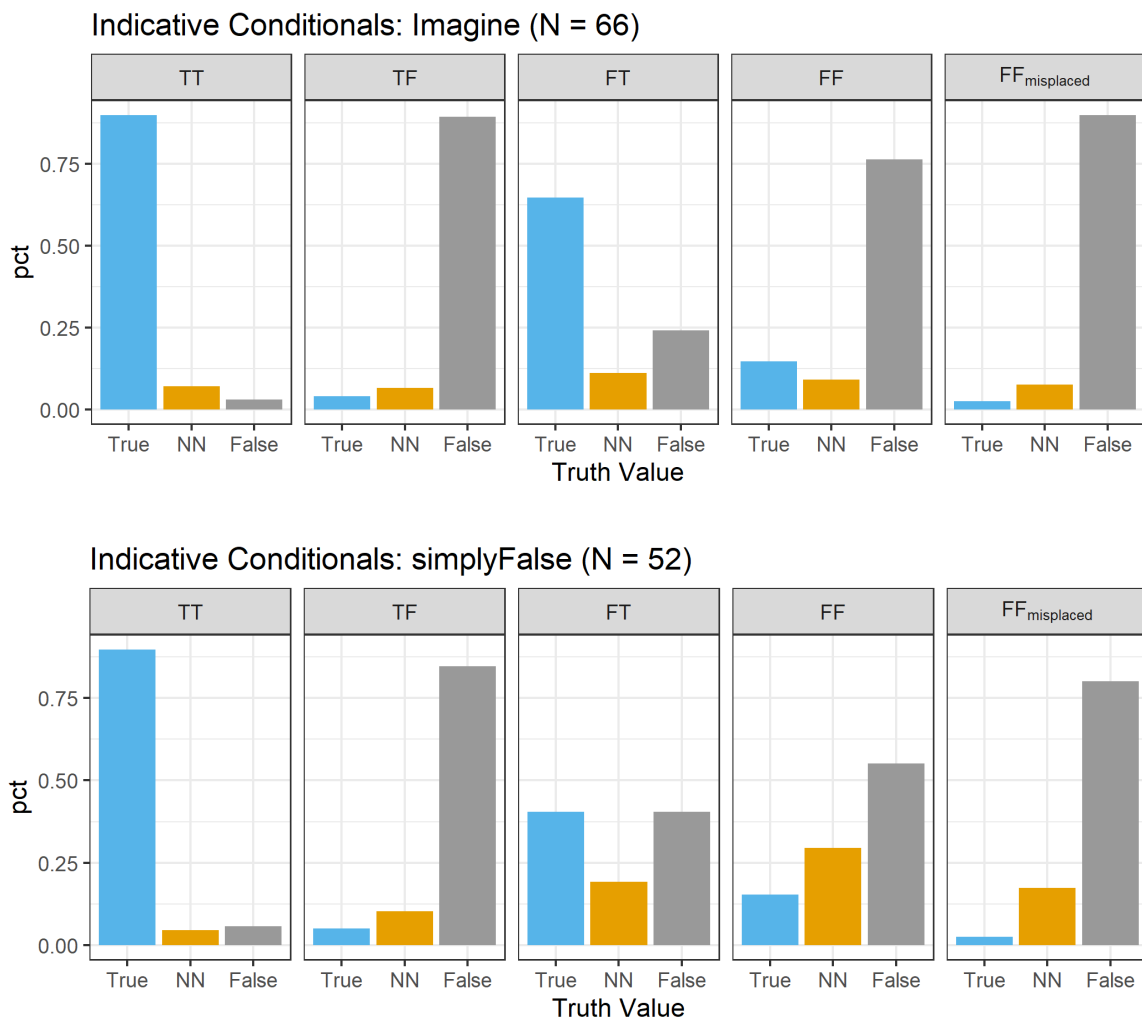

Figure 3. Truth Tables of Indicative and Subjunctive Conditionals. 'FF' = a conditional that is False False w.r.t. the Actual Picture shown but True True w.r.t. the salient Hypothetical Picture; 'FF$_{misplaced}$' = a conditional that is False False w.r.t. the Actual Picture shown but True False w.r.t. the salient Hypothetical Picture.

Again, the results closely resemble those from Experiment 1, which we report in Appendix A.
Participants (N = 102), who interpret the indicative conditionals with false antecedents as
referring to the Hypothetical Picture not shown all three times when asked, show the same
modal truth evaluation as with subjunctive conditionals. For the remaining participants (N =
120), the modal truth evaluation follows the evaluation of the consequent w.r.t. the Actual
Picture. To further investigate the latter group's truth evaluation, Figure 4 partitions their truth
evaluations of indicative conditionals as a function of how these participants responded to the
follow-up accommodation question. The outcomes displayed are based on the 118
participants (98%) who showed a preference for either the imagine or the simply false
strategy, when presented with the follow-up question.

Figure 4. Truth Tables of Indicative and Subjunctive Conditionals. 'FF' = a conditional that is False False w.r.t. the Actual Picture shown but True True w.r.t. the salient Hypothetical Picture; 'FF$_{\text{misplaced}}$' = a conditional that is False False w.r.t. the Actual Picture shown but True False w.r.t. the salient Hypothetical Picture.

As Figure 4 shows, participants were almost evenly divided between which of the two follow-up options to endorse. As the results show, the 55% selecting the imaging option were mainly responsible for the evaluation of the indicative conditional based on the truth value of the consequent when assessed based on the Actual picture shown. In contrast, of the 43% participants selecting the 'Simply False' option, the following response patterns shown in Table 7 were obtained.

Table 7. Response Patterns for Participants selecting Simply False

Pattern	T,F,F,F,F	T,F,T,F,F	T,F,NN,NN,NN	Unclassified
Proportion of 'simply false' (N = 45)	28.85% (15)	23.08% (12)	7.69% (4)	40.39% (21)
Proportion of whole sample (N = 222)	6.76%	5.41%	1.80%	9.46%

Note. Classification of participants who both selected that indicative conditionals with a false antecedent referred to the Actual Picture shown in the majority of cases and who responded that they followed the 'simply false' response strategy in the follow-up question. 'NN' = Neither True nor False.

As shown, less than a third of those selecting the 'simply false' option had actually produced truth-value assignments to indicative conditionals that matched this strategy. Almost the same number of participants produced the truth table pattern corresponding to the imagine option. 7.69% produced a pattern resembling the de Finetti table with 'neither true nor false' response for false antecedent cells and more than a third displayed a mixed pattern that could not be assigned to any of the above. These proportions of truth-value assignments within the subset of participants selecting the 'simply false' options differ significantly from a random distribution of the four classes $\chi^2(3) = 11.54$, $p = .009$, but they do not correspond to the predicted modal pattern of uniformly responding false in false antecedent cells.

Discussion

In Experiment 2, the results of Experiment 1 were replicated closely concerning modal truth-value assignments to indicative and subjunctive conditionals, evaluations of control items, assignments of conditional probabilities, and truth-value assignments to indicative conditionals as a function of performance on the accommodation task. In Appendix B, we showed that all the model fitting results of Experiment 1 also replicated.

To investigate the responses by participants, who did not accommodate in the false antecedent cells via the Hypothetical Picture, a follow-up question was introduced in Experiment 2 outlining two strategies. According to the first strategy, participants who indicate that indicative conditionals with false antecedents refer to the Actual Picture shown carry out a different type of accommodation by imagining that the antecedent is satisfied for the Actual Picture and therefore use it as a basis for the truth evaluation of indicative conditionals. According to the second strategy, participants who indicate that indicative conditionals refer to the Actual Picture shown despite the false antecedent invariably produce a 'False' response as a reaction to the falsity of the antecedent.

The results indicate that participants were split in which strategy they reported in the follow-up question. A close examination of the truth-evaluations as a function of the selected strategy, however, revealed that only the group that selected the imagining option had previously produced a modal truth-value assignment corresponding to their chosen strategy. For the participants selecting the 'simply false' strategy, less than a third had actually produced truth-values that would match this strategy. This proportion was only slightly bigger than the proportion of participants who displayed the pattern of the imaging response despite selecting the 'simply false' option. As a result, it can be concluded that the 'simply false' hypothesis does not consistently account for the responses of the participants who indicate that they evaluate indicative conditionals with false antecedents with respect to the Actual Picture shown. Indeed, Figure 3 shows that the 'simply false' response strategy is inconsistent

with the modal pattern in this subset of the participants' responses, and it was found that out of the whole sample (N = 222) only 6.76 % of the participants both selected the 'simply false' option and produced a truth-value assignment that matched it. On the contrary, the modal pattern in Figure 3 (lower panel) did converge with the imagining strategy, which a little more than half the participants also indicated that they followed.

We can therefore conclude that although participants do not show perfect reliability in identifying the strategy they followed in assigning truth-values to indicative conditionals with false antecedents, there is a central tendency to accommodate the antecedent by imagining that the Actual Picture satisfies the descriptive characterization, which supports the application of the possible worlds model to account for these responses. Nevertheless, our results did show evidence of individual variation, which is the topic we turn to next.

Experiment 3

Experiment 1 indicated that the possible worlds account was the only one of the investigated theories that did not misfit the data at the aggregate level for indicative conditionals. At the same time, the posterior predictive checks indicated that none of the investigated theories were capable of accounting for the variances and covariances across individuals. For this reason, Experiment 3 applied a Bayesian mixture distribution analysis to investigate individual variation in participants' truth evaluations.

To increase the number of trial replications per participants, only indicative conditionals were investigated in Experiment 3. In addition, Experiment 3 also measured whether participants viewed the antecedent as a reason for or against the consequent. This dependent variable was included to investigate whether making reason relations, or inferential relations, part of the truth conditions would help account for participants' responses, as posited by truth-conditional inferentialism (Douven et al., 2018).

Participants

288 people participated in the study. The final sample after applying the *a priori* exclusion criteria from Experiment 1 consisted of 155 participants. Mean age was 41.88 years, ranging from 20 to 73. 41.29% of the participants identified as male; the rest identified themselves as female. 78.71 % indicated that the highest level of education that they had completed was an undergraduate degree or higher. Applying the exclusion criteria had a minimal effect on the demographic variables.

Design

The experiment had a within-subject design with the following within-subject factors: Truth Table Cell (TT, TF, FT, FF, FF$_{misplaced}$). Since the experiment had six trial replications of indicative conditionals, participants saw a total of 30 within-subject conditions.

Procedure

Experiment 3 followed the procedure of Experiment 1 with one exception. In addition to the initial control questions concerning the truth/falsity of claims stating that three objects were on the pictures, and the conditional probability questions for the Jeffrey table, Experiment 2 included questions concerning reason relations. To illustrate, in one condition participants were asked to evaluate whether the statement "The picture shown is one of Jack's pictures of a study" is a reason for/against the statement "There is a shampoo in it"[/"There is a study lamp in it"/"There is a pair of bed lamps in it"]. Participants were provided with a five point labelled Likert scale {a strong reason against; a reason against; neither for nor against; a reason for; a strong reason for} to give their responses.

With this addition to the procedure of Experiment 1, the dependent variables for Experiment 2 were: (1) ternary truth evaluations, (2) conditional probabilities, (3), ordinal reason relation assessments, (4) three accommodation questions, (5) control questions ensuring that the participants understood the truth table cell by correctly identifying the presence/absence of named objects w.r.t. the displayed pictures.

Results

As shown in Figure 5, participants' ternary truth value judgments of indicative

conditionals roughly replicated those found in Experiments 1 and 2.
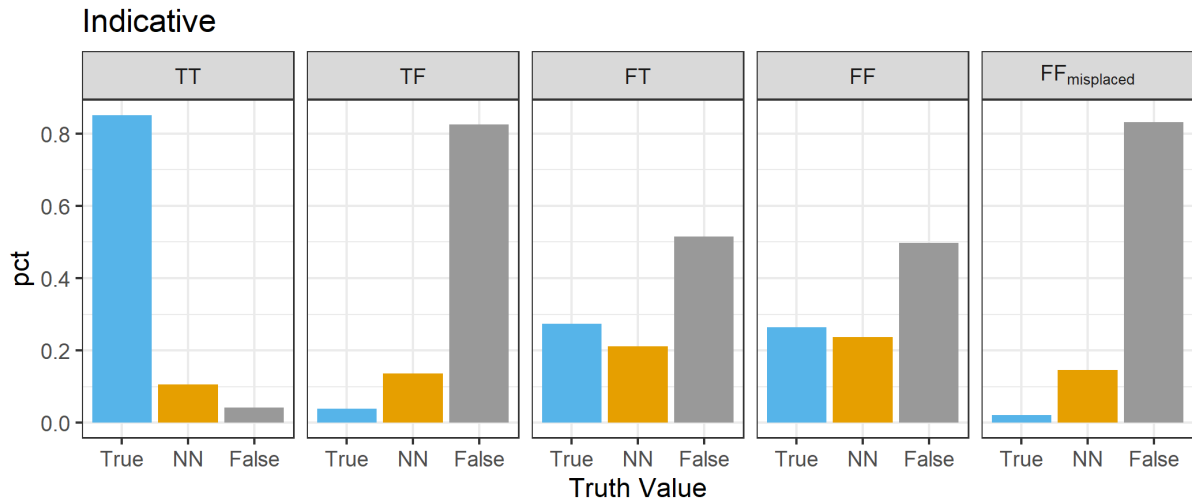
Indicative

Figure 5. Truth Tables of Indicative. 'FF' = a conditional that is False False w.r.t. the Actual Picture shown but True True w.r.t. the salient Hypothetical Picture; 'FF$_{misplaced}$' = a conditional that is False False w.r.t. the Actual Picture shown but True False w.r.t. the salient Hypothetical Picture.

According to truth-conditional inferentialism advanced in Douven et al. (2018), the indicative

conditional is true if the consequent can be inferred from the antecedent (possibly via

background assumptions) and false if it cannot be inferred. To be able to test this theory in the

present paradigm, participants were asked for the extent to which the antecedent provided a

reason for/against the consequent on a five-point Likert scale. For each pair of pictures, these

reason relation assessments were made by having the antecedent describe one of the pictures

of the pair and having the consequent either describing a matching object (e.g., a fruit bowl in

a picture of kitchen) or a mis-matching object (i.e., the warning sign of the hypothetical

picture or the misplaced object, e.g., a pair of bed lamps). The results are shown below
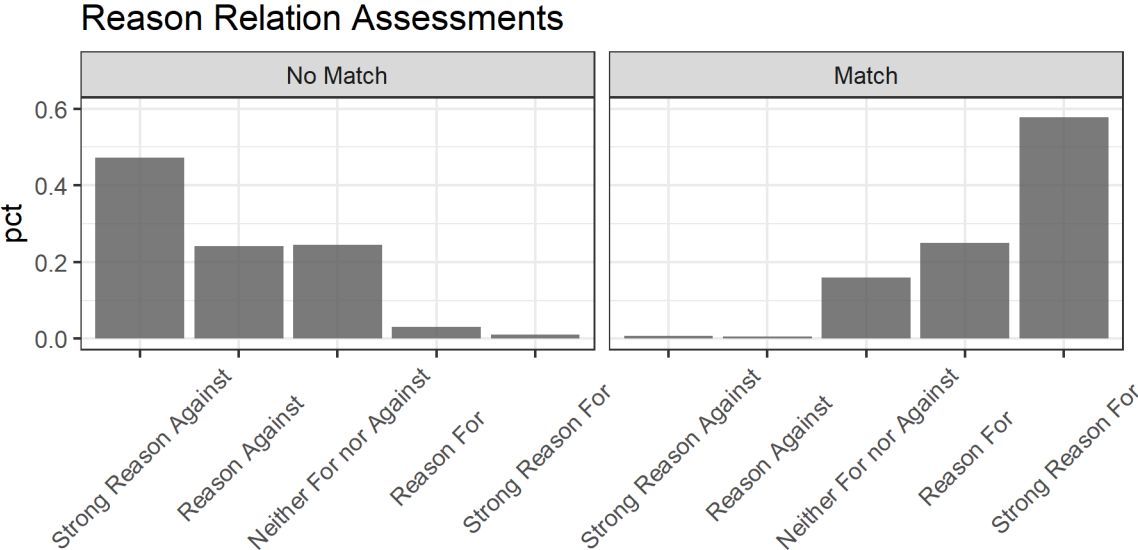
(Figure 6).

Reason Relation Assessments

Figure 6. Reason relation assessments on a five-point Likert scale. The TT and FF cell correspond to the 'matching cases', where an object from the Actual/Hypothetical Picture was mentioned in the consequent and the corresponding Actual/Hypothetical Picture was mentioned in the antecedent. The TF, FT, and FF_Misplaced cells correspond to the mismatching cases, where objects in the consequents were mentioned that violated the expectations concerning the pictures mentioned in the antecedent sentences.

The TT and FF cells correspond to the matching cases and the TF, FT, and FF_misplaced cells correspond to the mismatching cases. Accordingly, truth-conditional inferentialism makes the same predictions as M_Subjunctive in Table 3 for our experimental paradigm.

Like in Experiments 1 and 2, conditional probabilities were measured to apply the Jeffrey table which holds that participants assess conditional probabilities in the false antecedent cells. When these conditional probabilities are high, participants are predicted to treat the conditional as true and when the conditional probabilities are low, participants are predicted to treat the conditional as false, based on the auxiliary assumptions of Over (2020) and Over and Cruz (2018). Aside from a few outliers, a clear trend was recognizable when participants were asked to estimate conditional probabilities via the Ramsey test. It was found across conditions that the average conditional probabilities were .12 (SD = . 13), .92 (SD = .09), and .07 (SD = .10), for the Actual Object, Hypothetical Object, and Misplaced Object to

be on the picture under the supposition that the picture displayed was the Hypothetical Picture. This pattern thus gives rise to the same predictions as $M_{Subjunctive}$ in Table 3.[11]

Four Bayesian Mixture models were fitted in JAGS (Plummer, 2019) with a categorical variable deciding which of the mixture components an individual participant was assigned to. Three of these models included correlations among the MPT parameters in the hierarchical structure, following the latent trait model of Klauer (2010) applied in Experiment 1. The fourth model assumed that there were no correlations among these MPT parameters across participants, following the beta-MPT approach of Smith and Batchelder (2010, see e.g., Heck et al. 2018). Further technical details can be found in Appendix C.

Due to the partial overlap in the predictions of the truth tables, $M_{PossibleWorld}$ and $M_{Subjunctive}$, we fitted several versions of the latent trait model:

M1: Latent Trait model with three mixture components ($M_{Material}$, $M_{deFinetti}$, $M_{PossibleWorld}$).

M2: Latent Trait model with four mixture components ($M_{Material}$, $M_{deFinetti}$, $M_{PossibleWorld}$, $M_{Subjunctive}$).

M3: Latent Trait model with five mixture components ($M_{Material}$, $M_{deFinetti}$, $M_{PossibleWorld}$, $M_{Subjunctive}$, $M_{conjunction}$).

M4: Beta-MPT model with four mixture components ($M_{Material}$, $M_{deFinetti}$, $M_{PossibleWorld}$, $M_{Subjunctive}$).

In addition, all models featured a saturated mixture component to filter out noisy respondents, who did not fit any of the models in virtue of violating the shared predictions of 'True' and 'False' in the TT and TF cells.

---

[11]   Accordingly, for our experimental paradigm, this model encompasses truth-conditional inferentialism, the Jeffrey Table, MMT_{aux}, and Bradley's (2012) truth conditions. As we recall, the name '$M_{subjunctive}$' was chosen for this model, because the indexed truth tables make the same predictions for indicative conditionals in this specific task as the modal values of the subjunctive conditionals in Experiment 1.

The fifth mixture component in M3 was added, because previous research on individual variation with conditionals have shown that participants sometimes produce conjunctive responses (Evans et al., 2007). The fourth and final model was a beta-MPT version of M2, which differed from M2 by assuming that there were no correlations among these MPT parameters across participants in their *a priori* distributions (Smith & Batchelder, 2010). We then quantified the respective predictive performance of the four models by the leave-one-out cross validation criterion and WAIC (Table 8).

Table 8. Model Comparison

	LOOIC	Δelpd	SE	WAIC	Weight	p_{T1}	p_{T2}
M3: Latent Trait 5	2544.30	0	0	2219.53	1.00	0.26	0.007
M1: Latent Trait 3	2811.81	-133.76	22.77	2510.00	0.00	5e-06	1e-05
M2: Latent Trait 4	2836.13	-145.91	22.26	2600.22	0.00	0	0
M4: Beta-MPT 4	3436.42	-446.06	34.83	3324.74	0.00	0	0

Note. LOOIC = leave-one-out cross-validation information criterion. WAIC = Watanabe-Akaike information criterion. Weight = model averaging via staking of the predictive distributions. 'elpd' = expected log predictive density is a measure of the expected out-of-sample predictive accuracy. The test statistics T_1 and T_2 represent Bayesian p values and are based on the posterior predictive model checks in Klauer (2010). 'Beta-MPT 4' refers to the beta-MPT approach of Smith and Batchelder (2010) with 4 mixture components. 'Latent Trait 4' refers to the hierarchical latent trait approach of Klauer (2010) with 4 mixture components. The two models differ on whether they permit correlations in the MPT parameters across participants.

The information criteria show a preference for the latent trait model with correlated MPT parameters across participants over the Beta-MPT model without these correlations. In addition, they indicate that the latent trait model with five mixture groups (M3) is preferred in light of the fit and parsimony trade-off. In addition, M3 was also the only model that showed a satisfactory fit for the aggregate truth value frequencies (T_1).

As Figure 7 shows, in general participants classified according to these five mixture groups had a relatively high posterior probability of agreeing in their truth value judgments with their assigned truth table.
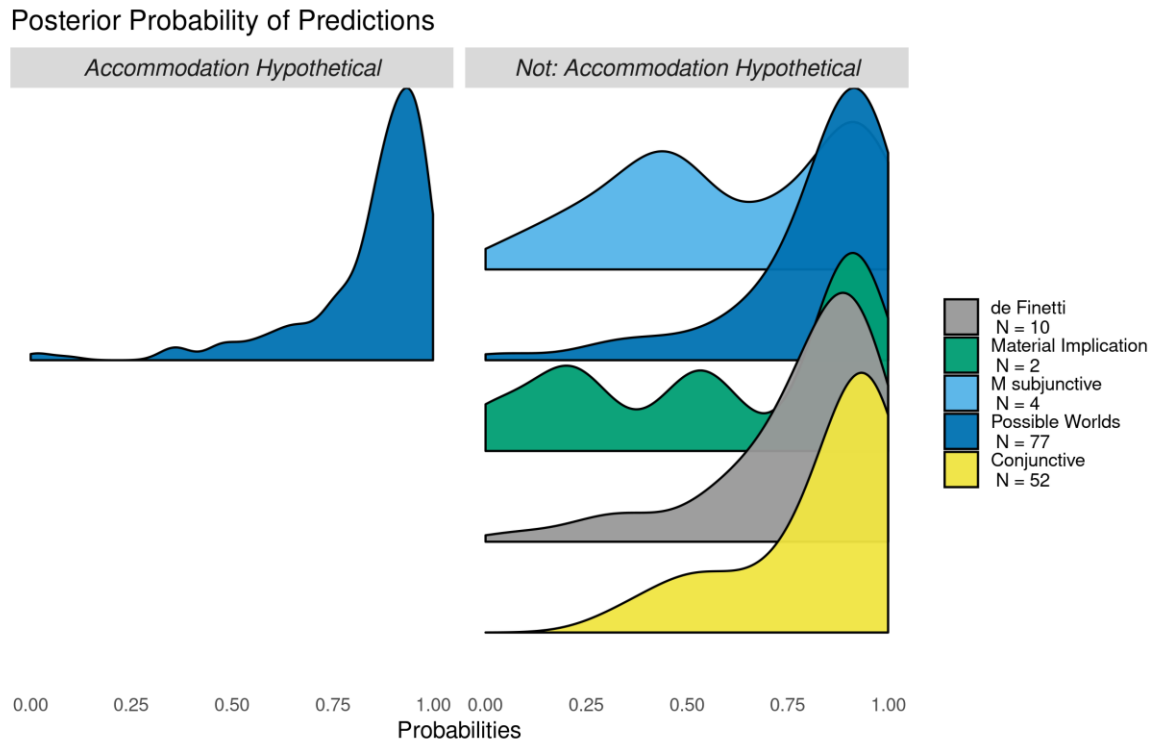
Figure 7. Posterior probability of the predictions of participants' truth value judgments agreeing with their predicted truth tables, for each of the five mixture groups of M3. 10 participants were captured by the saturated model used to filter out participants, who did not conform to the shared prediction of 'true' and 'false' in the TT and TF cells. The left and right hand side of the plot is separated by whether participants indicated that they accommodated the false antecedent to refer to the Hypothetical Picture in the accommodation task.

As the comparison shows, the largest group of participants were assigned to the possible worlds model. But individual variation was also found and so the model had to be supplemented with the other truth tables.

Discussion

As Figure 5 shows, Experiment 3 roughly replicates the qualitative patterns of truth value judgments of indicatives from Experiments 1 and 2. Of the different models investigated in Experiment 1, $M_{Subjunctive}$ makes the same predictions for indicative conditionals as the modal truth value assignment for subjunctives in Experiment 1. It was shown in Table 3 that this model characterized several competing theories for our experimental paradigm. Based on the measured conditional probability judgments, it was found, like in Experiment 1, that $M_{Subjunctive}$ also captures the predictions of the Jeffrey truth

table. In addition, based on the ordinal reason relation judgments of the participants, it was found in Experiment 3 that $M_{Subjunctive}$ also captures the predictions of truth-conditional inferentialism for our task.

In Experiment 1, $M_{PossibleWorld}$ was the only model not misfitting participants' responses at the group level (which is a finding that we replicated for Experiment 2 in Appendix B). In contrast, Experiment 3 sought to investigate via mixture distributions whether subgroups of participants could be identified that followed competing truth tables.

Since, however, $M_{PossibleWorld}$ and $M_{Subjunctive}$ have overlapping predictions for the subset of participants who accommodated the antecedent, we made a model comparison between four models (M1, M2, M3, M4). It was found that a latent trait model (M3) that included five mixture groups ($M_{Material}$, $M_{deFinetti}$, $M_{PossibleWorld}$, $M_{Subjunctive}$, $M_{conjunction}$) performed the best in light of the fit vs. parsimony trade-off. In this model, it was found that overall participants had a high posterior probability of following the assigned truth table in their truth value judgments, across mixture groups (Figure 7). In agreement with Experiment 1, it was found that the largest subgroup of participants could be captured by $M_{PossibleWorld}$ at the individual level. In contrast, very few participants were captured by $M_{Subjunctive}$.

Like previous studies investigating individual variation (Evans, et al., 2007), it was found that there was a sizable minority of participants who produced a conjunctive pattern, which would coincide with the 'simply false' strategy from Experiment 2. In Experiment 3, this pattern was stronger than in Experiments 1 and 2, and it is thus possible that the within-subject comparison with subjunctive conditionals in Experiments 1 and 2 suppressed this response tendency. A further contributing factor may have been that none of the experiments reported in this paper capitalized 'if then' as 'IF…, THEN…' when presenting the conditionals to participants, as is common in reasoning studies. It is likely that either of these

factors would have contributed to sensitize participants to a genuine conditional reading of the presented sentences, and suppressed a conjunctive reading.

General Discussion

Through our experiments, we have compared the possible worlds account to more than five competing truth tables and found that it could survive the competition.

In Experiment 1, we investigated both indicative and subjunctive conditionals and found that the possible worlds account was the only one of the investigated theories that did not misfit the data at the aggregate level for indicative conditionals. In Experiment 2, we replicated the central tendencies in these results for both subjunctive and indicative conditionals and ruled out an alternative hypothesis concerning the participants, who did not accommodate in Experiment 1 by reassigning the reference of indicative conditionals with false antecedents to a paired Hypothetical Picture not shown. In Experiment 3, we modeled participants' truth evaluations of indicative conditionals as a mixture distribution of competing truth tables and found that the possible worlds semantics accounted for the largest subgroup in the analysis of individual variation in our experimental paradigm.

Collectively, these studies obtained evidence in favor of possible worlds semantics over a wide range of popular alternatives. That indicative and subjunctive conditionals differ in content is illustrated by the following famous example (Adams, 1970), where most will accept the first indicative conditional while rejecting the second subjunctive:

(*indicative*)   If Oswald did not shoot Kennedy, someone else did.

(*counterfactual*)   If Oswald had not shot Kennedy, someone else would have.

The formulation of this minimal pair illustrating the indicative/subjunctive divide, has led to numerous attempts to either provide a unifying account (Stalnaker, 1975; Edgington, 2008; von Fintel, 2012; Spohn, 2013; Starr, 2014; Williamson, 2020), argue why disjunct accounts

are needed (Lewis, 1973, 1976; Bennett, 2003), or argue for a unifying account by questioning that this indeed constitutes a minimal pair (Quelhas et al., 2018).

For proponents of the unifying account, it is tempting to formulate one semantics of conditionals, like possible worlds semantics, and look to linguistic phenomena closer to pragmatics to account for the differences in the sentences. For instance, in Skovgaard-Olsen and Collins (2021), evidence was found that at least the difference in the status of epistemic openness towards the antecedent of the indicative and disbelief towards the antecedent of the subjunctive could be attributed to the pragmatic phenomenon of conversational implicatures.

Below we are going to show that the strategy of Stalnaker (1968, 1974) of formulating one type of truth conditions in an abstract form that can be shared by indicatives and subjunctives, while supplying a pragmatic principle that accounts for their differences, is a fruitful strategy for accounting for our experimental data on truth evaluations. To make this argument, the discussion below starts out by considering how our results bear on the non-truth functionality of conditionals and Stalnaker's (1975) claim that indicative and subjunctive conditionals have truth conditions that share the same abstract form, but which diverge by applying his pragmatic principle. Next, we compare this approach with mental model theory as applied to our experiments. Finally, we contrast our own approach to making possible worlds semantics empirically testable with some alternative strategies.

Non-Truth Functionality

In Experiment 1, we presented a novel experimental task for probing the possible worlds semantics of Stalnaker (1968) and Lewis (1973). The idea was to experimentally constrain assessments of closest possible worlds by offering participants a prototype of what the closest possible-*A* world could look like (e.g., "one of Jack's photos of a railroad station"). We did this to probe whether participants selectively made use of this additional source of information, rather than the displayed image (e.g., "one of Jack's photos of a kitchen"), when

assigning truth values to subjunctive conditionals. The results supported the constraint for a unifying account of indicative and subjunctive conditionals that their truth evaluations differ for conditionals with false antecedents.

An important aspect of the possible worlds account is the non-truth functionality of indicative and subjunctive conditionals (Edgington, 2006). On this account, the truth of a conditional is not only a function of the truth value of the antecedent and the consequent; rather it depends on what is true in the most similar antecedent situation.

Through the introduction of two types of FF cells with respect to the Actual Picture shown (the FF cells with the Hypothetical Objects and the FF$_{misplaced}$ with the Misplaced Objects), Experiments 1 and 2 were designed to test for non-truth functionality. Given the large differences between the FF and FF$_{misplaced}$ cells for both subjunctive and indicative conditionals, this feature of the possible worlds semantics is corroborated by the results. A constraint for a unifying account of indicative and subjunctive conditionals is therefore that our results suggest that they are both marked by the property of non-truth functionality. More generally, it was found in Experiments 1 and 2 that there was a clear tendency to not evaluate subjunctive conditionals truth functionally based on the Actual Picture shown (e.g., Jack's picture of a kitchen). Instead, most participants consider a hypothetical situation in which the Hypothetical Picture is displayed (e.g., Jack's picture of a railroad station), when the antecedent is false, and evaluate the subjunctive conditional based on it instead.

For indicative conditionals, participants were split between evaluating conditionals with false antecedents based on the Actual Picture displayed (T, F, T, F, F) or on the Hypothetical Picture (T, F, F, T, F), which the majority use for evaluating subjunctive conditionals (see Appendix A and Figure 3). In this task, there is a tension between the demonstrative reference to the Actual Picture shown (e.g., Jack's picture of a kitchen) and a description of the Hypothetical Picture (e.g., Jack's picture of a railroad station) in the false

antecedents. As a result, participants were found to oscillate between the truth evaluations based on the Actual Picture and the Hypothetical Picture.

One way of interpreting this oscillation is based on the pragmatic principle of Stalnaker (1975). The principle constrains the selection function to consider possible situations within the context set for indicative conditionals while permitting the selection of possibilities outside the context set for subjunctive conditionals. The picture that is shown is clearly part of the context set. Yet, the indicative conditionals with the false antecedents violate Stalnaker's (1975) pragmatic principle when describing a hypothetical picture (e.g., Jack's picture of a railroad station) while demonstratively referring to the picture shown. Participants react differently to this: some are found to accommodate via the Hypothetical Picture and take the indicative conditional as strictly referring to it (despite the demonstrative reference); others evaluate the conditional based on the Actual Picture and accommodate the antecedent by imagining changes to the Actual Picture, as shown in Experiment 2.

Experiments 1 and 2 thereby find evidence of different strategies for dealing with indicative conditionals with false antecedents. On the aggregate level, a bi-modal pattern emerges which is not compatible with any of the other truth tables considered in Table 3. For subjunctive conditionals with false antecedents, in contrast, participants evaluate the conditionals based on the Hypothetical Picture, in accordance with Stalnaker's (1975) pragmatic principle, which permits them to consider possibilities outside the context set.

On Stalnaker's (1975) theory, the semantics specifies the *form* of truth conditions of conditionals by referring to selection functions (i.e. '$A > C$' is true in w, iff C is true in the possible world/s returned by f(A, w)). *Which* proposition is expressed by the assertion of a conditional is, however, influenced by the context of utterance, given the pragmatic rule that f(A, w) outputs possible worlds *within* the context set for indicative conditionals, and that this selection function can output possible worlds *outside* the context set for subjunctive

conditionals. Our findings support this strategy of using the *same abstract form* of the truth conditions for indicative and subjunctive conditionals but allowing their truth conditions to diverge through the application of a pragmatic principle that constrains the selection of the most similar situations.

In connecting the results on truth value assignments of Experiments 1 and 2 with the finding in Skovgaard-Olsen and Collins (2021) that the falsity of the antecedent expressed by a subjunctive conditional is a conversational implicature, the following observations can be made. If use of the subjunctive mood, and the fake past tense of subjunctive conditionals, generates the conversational implicature that the antecedent is false, then this has implications for the truth evaluations of subjunctive conditionals, on a possible worlds account. The conversational implicature that the antecedent is false warrants interlocutors to use possible worlds outside the common ground (in our case: the Hypothetical Picture) in assigning truth values to the conditional. The conversational implicature is, however, cancellable, which means that the search for the closest possible *A*-worlds outside the context-set can be overridden. One case in point is the famous example of so-called Anderson conditionals:

(4)   "If Jones had taken arsenic, he would have shown exactly those symptoms which he does in fact show" (Anderson, 1951, p. 37).

Since a speaker of this conditional could use (4) to argue that Jones had, in fact, taken arsenic, the conversational implicature of the subjunctive conditional that Jones did not take arsenic is cancelled in this case (von Fintel, 1997, 2012; Stalnaker, 1975, 2014).

Aside from Anderson conditionals, this type of cancellation takes place when participants are asked to evaluate subjunctive conditionals in TT and TF cells, where they need not go beyond the possible world offered by the displayed picture to identify the closest possible *A*-world. Through the constraint that the similarity relation is *centered*, so that the

closest possible A-world to w is w itself, whenever A is true at w, this requirement is built into possible worlds semantics by fiat (Stalnaker, 1968; Lewis, 1973).

In the discussion sections of the individual experiments, we have already considered in detail what bearing our results have on a wide range of different accounts of indicative conditionals (e.g., the material implication, de Finetti Table, the Jeffrey Table, Bradley's (2012) account). By measuring participants' reason relation assessments, Experiment 3 added a further account to this list; truth-conditional inferentialism (Douven et al., 2018), which holds that inferential relations from the antecedent to the conclusion are part of indicative conditionals' truth conditions. Skovgaard-Olsen et al. (2017) already provided evidence against this account using a different experimental paradigm. In Experiment 3 we conducted a new test of truth-conditional inferentialism by showing that it is not the case that our task implicitly manipulates inferential relations as a confound that could account for participants' truth value assignments in this task. Below we consider whether mental model theory would be able to account for our results concerning subjunctive conditionals in Experiments 1 and 2.

Mental Model Theory and Possible Worlds Semantics

On the revised mental model theory, conditionals are conjunctive assertions about possibilities (i.e., "$A\&C$ is possible and $A\&\neg C$ is not possible..."). That not-A is possible is a shared presupposition of true and false conditionals. In the case of counterfactual conditionals, the "$\neg A\&\neg C$" possibility acquires the status of being *a fact* and the other possibilities change status to express "counterfactual possibilities" (see Table 4). Counterfactual possibilities concern states that were once possible but did not obtain.

It is a well-known observation that participants often exhibit a biconditional interpretation of conditionals (see e.g., Goodwin & Johnson-Laird, 2018). If we consider only the standard truth table cells in Experiments 1 and 2 (TT, TF, FT, FF), then participants are found to interpret subjunctive conditionals bi-conditionally (T, F, F, T). The addition of the

FF$_{\text{misplaced}}$ cell demonstrates, however, that it is not really a bi-conditional interpretation that is found, because the modal truth value makes a strong flip from True to False in what is an FF cell, when evaluated w.r.t. the Actual picture shown (see Figures 1 and 2).

Could the mental model account handle the results for subjunctive conditionals of Experiments 1 and 2? To be able to account for the full range of conditions (TT, TF, FT, FF, FF$_{\text{misplaced}}$), mental model theory would have to apply its possibility table (see Table 4) for both indicative and counterfactual conditionals. What enforces this is the constraint of counterfactuals that "$\neg A \& \neg C$ is a fact", which is only met in the FF and FF$_{\text{misplaced}}$ cells. That is to say, the subjunctive conditionals in the TT, TF, and FT cells are effectively treated as indicative conditionals, if the account is to be applicable. Under these assumptions, mental model theory is equipped to account for the modal pattern (T, F, F, T, F) in participants' responses, but problems emerge, as shown below.

A Technical Problem Concerning Subjunctives

The first problem is that mental model theory is forced to misrepresent the subjunctive conditional in the FT cell as making a (false) claim about a *real* possibility, instead of making a (false) claim about a *counterfactual* possibility (now that "$\neg A \& C$" is known as *a fact*). But treating the subjunctive conditional in the FT cell as an indicative conditional is problematic for other reasons as well. As we have seen, Goodwin and Johnson-Laird (2018) suggest that indicative conditionals are true in the false antecedent cases if the corresponding counterfactual conditionals are true. However, if we cannot evaluate the subjunctive conditional in the FT cell as a counterfactual on mental model theory (because "$\neg A \& \neg C$" is *not* a fact), and we can only evaluate it as an indicative, if we already know the truth value of the corresponding counterfactual, then we have landed in a circle. It may thus prove difficult to derive a truth value in this case by following the proposed strategy.

In contrast, a selection function in possible worlds semantics can switch between the pictures more smoothly, and not misconstrue the modal status of the possibilities under evaluation, as shown in Table 9.

Table 9. Selection function applied to the subjunctives of Experiment 1

Cell	$f(w_1, A) = w_i$		w_1 = Actual P	Conditional	Response
TT	w_i = Actual P	TT	TT	Actual P > Actual object	T
TF	w_i = Actual P	TF	TF	Actual P > Hypothetical object	F
FT	w_i = Hypothetical P	TF	FT	Hypothetical P > Actual object	F
FF	w_i = Hypothetical P	TT	FF	Hypothetical P > Hypothetical object	T
FF$_{mis}$	w_i = Hypothetical P	TF	FF	Hypothetical P > Misplaced object	F

Note. "$A > B$" = "If A had been the case, then B would have been the case." 'Actual P' = the Actual Picture is displayed. 'Hypothetical P' = the Hypothetical Picture is displayed. 'FF$_{mis}$' = FF$_{misplaced}$. $f(w_1, A)$ selects the closest possible A-world. For the TT, TF conditions, this is: a situation where the Actual Picture is shown. For the FT, FF, FF$_{misplaced}$ conditions, this is: an imagined situation where the Hypothetical Picture is shown. The conditional is then evaluated w.r.t. the selected world.

As seen, the selection function correctly selects the Actual Picture as the world of evaluation in the TT and TF conditions, and the Hypothetical Picture as the world of evaluation in the FT, FF, and FF$_{misplaced}$ conditions. A further problem for the mental model theory is that, strictly speaking, it is *both possible* that the Hypothetical Picture (i.e., "one of Jack's photos of a railroad station") has a warning sign on it *and possible* that it does *not* have a warning sign on it. Yet, to capture the participants' responses in Experiment 1, the second combination would have to be treated as *impossible*. It is unclear why any of the combinations of Jack's pictures would merit this description under the account by mental model theory.

It is easy to read this as the implausible suggestion that participants treat it as *epistemically impossible* that there would be no warning sign on one of Jack's photos of a railroad station. Goodwin and Johnson-Laird (2018, p. 2528) are aware of the difficulty and thus caution that they do not intend impossibility to be taken in an absolute sense but rather as that the conditional in a specific context is incompatible with a state of affairs.

However, in addition to understanding that a given conditional is *incompatible* with a specific possibility, accepting a conditional as true requires that that possibility is set aside as not pertinent in a given context. It is at this point that possible worlds semantics appeals to

comparisons between which possibilities are more similar to the actual situation to allow for a *relative notion* of possibility.[12] Given the earlier encountered prototypes of Jack's photos, it would be more compatible with everything that is known to assume that one of Jack's photos of a railroad station would have had a warning sign on it than to assume that it would lack one. Most participants therefore converge on their truth evaluations of the subjunctive conditionals. But there would be nothing contradictory, or incoherent, about assuming that Jack could also have taken photos of railroad stations without warning signs. We therefore cannot exclude it across contexts, even if we treat it as a *more distant* possibility in this context. In contrast, while Goodwin and Johnson-Laird (2018) may stress that they do not mean 'impossibility' to be taken in an absolute sense, they have not, on the other hand, given the theoretical means for accounting for relative comparisons of possibility/impossibility.

Possible worlds semantics is thus better equipped to account for the results in Experiments 1 and 2 concerning subjunctive conditionals. Johnson-Laird and Ragni (2019) argue against the psychological plausibility of possible worlds semantics on the grounds that it would require reasoners to consider an infinite number of possible worlds. It is, therefore, important to stress that, in fact, participants need not consider an infinite number of possible worlds. On the contrary, one can apply the semantics more locally based on a small number of situational models considered by the participants, as we have done in Experiments 1 and 2. When modal operators, like conditionals, are given a *global interpretation*, conditionals are assigned truth values *at all possible worlds*. But in natural-language semantics, the set of possible worlds may be restricted so that sentences containing conditionals are only assigned truth values at a small set of worlds that are deemed contextually relevant (Garson, 2013, p.

---

[12]   Formally, modal logic introduces a relative notion of relevant possibilities by introducing accessibility relations that constraint which possible worlds can be accessed from a given world for a specific modal operator (Garson, 2013; Ch. 5, 20). For this, Lewis (1973) uses spheres of similarity and Stalnaker (1968) applies selection functions.

63, see further Portner, 2009). Formally, the set of possible worlds just needs to be non-empty (Garson, 2013, p. 93). Moreover, these situational models need not contain more structure than what is needed to address the current questions under discussion (Roberts, 1996; Stalnaker, 2019, p. 185, fn. 9).

Making Possible Worlds Semantics Testable

Above, we outline some technical problems that arise for the revised mental model theory in attempting to account for the results for subjunctive conditionals in Experiments 1 and 2. We take this, as well as the parallel finding for indicative conditionals (see Table 3), as evidence that possible worlds semantics specifies an input-output function that fit the participants' mean response tendencies responses better than the alternatives we have looked at. However, the failure of all tested theories to account for the variances and covariances in assigned truth values to indicative conditionals across individuals suggests that there is further individual variation that is left unaccounted for by assuming any given truth table for all participants. In Experiment 3, we investigated this possibility via mixture distributions based on the truth tables used in Experiment 1. It was found that while the model implementing possible worlds semantics did not account for all the participants, it accounted for the largest subgroup. In addition, minorities following both the material implication, the de Finetti truth table, and a conjunctive pattern could also be identified.

Given that there was a lower proportion of participants answering both SAT questions correctly in Experiment 3 than in Experiments 1 and 2, we suspect that there was a higher proportion of participants, who invested less effort in the task in Experiment 3. This in turn could explain why a higher proportion of conjunctive respondents were found in Experiment 3, since other studies have linked a conjunctive response pattern to low effort, low cognitive ability, and developmentally early interpretations of conditionals (Evans et al., 2007).

For the other interpretations, their replicability and stability would need to be further assessed. For instance, via the Scorekeeping Task and multiple test sessions used to study individual variation in participants' interpretations of conditionals in Skovgaard-Olsen et al. (2019). Provided that such evidence could be obtained, we are perfectly ready to embrace the possibility that different people come up with different interpretations and usages of conditionals that are precursors or laypersons' versions of what philosophers and psychologists have elaborated on in the different formal models. After all, these formal models must have an origin, one important source presumably being language practices observed by these theoreticians (as argued, e.g., in Skovgaard-Olsen, 2023, Ch. 1).

Before any conclusions can be reached about the psychological implementation, an important first step is to make the main theoretical alternatives *empirically testable* by designing new experimental tasks. Given its persistent popularity in other disciplines, in few other cases is this need arguably more urgent than for the possible worlds semantics of Stalnaker (1968) and Lewis (1973). It is to fill this gap between the types of theories considered in psychology and philosophy/linguistics that the present paper contributes.

Another way to accomplish this feat is to follow Wijnbergen-Huitink et al.'s (2015) strategy and apply notions of similarity from psychology to possible worlds semantics. For instance, Wijnbergen-Huitink et al. follow the approach to similarity of Tversky (1977) of counting the number of features that objects have in common. Relatedly, other similarity or distance measures such as the following could be considered for the cognitive representations of possibilities by participants (Table 10).

Table 10. Similarity and Distance Measures

Name	Formula	Comment
Stimulus generalization in configural learning (Pearce, 1987, 1994).	$_P S_{P'} = \dfrac{N_C^2}{N_P * N_{P'}}$	The stimulus generalization from the stimulus (P) AB to (P′) ABC is quantified based on their similarity, $_P S_{P'}$. In this case, P and P′ share two components ($N_C = 2$) and the number of elements of the stimuli is 2 ($N_P = 2$) and 3 ($N_{P'} = 3$).
Jaccard similarity (Leskovec et al., 2020).	$Sim(C_1, C_2) = \dfrac{|C_1 \cap C_2|}{|C_1 \cup C_2|}$	C_1 and C_2 represent two sets. The Jaccard similarity is the size of the intersection divided by the size of their union.
L_r-norm distance measures (Leskovec et al., 2020).	$d([x_1, .., x_n], [y_1, .., y_n])$ $= \left(\sum\limits_{i=1}^{n} |x_i - y_i|^r \right)^{\frac{1}{r}}$	Distance between points as vectors of n real numbers. For $r = 1$, the Manhattan distance is given. For $r = 2$, the Euclidean distance is given.
Cosine similarity (Leskovec et al., 2020).	$cos(\theta) = \dfrac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|}$	Similarity of two vectors of n real numbers, or Boolean components (0 or 1), based on the angle, θ, formed between the vectors. The cosine of this angle is given by the dot product of the two vectors divided by the product of their lengths.

Note. Non-exhaustive list of some similarity and distance measures.

Depending on the type of cognitive representation, different similarity measures are applied. For instance, measures of similarity exist in spatial representations and semantic spaces based on distance measures in high-dimensional spaces (Markman, 1999, Ch. 2), structured representations (ibid., Ch. 5), models of categorization (ibid. Ch. 8), and relational representations (ibid., Ch. 10). Finally, Pearl (2009: Ch. 7) shows how to use an account of interventions in causal models to explicate the notion of similarity in Lewis (1973) and shows that it is possible to derive the same conditional logics as on Lewis' account.

Whether any of these various approaches to similarity fully captures what possible world semantics intends, we take to be an open and controversial question. In Lewis (1979), a system of weights and priorities was presented, which was supposed to explicate how to weight violations of natural laws and agreement on particular facts in the comparison of the similarity of worlds. Like Pearl (2009), this focuses more on underlying causal structure than on physical resemblance but is much broader. In linguistics and philosophy, endeavors to improve upon these qualitative criteria continue (Rescher, 2007; Ippolito, 2016; Starr, 2019).

In Stalnaker (2019, Ch. 11) yet other another notion of similarity is explicated based on objective chances in branching tree representations of events unfolding over time.

Because these issues are so controversial, the strategy that we adopted in this paper was to attempt side-step these open questions by designing an experimental paradigm in which the structure of the task constraints what the most salient similar alternative is – and then to test whether participants selectively make use of this source of information when assigning truth values to indicative and subjunctive conditionals.

Our findings indicate that participants do utilize a salient similar alternative when it is contextually available in their truth evaluations of indicative and subjunctive conditionals. For subjunctive conditionals, it directly predicted the modal truth value assignments, and for indicative conditionals, it was found to play a role in a pragmatic process of accommodating false antecedents (Experiment 1). In Experiment 2, it was investigated which strategy participants employ, when they do not make use of this source of information, but instead resolve the referential ambiguity differently of being presented with an indicative conditional with a demonstrative reference and a descriptive characterization that conflicts with the presented picture. It was found that a second process of accommodation played a role, whereby participants prioritize the demonstrative reference and imagine that the actual picture satisfies the descriptive characterization of indicative conditionals to evaluate their truth.

Regarding indicative conditionals, participants thus acknowledged the use of two routes to accommodating a conditional with false antecedent – by taking the antecedent to refer to the Hypothetical Picture that is not shown or by taking the antecedent to refer to the picture that is shown but imagining a way in which it might still satisfy the antecedent. In both cases, participants used the Hypothetical Picture not shown to guide their truth-table evaluations attesting to the effectiveness of our manipulation of salient possible worlds. At the same time, the paradigm thereby raises the possibility of demand effects, that is the possibility

that participants felt that they were to use the Hypothetical Picture and that they might not conform to the account by possible worlds if salient possible worlds had not been prompted.

Yet, demand effects of this kind alone cannot account for the present pattern of results. A task demand to use the Hypothetical Picture can thus not explain why there was an asymmetry between indicative and subjunctive conditionals in Exps. 1 and 2, suggesting that participants did use the Hypothetical Picture for subjunctive conditionals, but less so for indicative conditionals. Moreover, given such a task demand, we would have expected the truth evaluations of indicative conditionals to be based on the Hypothetical Picture in Exp. 3, in which case 'M subjunctive' should have accounted for participants' responses (whereas in fact it only account for the performance of 4 participants). Indeed, the task demand hypothesis cannot explain why participants selectively felt a task demand to use the hypothetical pictures for the truth evaluations of subjunctive conditionals in Experiments 1 and 2. Nevertheless, we acknowledge that task demands are a concern and that it would be desirable to devise alternative manipulations that constrain the construction and choice of possible worlds in other ways.

Like in most reasoning experiments, a number of different additional task-demand effects are conceivable: Since participants were also asked to evaluate the truth or falsity of the antecedents and consequents of conditionals before evaluating the conditionals, one could argue that there should have been a task demand for applying a truth-functional semantics to conditionals like the material implication. Similarly, since participants were asked to assign conditional probabilities before evaluating the truth of conditionals, one could argue that there should have been a task demand for applying a Jeffrey truth table. Moreover, since participants had the option to use the third truth value response (Neither Nor), one could argue that there was a task demand for applying the de Finetti truth table. Finally, in Exp. 3, since participants were asked to evaluate reason relations before assigning truth values to indicative

conditionals, one could argue that there was a task demand for applying truth-conditional inferentialism. Yet, neither the material implication, the Jeffrey Table, the de Finetti Table, nor Truth-Conditional Inferentialism could account for our results. Given these possibilities and the shortcomings of each of these task-demand hypotheses to account for our results, it is desirable to elaborate on these task-demand hypotheses and to investigate their possible role in shaping our results in future work.

Likewise, it is desirable that future work investigates further aspects of possible worlds semantics using different experimental paradigms. While our findings alone may not settle the final score on the possible worlds semantic as applied to indicative conditionals, they at the very least show that Stalnaker (2011, 2014) and Nute (1980) were right to emphasize the role of accommodation in relation to the processing of conditionals. According to Thomason (1990), accommodation is a type of obstacle removal of cooperative interlocutors, who recognize the plan and intentions of the speaker and help eliminate potential obstacles, which occurs in a broad range of different linguistic phenomena (see also Lewis, 1979; Roberts, 2015). Despite its frequent occurrence, it has hitherto not played a role in the psychology of reasoning, however. Accordingly, influential theories in the psychology of reasoning, like the suppositional theory or mental model theory, are well-advised to extend their theories to take this important aspect of communication into account.

Conclusion

In this paper we presented a new experimental task for investigating the possible worlds account of conditionals. This enabled us to present some of the first direct empirical corroboration for its truth conditions as capturing the mean response frequencies for both indicative and subjunctive conditionals. In contrast, none of the salient alternatives, like the material implication, the de Finetti table, the Jeffrey Table, the multi-dimensional approach of

Bradley (2012), nor the revised mental model theory, were able to capture the mean response frequencies of indicative conditionals in Experiments 1 and 2.

What enabled the possible worlds account to do better than the other accounts was its prediction of a bi-modal response pattern based on whether or not participants accommodated in the case of indicative conditionals with false antecedents. By incorporating this factor of whether participants accommodated the reference of the false antecedent, the possible worlds account was able to account for when participants evaluated indicative and subjunctive conditionals alike, and when their truth evaluations diverged, unlike the other accounts.

In Experiment 2, we examined an alternative hypothesis concerning this bi-modal pattern which held that it was attributable to a reaction to treat indicative conditionals with false antecedents as invariable false. In an experiment that closely replicated the findings of Experiment 1, it was found, however, that the modal tendency of participants, who indicated that they did not accommodate by reassigning the reference of indicative conditionals with false antecedents to a paired Hypothetical Picture not shown, was to instead accommodate by imagining some alternative scenario such that the Actual Picture did indeed satisfy the antecedent.

In Experiment 3, participants' truth evaluations of indicative conditionals were modeled as coming from a mixture distribution of competing truth tables. Through model comparisons, it was found that possible worlds semantics accounted for the truth evaluations of the largest subgroup of participants within our experimental task. At the same time, minorities of participants were classified as following competing truth tables, which could account for why the possible worlds account by itself only had a satisfactory fit for aggregate truth value frequencies in Experiment 1 but did not account for the pattern of individual differences in the data.

References

Adams, E. (1970). Subjunctive and Indicative Conditionals. *Foundations of Language, 6*, 89-94.

Adams, E. (1975). The logic of conditionals: an application of probability to deductive logic. Dordrecht: Reidel.

Anderson, A. R. (1951). A note on subjunctive and counterfactual conditionals. *Analysis*, *12*, 35–38.

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

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review, 6*, 57–86.

Bennett, J. (2003). *A Philosophical Guide to Conditionals*. Oxford: Oxford University Press.

Bradley, R. (2012). Multidimension Possible-world Semantics for Conditionals. *The Philosophical Review, 121*(4), 539-71.

Byrne, R. M. J. (2016). Counterfactual Thought. *Annual Review of Psychology*, *67*(1), 135–157. https://doi.org/10.1146/annurev-psych-122414-033249

Cariani, F. and Rips, L. J. (2017). Conditionals, Context, and the Suppression Effect. *Cognitive Science*, *41*, 540-589.

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

Douven, I., Elqayam, S., Singmann, H., Wijnbergen-Huitink, JV., (2020). Conditionals and inferential connections: toward a new semantics. *Thinking and Reasoning*, *26*(3), 311-351.

Edgington, D. (1995). On conditionals. *Mind*, *104*, 235-329.

Edgington, D. (2006). Conditionals. In: E.N. Zalta (Eds.), *The Stanford Encyclopedia of*

Philosophy (Winter 2008 edn.). Retrived from: http://plato.stanford.edu/archives/win2008/entries/conditionals/.

Edgington, D. (2008). I-Counterfactuals. *Proceedings of the Aristotelian Society*, *108*(1), 1–21.

Erdfelder, E., Auer, T., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models. *Zeitschrift fur Psychologie / Journal of Psychology, 217*, 108–124.

Evans, J. S. B., & Newstead, S. E. (1977). Language and reasoning: A study of temporal factors. *Cognition*, *5*(3), 265-283.

Evans, J. St. B. T., Handley, S. J., Neilens, H., and Over, D. E. (2007). Thinking about conditionals: A study of individual differences. *Memory & Cognition*, *35*(7), 1772-84.

Evans, J. St. B. T. & Over, D. (2004). *If*. Oxford: Oxford University Press.

Fugard, A. J. B., Pfeifer, N., Mayerhofer, B., and Kleiter, G. D. (2011). How People Interpret Conditionals: Shifts Toward the Conditional Event. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(3), 635-648.

Garson, J. W. (2013). *Modal Logic for Philosophers.* (Second Edition) New York: Cambridge University Pres.

Gelman, A., Carlin, J. B., Stern, H. B., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis* (Third Edition). Boca Raton: Chapman & Hall/CRC.

Goodwin, G. P., and Johnson-Laird, P. N. (2018). The Truth of Conditional Assertions. *Cognitive Science*, *42*, 2502-2533.

Heck, D. W., Arnold, N. R., and Arnold, D. (2018). TreeBUGS: An R package for hierarchical multinomial-processing-tree modelling. *Behavior Research Methods, 50*, 264-84.

Iatridou, S. (2000). The Grammatical Ingredients of Counterfactuality. *Linguistic Inquiry*, *31*(2), 231–270. https://doi.org/10.1162/002438900554352

Ippolito, M. (2016). How Similar Is Similar Enough? *Semantics and Pragmatics*, *9*(6), 1–60.

Jeffrey, R. C. (1991). Matter of fact conditionals. Aristotelian Society Supplementary Volume, 65, 161-183.

Johnson-Laird, P.N., & Tagart, J. (1969. How implication is understood. *American Journal of Psychology, 2*, 367-373.

Johnson-Laird, P.N., & Byrne, R.M.J. (1991). *Deduction.* Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Johnson-Laird, P. N. and Ragni, M. (2019). Possibilities as the foundation of reasoning. *Cognition*, *193*, 103950. doi: 10.1016/j.cognition.2019.04.019

Karabatsos, G. (2005). The exchangeable multinomial model as an approach for testing axioms of choice and measurement. *Journal of Mathematical Psychology, 49*, 51-69.

Khemlani, S. S., Byrne, R. M. J., & Johnson-Laird, P. N. (2018). Facts and Possibilities: A Model-Based Theory of Sentential Reasoning. *Cognitive Science*, *42*(6), 1887–1924. https://doi.org/10.1111/cogs.12634

Klauer, K. C. (2010). Hierarchical Multinomial Processing Tree Models: A Latent-Trait Approach. *Psychometrika, 75*(1), 70-98.

Klauer, K. C., Singmann, H., & Kellen, D. (2015). Parametric order constraints in Multinomial Processing Tree Models: An extension of Knapp & Batchelder (2004). *Journal of Mathematical Psychology*, 64, 1-5.

Kratzer, A. (2012). *Modals and Conditionals: New and Revised Perspectives*. New York: Oxford University Press.

Lassiter, D. (2017). Probabilistic language in indicative and counterfactual conditionals. *Proceedings of SALT, 27*, 525-546.

Leskovec, J., Rajaraman, A., and Ullman, J. D. (2020). *Mining of Massive Datasets* (3rd Edition). Cambridge: Cambridge University Press.

Lewis, C. I. (1912). Implication and the Algebra of Logic. *Mind*, *21*(84), 522–531.

Lewis, D. (1973). *Counterfactuals*. Cambridge, MA: Harvard University Press.

Lewis, D. (1976). Probabilities of conditionals and conditional probabilities. *Philosophical Review*, *85*, 297-315.

Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic*, *8*, 339-59.

Markman, A. B. (1999). *Knowledge Representations*. New York: Psychology Press.

Nickerson, R. (2015). *Conditional Reasoning. The Unruly Syntactics, Semantics, Thematics, and Pragmatics of "If"*. Oxford: Oxford University Press.

Nute, D. (1980). Conversational Scorekeeping and Conditionals. *Journal of Philosophical Logic*, *9*, 153-166.

Oaksford, M., Over D. E., & Cruz, N. (2019). Paradigms, possibilities, and probabilities: Comment on Hinterecker et al. (2016). *Journal of Experimental Psychology: Learning Memory and Cognition*, *45*, 288-297.

Over, D. E., Hadjichristidis, C., Evans, J. S. B. T., Handley, S. J., & Sloman, S. A. (2007). The probability of causal conditionals. *Cognitive Psychology*, *54*(1), 62–97.

Over, D. E., & Baratgin, J. (2017). The "defective" truth table: Its past, present, and future. In N. Galbraith, E. Lucas, & D. E. Over (Eds.), *The Thinking Mind: A Festschrift for Ken Manktelow* (pp. 15-28). Abingdon, UK: Routledge.

Over, D. E. (2020). The development of the new paradigm in the psychology of reasoning. In S. Elqayam, I. Douven, J. St. B. T. Evans, & N. Cruz (Eds.), Logic and uncertainty in the human mind (pp. 243-263). Abingdon: Routledge.

Over, D. E., & Cruz, N. (2018). Probabilistic accounts of conditional reasoning. In Linden J. Ball and Valerie A. Thompson (Eds.), International handbook of thinking and reasoning (pp. 434-450). Hove: Psychology Press.

Pearce, J. M. (1987). A model of stimulus generalization for Pavlovian conditioning. *Psychological Review*, *94*, 61-73.

Pearce, J. M. (1994). Similarity and discrimination: a selective review and a connectionist model. *Psychological Review*, *101*, 587-607.

Pearl, J. (2009). *Causality: models, reasoning, and inference* (2th Ed.). Cambridge: Cambridge University Press.

Pearl, J. (2013). Structural Counterfactuals: A Brief Introduction. *Cognitive Science*, *37*(6), 977-985.

Pfeifer, N. & Tulkki, L. (2017). Conditionals, Counterfactuals, and Rational Reasoning: An Experimental Study of Basic Principles. *Minds and Machines, 27*(1), 119-165.

Plummer, M. (2019). rjags: Bayesian Graphical Models using MCMC. R package version 4-10. https://CRAN.R-project.org/package=rjags

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

Portner, P. (2009). *Modality*. Oxford: Oxford University Press.

Potts, C. (2005). *The Logic of Conventional Implicatures.* Oxford: Oxford University Press.

Quine, W. V. O. (1960).Word and object. Cambridge, MA: MIT Press.

Roberts, C. (2015). Accommodation in a language game. In Loewer, B. and Schaffer, J. (Eds.), *A Companion to David Lewis* (pp. 345-66). Malden, MA: Wiley-Blackwell.

Quelhas, A. C., Rasga, C., & Johnson-Laird, P. N. (2018). The Relation Between Factual and Counterfactual Conditionals. *Cognitive Science*, *42*(7), 2205–2228. https://doi.org/10.1111/cogs.12663

Reips, U. D. (2002). *Standards for Internet-based experimenting. Experimental Psychology, 49* (4), *243*-256.

Rescher, N. (2007). *Conditionals*. Cambridge, Massachusetts: The MIT Press.

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review, 95*, 318–339.

Rips, L. J. (2010). Two Causal Theories of Counterfactual Conditionals. *Cognitive Science*, *34*(2), 175-221.

Rips, L. J., and Marcus, S. L. (1977). Suppositions and the analysis of conditional sentences. In Just, M. A., and Carpenter, P. A. (Eds.), *Cognitive Processes in Comprehension* (pp. 185-220). New York: Lawrence Erlbaum Associates.

Roberts, C. (1996). Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics & Pragmatics*, *5*(6), 1-69.

Roberts, C. (2015). Accommodation in a language game. In Loewer, B. and Schaffer, J. (Eds.), *The Blackwell Companion to David Lewis* (pp. 345-366). Oxford: Blackwell.

Santamaría, C., Espino, O., & Byrne, R. M. J. (2005). Counterfactual and Semifactual Conditionals Prime Alternative Possibilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 1149–1154. https://doi.org/10.1037/0278-7393.31.5.1149

Schroyens, W. (2010). A meta-analytic review of thinking about what is true, possible, and irrelevant in reasoning from or reasoning about conditional propositions. *European Journal of Cognitive Psychology*, 22 (6), 897-921.

Skovgaard-Olsen, N. (2021). Relevance and Conditionals: A Synopsis. In: Elqayam, S., Douven, I., Cruz, N., Evans, J. (Eds.), *Festschrift for David Over*. Routledge (pp. 192-206).

Skovgaard-Olsen, N. (2023). *Reasoning Studies. From Single Norms to Individual Differences.* (Habilitation in Psychology, University of Freiburg)

Skovgaard-Olsen, N. and Collins, P. (2021). Indicatives, Subjunctives, and the Falsity of the Antecedent. *Cognitive Science, 45*(11), https://doi.org/10.1111/cogs.13058.

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

Skovgaard-Olsen, N., Kellen, D., Krahl, H., and Klauer, K. C. (2017). Relevance differently affects the truth, acceptability, and probability evaluations of 'And', 'But', 'Therefore', and 'If Then'. *Thinking and Reasoning*, *23*(4), 449-482.

Skovgaard-Olsen, N., Stephan, S., and Waldmann, M. (2021). Conditionals and the Hierarchy of Causal Queries. *Journal of Experimental Psychology: General*, *150*(12), 2472-2505.

Sloman, S. A. and Lagnado, D. A. (2005). Do We "do"? *Cognitve Science*, *29*, 5-39.

Smith, J. B., & Batchelder, W. H. (2010). Beta-MPT: Multinomial processing tree models for addressing individual differences. *Journal of Mathematical Psychology, 54*, 167–183.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and Cognition* (Second Edition). Oxford, England: Blackwell.

Spohn, W. (2013). A ranking-theoretic approach to conditionals. *Cognitive Science*, *37*, 1074–1106.

Stalnaker, R. C. (1968). A Theory of Conditionals. In: Rescher, N. (Eds.), *Studies in Logical Theory* (pp. 98-112). Oxford: Basil Blackwell.

Stalnaker, R. (1974). Pragmatic presuppositions. In M. Munitz & P. Unger (Eds.), *Semantics and philosophy* (pp. 197–214). New York, NY: New York University Press.

Stalnaker, R. C. (1975). Indicative conditionals. *Philosophia, 5*(3), 269-286.

Stalnaker, R. (2011). Conditional propositions and conditional assertions. In: Egan, A. and Weatherson, B. (Eds.), *Epistemic Modality* (pp. 227-48). Oxford: Oxford University Press.

Stalnaker, R. (2014). *Context*. Oxford: Oxford University Press.

Stalnaker, R. (2019). *Knowledge and Conditionals*. Oxford: Oxford University Press.

Stalnaker, R., & Jeffrey, R. (1994). Conditionals as random variables. In E. Eells & B. Skyrms (Eds.), *Probability and conditionals* (pp. 31-46). Cambridge: Cambridge University Press.

Starr, W. B. (2014). A Uniform Theory of Conditionals. *Journal of Philosophical Logic*, *43*(6), 1019-1064.

Starr, W. (2019). Counterfactuals. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition). Retrieved from forthcoming URL = <https://plato.stanford.edu/archives/fall2019/entries/counterfactuals/>.

Thomason, Richmond H. (1990). "Accomodation, meaning, and implicature: Interdisciplinary foundation for pragmatics". In Cohen, P. R, Morgan, J. L, and Pollack, M. E. (Eds.), *Intentions in Communication* (pp. 325–363). Cambridge, Mass.: MIT Press.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*(4), 327-352.

van Wijnbergen-Huitink, Elqayam, S. and Over, D. E. (2015). The probability of iterated conditionals. *Cognitive Science, 39*(4), 788-803.

von Fintel, K. (1997). The presupposition of subjunctive conditionals. In *MIT working papers in linguistics 25* (pp. 29–44). Cambridge, MA: MIT Working Papers in Linguistics.

von Fintel, K. (2012). Subjunctive conditionals. In G. Russell & D. Graff Fara (Eds.), *The Routledge companion to the philosophy of language* (pp. 466–477). New York, NY: Routledge.

Williamson, T. (2020). *Suppose and Tell. The Semantics and Heuristics of Conditionals.* Oxford: Oxford University Press.

Appendix A: Accommodation

The figures below illustrate the bimodal pattern in the truth evaluation of indicative conditionals, which arose depending on whether participants accommodated the reference of the antecedent to refer to the Hypothetical Picture across three trials.
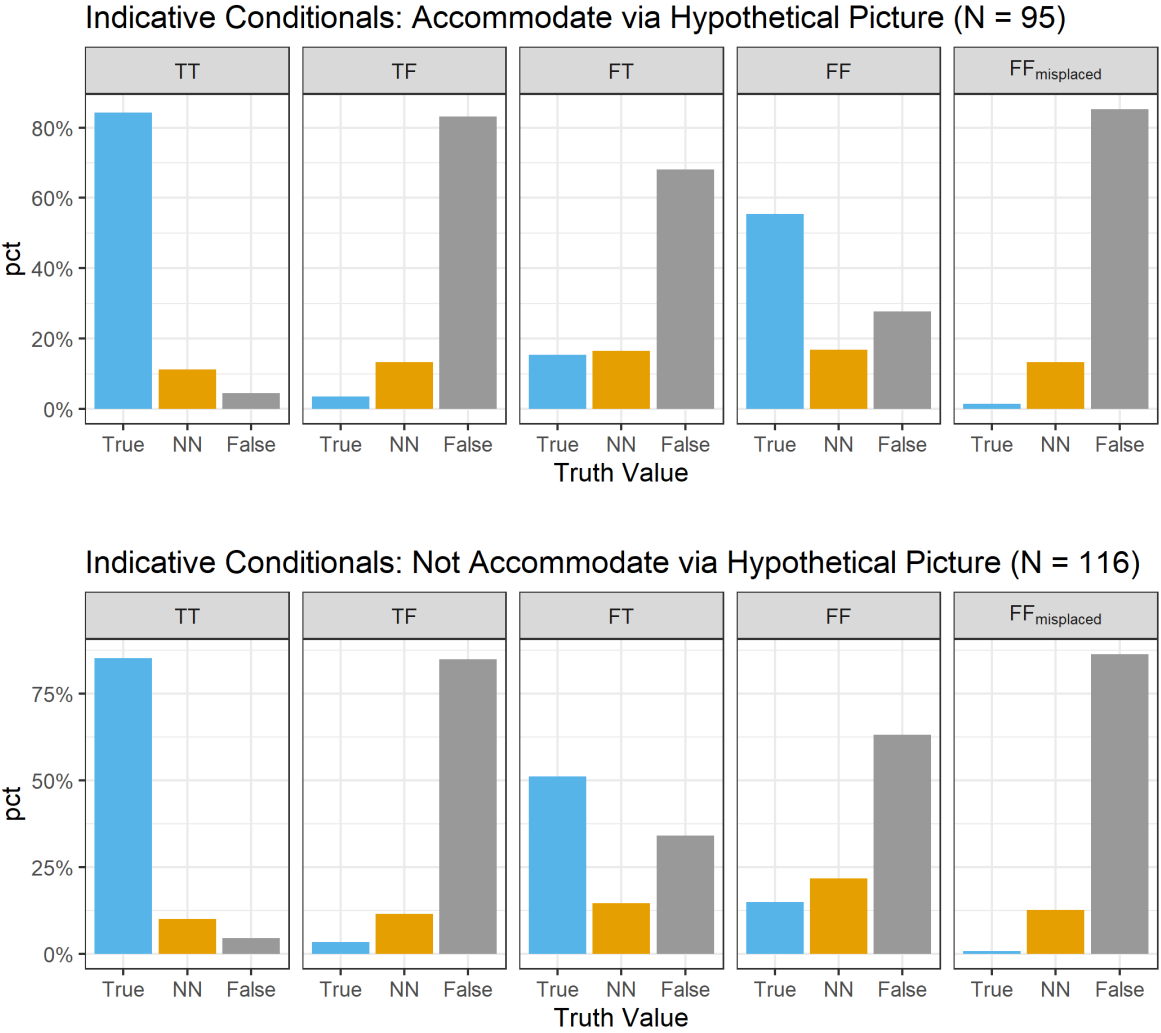

Figure A1. Truth Tables of Indicative Conditionals as a Function of Accommodation. 'FF' = a conditional that is False False w.r.t. the Actual Picture shown but True True w.r.t. the salient Hypothetical Picture; 'FF$_{misplaced}$' = a conditional that is False False w.r.t. the Actual Picture shown but True False w.r.t. the salient Hypothetical Picture.

As shown, for the 95 participants who did accommodate, the indicative conditionals were evaluated corresponding to the modal values for subjunctive conditionals (see Figure 1). In contrast, the modal truth values flipped in the FT and FF cells for the 116 participants who

chose not to accommodate. When the truth evaluations for these two groups are merged, the aggregate pattern for indicative conditionals displayed in Figure 1 arises.

As a result, it is found that participants' decisions of whether to accommodate the antecedent across three trials are predictive of the qualitative differences in truth evaluations of indicative conditionals displayed in Figure A1.

Appendix B: Replication of Experiment 1

Below we show that Experiment 2 replicates the results of Experiment 1.

Table B1. Model-Comparison Results, Experiment 2

Model	p_{T1}	p_{T2}	WAIC	Weight
$M_{FT,FF,ind}$	.16	.00	4858.6	0
$M_{saturated}$	.49	.13	4132.3	1
$M_{sentence}$	.00	.00	6760.7	0
$M_{FF,FFmis,ind}$	.00	.00	4746.2	0
$M_{PossibleWorld}$	.22	.00	2264.9	1
$M_{Subjunctive}$	.00	.00	2639.1	0
$M_{Material}$	.00	.00	3505.5	0
$M_{deFinetti}$	.00	.00	3776.2	0

Note. Note that the test statistics T_1 and T_2 represent Bayesian *p* values and are based on the posterior predictive model checks in Klauer (2010). WAIC = Watanabe-Akaike information criterion. Weight = Akaike weight of WAIC. This table reports the same model comparisons for Experiment 2 as Tables 5 and 6 report for Experiment 1.

Appendix C: Hierarchical MPT Models

Multinomial processing trees models model response frequencies for a set of mutually exclusive categorical response outcomes (Riefer & Batchelder, 1988). In this paper, multinomial processing tree models are used to analyze the observed response frequencies in truth value assignments (T, F, NA). To implement the various truth tables, inequality constraints are introduced for the MPT parameters following the analytic approach of Klauer et al. (2015, Appendix A). For example, if response "T" is predicted to be the modal response, the response probabilities $\eta_1, \eta_2,$ and η_3 for the responses "T", "F", and "NA", respectively,

should satisfy the inequalities $\eta_1 \geq \eta_2$ and $\eta_1 \geq \eta_3$. This is guaranteed by parameterizing the three response probabilities as follows (Eq. 1):
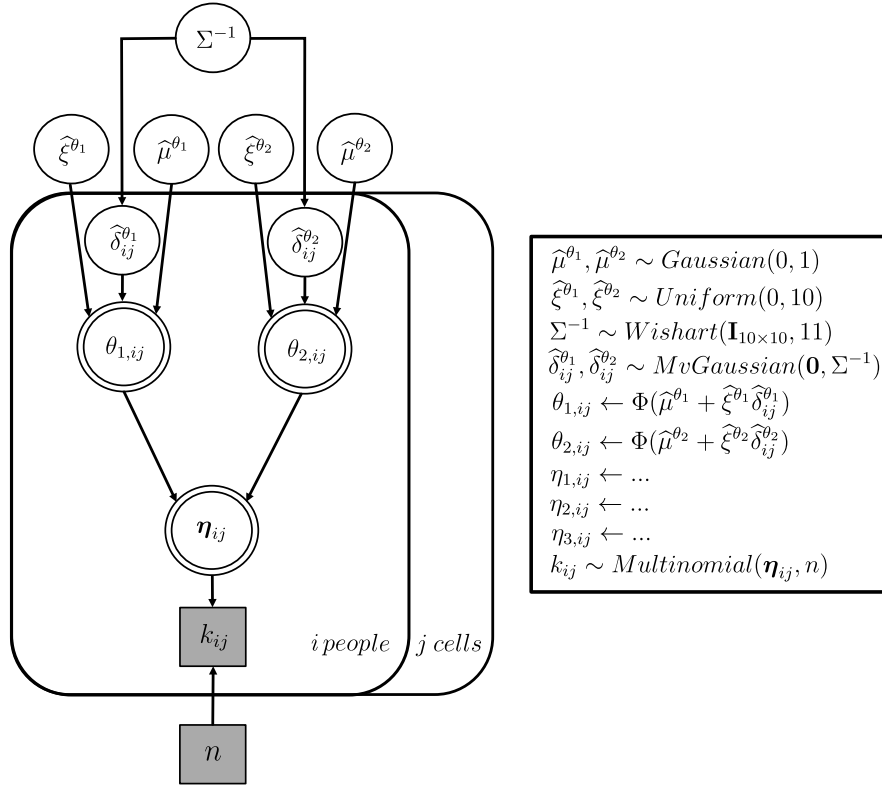
$$\begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} = (1 - \theta_1)(1 - \theta_2) \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \theta_1(1 - \theta_2) \begin{pmatrix} 1/2 \\ 1/2 \\ 0 \end{pmatrix} + (1 - \theta_1)\theta_2 \begin{pmatrix} 1/2 \\ 0 \\ 1/2 \end{pmatrix} + \theta_1\theta_2 \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix}$$

Instead of aggregating the categorical outcomes across participants, the hierarchical latent trait approach of Klauer (2010) is followed in Experiment 1, which adds a hierarchical structure in which the participants' MPT parameters are constrained to be samples from a population-level probability distribution.

On this approach, a probit link function is used to transform MPT parameters (representing probabilities between 0 and 1) to the real line, $\Phi^{-1}(\theta)$. The transformed parameters are then modelled via a multivariate normal distribution while estimating mean, μ, and covariance matrix, Σ, from the data. The advantage of this approach is that heterogeneity in parameter estimates across participants and correlations among MPT parameters can be accommodated while allowing for partial aggregation of statistical information across participants in the posterior parameters of the multivariate normal distribution (Klauer, 2010). Accordingly, for each participant, i, the probit-transformed parameters are additively decomposed into a group mean, μ, and a random effect, $\Phi^{-1}(\theta) = \mu + \delta_i$.

In Experiment 1, a set of hierarchical multinomial models following this approach with different order constraints implementing competing truth tables were contrasted in a model-comparison exercise. Table 1C illustrates the general form of these models.

Table 1C. Hierarchical Latent Trait MPT Model with Inequality Constraints

$$\widehat{\mu}^{\theta_1}, \widehat{\mu}^{\theta_2} \sim Gaussian(0,1)$$
$$\widehat{\xi}^{\theta_1}, \widehat{\xi}^{\theta_2} \sim Uniform(0,10)$$
$$\Sigma^{-1} \sim Wishart(\mathbf{I}_{10 \times 10}, 11)$$
$$\widehat{\delta}^{\theta_1}_{ij}, \widehat{\delta}^{\theta_2}_{ij} \sim MvGaussian(\mathbf{0}, \Sigma^{-1})$$
$$\theta_{1,ij} \leftarrow \Phi(\widehat{\mu}^{\theta_1} + \widehat{\xi}^{\theta_1}\widehat{\delta}^{\theta_1}_{ij})$$
$$\theta_{2,ij} \leftarrow \Phi(\widehat{\mu}^{\theta_2} + \widehat{\xi}^{\theta_2}\widehat{\delta}^{\theta_2}_{ij})$$
$$\eta_{1,ij} \leftarrow \dots$$
$$\eta_{2,ij} \leftarrow \dots$$
$$\eta_{3,ij} \leftarrow \dots$$
$$k_{ij} \sim Multinomial(\boldsymbol{\eta}_{ij}, n)$$

Note. $\boldsymbol{\eta}_{i,j}$ is a vector of $\eta_{1,i,j}$, $\eta_{2,i,j}$, and $\eta_{3,i,j}$. See Eq. 1 above for the parameterization of these parameters in terms of $\theta_{1,i,j}$ and $\theta_{2,i,j}$. The j cells correspond to {TT, TF, FT, FF, FF$_{\text{misplaced}}$}. Since there are five truth table cells, with three categorical responses parameterized in terms of 2 theta parameters each, there are $5 \times 2 = 10$ theta parameters in total and 10+1 degrees of freedom of the inverse Wishart distribution together with a 10×10 identity matrix, as scale matrix.

In Experiment 3, these hierarchical multinomial processing trees were extended by representing several competing truth tables as mixture components of a mixture distribution. A categorical variable with a Dirichlet prior assigned participants to each of the different mixture components and thereby permitted the assignment of individual truth tables to participants to examine individual differences.

In Experiment 3, three implementations of the hierarchical mixture distributions were contrasted in a model comparison (containing 3, 4, or 5 mixture components, respectively), which used a scaled inverse-Wishart prior for modelling the covariance matrix, following the the hierarchical latent trait approach of Klauer (2010) from Experiment 1, illustrated in Table 1B. In addition, a fourth model with 4 mixture components was fitted which followed the

approach of Smith and Batchelder (2010) of using independent beta distributions for the different MPT parameters. Of the two, the latter assumes independence of MPT parameters in the hierarchical prior distribution while the former explicitly accommodates correlations among the MPT parameters already in the prior distribution.

All models featured a saturated mixture component to filter out noisy respondents, which did not fit any of the models in virtue of violating the shared predictions of 'True' and 'False' in the TT and TF cells.

The models were fitted in a Bayesian framework through the Gibbs sampler implemented in JAGS (Plummer, 2019), which estimates the posterior distributions of model parameters by means of Monte Carlo-Markov chains.