Master's Thesis

ϵ-MMT
A Cognitive Computational Model Extending the Mental Model Theory with ϵ-semantics

Sara Todorovikj

Examiner: apl. Prof. Dr. Dr. Marco Ragni
Adviser: apl. Prof. Dr. Dr. Marco Ragni

Albert-Ludwigs-University Freiburg
Faculty of Engineering
Department of Computer Science
Chair for Foundations of Artificial Intelligence
Cognitive Computation Lab

September 21, 2020

Writing period

17 .02. 2020 – 21. 09. 2020

Examiner

apl. Prof. Dr. Dr. Marco Ragni

Second Examiner

Prof. Dr. Gabriele Kern-Isberner

Adviser

apl. Prof. Dr. Dr. Marco Ragni

Declaration

I hereby declare, that I am the sole author and composer of my thesis and that no other sources or learning aids, other than those listed, have been used. Furthermore, I declare that I have acknowledged the work of others by providing detailed references of said work.
I hereby also declare, that my Thesis has not been prepared for another examination or assignment, either wholly or excerpts thereof.

Place, Date

Signature

Abstract

One of the goals of cognitive science is to comprehend the human reasoning processes. In order to do that a significant amount of cognitive models have been developed over the years. In the 60s scientists assumed that standard deductive logic is the basis for reasoning, creating a so-called *deductive* paradigm. However, through psychological experiments it has been shown that human reasoning does not always conform to logical rules. Acknowledging uncertainties when reasoning led to the development of a new *probabilistic* paradigm. This thesis examines how the performances of models for conditional reasoning from the different paradigms compare to each other. I developed a *benchmark* that tests the performance of two deductive and three probabilistic theories on eight experiments with diverse data. The selected experiments aim to manipulate reasoning properties leading to a larger variance in conclusion endorsements made by individuals. Additionally, I propose a cognitive computational model – ϵ-MMT, that combines the deductive and probabilistic modeling approaches. I take the mental model representation of conditionals and extend it with probabilities based on Pearl's ϵ-semantics. I include the new model in the benchmark in order to determine its competence among state-of-the-art cognitive models. Moreover, I examine the ability of ϵ-MMT to account for various reasoning effects and properties through its parameters.

The benchmark results show that overall the probabilistic models achieve the best fits. However, by analyzing the models' performances on each experiment separately it is shown that in some scenarios they fall behind the deductive ones. Ultimately, this supports the fact that there is not one single theory that can explain reasoning processes in all circumstances. ϵ-MMT is one of the two best-performing models in the benchmark. Additionally, I present a psychological interpretation of ϵ-MMT's parameters. A statistical analysis on the parameter values delivers significant results which lead to the conclusion that the model can successfully provide insight into human reasoning and how they interpret conditionals.

Contents

List of Figures

List of Tables

1 Introduction and Theoretical Background

> "Humans, it seems, know things; and what they know helps them do things." (Russel & Norvig, 2010, pg. 234)

We are constantly gathering knowledge and information about ourselves and our surroundings. We then use them to infer conclusions by processes of *reasoning* on a daily basis. For example, consider a village where the people are trying to judge whether a woman is a witch or not. They have learned that if she weighs the same as a duck, then she is a witch. So, they put her on a scale against a duck. It turns out that she does indeed weigh the same as a duck, therefore, she must be a witch[1] – a logically valid inference. However, consider having a larger knowledge base, introducing a very strong belief in the fact that witches are not real. In that case one would most probably no longer be in favor of the conclusion that she is a witch, meaning that a logically valid inference would be refuted.

On the other hand, imagine a card that has a letter on one side and a number on the other. You learn that if the letter is A then the number is 3. Then you are given a card with the letter A, and you conclude that the number on the other side is 3. Given such an abstract topic, no background knowledge or any other factor would prevent a person from making that inference, even though it is the same inference as in the first example from a logical point of view.

But, how do people reach these conclusions? What are the cognitive processes that lead to them? Throughout the years, a vast amount of cognitive models and theories have been developed in an effort to gain a better understanding of human reasoning. In the 60s, cognitive scientists were following a *deductive* path assuming that logic is the basis for reasoning (Evans & Over, 2004).

One of the most prominent deductive reasoning theories is the Mental Model Theory (Johnson-Laird & Byrne, 1991). It assumes that individuals use *mental models* to

[1] You might recognize this inference from the 1975 movie "Monty Python and the Holy Grail". For an explanation as to why this rule is true according to the movie, please see Appendix A.6.

represent knowledge. The rule “If she weighs the same as a duck then she is a witch” used in the previous example is called a *conditional rule.* It describes the causal relationship between the propositions “she weighs the same as a duck” and “she is a witch”. Its mental model representation would be:

duck	witch
...	

The first proposition is represented with “duck” and the second with “witch”. The three dots mean that this representation is not entirely complete. Imagine we know that she could not weigh the same as a duck but she could still be a witch, then the model “¬duck witch” (“¬” for negation) would also be added to the representation. Additionally, we know that she could not weigh the same as a duck and she could not be a witch. That world is represented with the model “¬duck ¬witch”. By adding these two models we reach the complete *fleshed-out* mental model representation for which the conditional rule *holds*:

duck	witch
¬duck	witch
¬duck	¬witch

But, as we have seen in the example – what if we believe that she might weight the same as a duck, yet she is not a witch? From a logical point of view, that state falsifies the conditional. Taking into consideration only the propositions’ truth states for which the conditional is true is called *material implication.* Reasoning theories in the deductive paradigm follow the material implication characterization when interpreting conditional rules (Elqayam & Over, 2013). For instance, the Mental Model Theory does not support adding the model “duck ¬witch” to the mental representation (Johnson-Laird & Byrne, 1991).

Nevertheless, we still strongly believe that there is a *possibility* she could weigh the same as a duck but she is not a witch. That way, we question the conditional’s certainty in a way that standard deductive logic cannot grasp (Oaksford & Chater, 1995). In the present, we interpret these uncertainties as *probabilities* (Over, 2004), leading to a major shift in the cognitive modeling research approach. This new paradigm of reasoning comprised of probabilistic models is also called the *Bayesian paradigm.* Now, the focus diverts from asking whether a conclusion is true to how *likely* it is.

In this thesis, I propose a new cognitive reasoning model which combines the deductive and probabilistic modeling approaches. I take the mental model representation of conditionals, include the logically invalid mental model and interpret them as *possible worlds*. For example, the conditional rule "If she weighs the same as a duck then she is a witch" is represented with the worlds:

duck	witch
¬duck	witch
¬duck	¬witch
duck	¬witch

Following Pearl's (1991) ϵ-semantics, this representation is extended with a probability distribution that reflects the likelihood of each world. From now on, I will refer to the proposed model as ϵ-MMT. To illustrate the main idea behind ϵ-MMT, imagine that you are skeptical about the existence of witches. Then, the world where she weighs the same as a duck and she is a witch ("duck witch") could be 1% probable, according to you. On the other hand, we strongly believe that it is possible she weighs the same as a duck but she is not a witch ("duck ¬witch"), so that world could be 99% probable. Your probability distribution of the possible worlds would be represented as:

duck	witch	1%
¬duck	witch	0%
¬duck	¬witch	0%
duck	¬witch	99%

The inferences we make in the end depend on the probability distribution we have for the possible worlds, which in turn depends on our subjective beliefs about the content.

Now, suppose it is up to you to determine the likelihood that she is a witch. You learn that she weighs the same as a duck. What would your answer be if beforehand you were told that "if she weighs the same as a duck then she is a witch"? How would your answer differ in case you were never presented with the rule? Singmann, Klauer, and Beller (2016) report the results of experiments where the *conditional presentation form* was manipulated in such a way, within participants. This is one of many reasoning properties that psychological experiments aim to manipulate in order to provide insight into how people's inferences deviate in different circumstances.

Cognitive models can be applied to such experimental data and their performance can be evaluated. In the scope of this thesis I am interested in determining how well do certain models perform in contrast to others. In order to do that I developed a *benchmark*. We have now reached the first research question of this thesis:

RQ 1: What would be a good form of a benchmark?

A benchmark would provide an answer to the question "Which model provides a best experimental data fit overall?". However, in the world of cognitive modeling there are many more questions that can be asked. For example, how do deductive models compete with new probabilistic ones? Would the latter always perform better or would they fall behind in some scenarios? Ultimately, is there one model that can explain the variety in individuals' answers in different circumstances? On the other hand, do all of the cognitive theories' assumptions hold? For instance, one of the Mental Model Theory's assumptions is that the material implication interpretation holds, but does experimental data corroborate that? The benchmark that I develop for this thesis gives answers and insight into these questions.

Now that a benchmark is developed, I can pose the second research question of this thesis:

RQ 2: How do various cognitive models perform when evaluated on sensible data?

The benchmark contains data from experiments that manipulate various reasoning properties, e.g., the conditional presentation form. Additionally, some experiments provided participants with *abstract* materials and some with *everyday* contents. To which extent are different models able to adapt to this variety in data? The benchmark provides an answer to which model has the best capability to do that. Moreover, I will analyze the models' performances on each experiment separately. By doing that I will determine in which circumstances certain models reach their peak performance and when do they fall short.

The new cognitive model that I propose, ϵ-MMT, is also included in the benchmark in order to assess its competence, leading to the next research question:

RQ 3: Can ϵ-MMT compete with state-of-the-art models?

ϵ-MMT extends a deductive way of conditional representation with probabilistic assumptions from the new paradigm. Does this merger allow for competent performance? How does ϵ-MMT compare to other probabilistic, state-of-the-art models?

> "To explain reasoning with conditionals, however, we need to understand how they are understood." (Johnson-Laird & Byrne, 1991, pg. 63)

This quote leads us to the final major focus and research question of this thesis – ϵ-MMT's parameters and their psychological interpretation.

RQ 4: Can ϵ-MMT account for various reasoning effects and properties and provide insight into human reasoning and interpretation?

Can this model help us learn more about human reasoning? Can it adapt to the experiments' manipulations and represent reasoning effects adequately? I answer this question by providing a psychological interpretation for the models' parameters and performing statistical analysis on their values.

Now that I introduced the four main focus points, I will provide an outline of the thesis structure in the following. Then, in order to understand the research questions better and answer them, we will go on a brief stroll through the relevant theoretical background. In the upcoming sections, I provide the definition of conditionals and possible inference forms. Afterwards, I explore how conditionals have been characterized throughout the years and discuss issues and paradoxes that led to a development of a new reasoning paradigm.

1.1 Outline

The thesis is organized as follows:

This introductory chapter presents the main research objectives of this thesis and provides all the necessary theoretical background.

When developing the benchmark, I selected five existing theories, two from the deductive paradigm and three from the probabilistic one. I present their theoretical background and assumptions in **Chapter 2**.

Additionally, I selected eight experiments whose data is used in the benchmark. All of them focus on manipulating different reasoning properties and effects. In **Chapter 3** I first present those properties and give examples. Afterwards, the motivation behind each experiment and its methods are covered.

Once all the relevant theories and data is introduced, I describe the benchmark I developed in **Chapter 4**. I explain the in-depth details of how the theories are implemented and provide their modeling equations. Afterwards I list the used evaluation methods. I use three different *goodness-of-fit* measures when answering RQ 2 and RQ 3. For RQ 4 I use statistical analysis methods for determining significance of value changes and correlation. The chapter ends with a brief description of the general functionality of the benchmark program.

Following is **Chapter 5** where I formally propose the new cognitive computational model ϵ-MMT. I provide the relevant definition from Pearl's (1991) ϵ-semantics and explain how it is applied to the mental model representation. Additionally, I discuss

ϵ-MMT's *possible worlds* interpretation along with the abandonment of the material implication.

I show the results of the benchmark analysis in **Chapter 6**. First, I analyze the overall models' performances. Those results give us an answer to the question which model has the best capability to adapt to various data. Additionally, we get a glimpse of ϵ-MMT's competitiveness among state-of-the-art models. Afterwards, I examine each experiment separately in two stages. I start with a comparison of all models' performances on one experiment only. With that, I can determine the circumstances that give certain models potential (dis-)advantages. The second stage is comprised of statistical analysis of ϵ-MMT's parameter values aiming to answer RQ 4 and learn more about human reasoning.

I extensively discuss and interpret the meaning of all analysis results in **Chapter 7**. Additionally, I also consider relevant general limitations in this field and propose ideas for future work.

And, finally, I conclude this thesis in **Chapter 8** by providing answers to the research questions.

1.2 Conditionals

Conditionals are statements usually of the form "If X then Y" (also written as X$\rightarrow$ Y, where X is called the *antecedent* and Y, the *consequent*), used to describe a causal relationship between two propositions X and Y.

Given a conditional (also called a *major premise*) and the current state of a proposition (called a *minor premise*), a *conclusion* can be inferred about the state of the other proposition. There are four inference forms, as shown in Table 1. Example 1 shows the Modus Ponens inference form. Table 2 shows the so-called converse inferences (Oaksford, Chater, & Larkin, 2000). They have the same minor premise as the corresponding original inference form, but the polarity of the conclusion is inverted, i.e. its negation is taken.

Taking the traditional logical interpretation of conditionals into consideration, the MP and MT inference forms are logically valid inferences, whereas AC and DA are not and are often referred to as *fallacious* (Singmann & Klauer, 2011).

Example 1. If the number on the card is 3, then the card is colored red.
The number on the card is 3.

Therefore, the card is colored red.

Table 1: Inference Forms

Inference Form	Conditional	Minor Premise	Conclusion
Modus Ponens (MP)	X→Y	X	Y
Modus Tollens (MT)	X→Y	¬Y	¬X
Affirmation of the Consequent (AC)	X→Y	Y	X
Denial of the Antecedent (DA)	X→Y	¬X	¬Y

Table 2: Converse Inference Forms

Converse Inference Form	Conditional	Minor Premise	Conclusion
Modus Ponens (MP')	X→Y	X	¬Y
Modus Tollens (MT')	X→Y	¬Y	X
Affirmation of the Consequent (AC')	X→Y	Y	¬X
Denial of the Antecedent (DA')	X→Y	¬X	Y

1.3 Characterization of Conditionals and Reasoning Paradigms

Conditionals are sentences made up of propositions, to which truth values can be assigned. That led to a basic characterization of conditionals, sometimes called *material implication* and it is logically equivalent to ¬X ∨ Y. The truth states of material implication are shown in the "Material" column in Table 3 (Manktelow, 1999). These conditionals are also called *truth functional conditionals.* As discussed by Evans and Over (2004), the invalidity of this approach leads to certain 'paradoxes':

P1: Given ¬X, it follows that if X then Y.
P2: Given Y, it follows that if X then Y.

These inferences are valid according to Table 3, yet it can be absurd to assume that they are valid for conditionals in natural language. Evans and Over (2004) provide an example of this invalidity using the conditional "If it rains, the plants will die.". When applying **P1** to this conditional, the following is obtained:

It will not rain.
Therefore, if it rains, then the plants will die.

In the case that there has been a drought, which leads the plants to die, the conditional would not be true, because it would not rain. However, treating this conditional as a truth functional conditional would justify accepting the conditional as true, on the basis that there is a drought.

The same is valid in the case of applying **P2** to the conditional:

> The plants will die.
> Therefore, if it rains, then the plants will die.

By imagining the same situation that there is a drought, which would be the reason for the plants dying, again it does not make sense that the death of the plants would be a basis for accepting the conditional as true.

Table 3: Characterization of a conditional "If X then Y" with a truth-table

X	**Y**	**If X then Y (Material)**	**If X then Y (Stalnaker)**	**If X then Y (Defective)**
True	True	True	True	True
True	False	False	False	False
False	True	True	True/False	Irrelevant
False	False	True	True/False	Irrelevant

Furthermore, consider the conditional "If the moon is blue, then the moon is green.". Another issue with material implication when considering everyday natural language conditionals is that Table 3 shows that this conditional is true, just because both the antecedent and the consequent are false, however it is obvious to any human individual that it is an absurd sentence, and no-one would endorse it (Oaksford & Chater, 2007).

Additionally, even more problems arise when taking the so-called "strengthening of the antecedent" into consideration. Given a conditional "If X then Y", then "If X and Z then Y" can also be concluded, which is mathematically acceptable. However, in the world of natural language conditionals this can lead to complications. Consider the conditional "If it's a bird then it flies". It should not allow an individual to then endorse that "If it's a bird and it's a penguin then it flies". However, if on the other hand we learn that it is a parrot, then the conclusion still holds. An important point is that whether the additional information Z has an effect on the conclusion or not is entirely content dependent (Oaksford & Chater, 2003b).

Psychologists and philosophers argue for (Grice, 1989; Johnson-Laird & Byrne, 1991, 2002) and against (Stalnaker, 1968; Rips & Marcus, 1977; Evans, Handley, &

Over, 2003; Over & Evans, 2003) the theory of natural language indicative conditionals being truth functional (Evans & Over, 2004).

Evans and Over (2004) discuss early advancements towards not viewing indicative conditionals as truth functional, as primarily done by Ramsey (1931/2013). He suggests that when people are reasoning about a conditional "If X then Y", they are actually trying to judge the probability of Y given X, i.e. $P(Y|X)$, widely known as *The Ramsey Test*. More specifically, the 'test' states that humans add X to their knowledge, and then argue about Y based on that. Stalnaker (1968) has also extended this test stating that before adding X to the knowledge, humans might need to modify their assumptions and belief by taking into consideration previous knowledge they might have on the topic, e.g., by believing $\neg X$. That would lead to assigning a very low value to $P(Y|X)$. These approaches eliminate the paradoxes **P1** and **P2**. Stalnaker (1968, 1975) developed a representation of conditionals differing from the truth functional conditional representation of "If X then Y" in the cases where X is false. He describes the possible truth states of X and Y as *possible worlds*. He then discusses the idea of 'closeness' between two worlds. For instance, if we have $\neg X$ Y, he states that the conditional might be true or it might also be false (contrary to the definite true as given by the material implication) – it would be true if the world where X and Y both happen is a closer possibility to $\neg X$ Y than X $\neg Y$. In order to understand this notion of 'closeness' better, we will go back to the example "If it rains, then the plants will die" and imagine we are in a state that "it will not rain" ($\neg X$) and "the plants will die" (Y). Now, consider the worlds X $\neg Y$ and X Y. Suppose that there is a drought. In that situation X $\neg Y$ means that rain will end the drought and the plants will not die, and then the conditional will be false. However, if the ground is waterlogged, the world X Y would mean that if it rains the plants will die which makes the conditional true. This characterization relaxes the prerequisites for a conditional to be considered true, in comparison to the material implication. It also shows that in some cases, the contents of the rule would have a bigger influence on the perceived truth state of the conditional, even if it is not logically valid.

Additionally, Wason (1966) suggests that when judging whether a conditional is true or not, the cases where the antecedent is false are irrelevant. That led to what he called a 'defective' truth table as shown in the "Defective" column in Table 3. This is also referred to as a *de Finetti* table, after de Finetti (1936/1995), who is the first one to construct it in a major theory of the conditional (Baratgin, Over, & Politzer, 2013). The 'defective'/*de Finetti* truth table has been corroborated through experiments where participants were either provided with truth value combinations for the propositions and asked to decide which ones render the conditional true or false, or they were asked to construct them themselves (Johnson-Laird & Tagart, 1969; Evans, 1972).

As pointed out by Manktelow (2012), tasks related to truth-tables, in general, have not been as beneficial to conditional reasoning research the way that inference tasks have. Nevertheless, both of them carry great importance towards understanding how humans interpret conditionals.

In the 1960s a *deductive paradigm* was developed, which assumed that logic accounts for rational reasoning, and the truth functional conditional was valid (Evans & Over, 2004). Cognitive models developed during that time focus on binary truth and truth preservation from assumptions (Elqayam & Over, 2013). When reasoning, however, humans do not always follow the norms provided by standard deductive logic. Their conclusions can depend on a lot more than logical validity of an inference. In the real world, inferences are thought of as uncertain and depend on judgements of probability, therefore psychological theories should provide an account of the *subjective probability* of conditionals (Over, 2004). This realization leads to the development of a new *probabilistic reasoning paradigm.*

In experiments about reasoning with conditionals, participants are usually given a conditional (major premise), a minor premise and then presented with a conclusion, according to the relevant inference form. In the old, deductive paradigm participants were asked if they accept an inference form or not. The conditionals' contents are usually abstract, therefore the individual would not have any background knowledge about the topic. However, in the new, probabilistic paradigm they are given more realistic every-day content and are asked about their endorsements of the inference forms, i.e., probabilities, while being encouraged/allowed to take their background knowledge into consideration (Singmann & Klauer, 2011). Simultaneously, the endorsement of inference forms in the probabilistic paradigm is assumed to be proportional to the conditional probability of the conclusion given the minor premise, as expressed with Bayes' theorem (Eq. 1) (Oaksford et al., 2000). Thus the alternative name for the paradigm is *Bayesian.*

$$\mathrm{P}(\beta|\alpha) = \frac{\mathrm{P}(\alpha|\beta) \cdot \mathrm{P}(\beta)}{\mathrm{P}(\alpha)} \tag{1}$$

Often, experiments manipulate contents and reasoning properties with the goal to clearly demonstrate how participants' endorsements can change, sometimes even drastically. In order to account for these kinds of effects that promote 'illogical' reasoning, models with a probabilistic approach seem to be more promising, as argued by Oaksford and Chater (2007), who provide a substantial discussion on how much deduction there really is in everyday human reasoning.

2 Existing Models and State of the Art

Both reasoning paradigms are comprised of a vast amount of cognitive models. I have selected five existing theories whose performances are analyzed in the benchmark. Their theoretical background follows in this chapter. The deductive paradigm is represented by the Mental Model Theory (Johnson-Laird & Byrne, 1991) and the Suppositional Theory (Evans & Over, 2004). Both of them have multiple formalization variants that challenge some of their assumptions, as presented further in this chapter. Even though these two models belong to the *old*, deductive paradigm, they are still highly relevant due to their psychological assumptions. Consequently, they have potential to bring us a step closer to resolve the mystery of the human reasoning process. Nevertheless, taking into consideration the uncertainties in our everyday world, the individuals' 'irrationality' and occasional inability to be logical, a better account is provided by the *state-of-the-art* models belonging to the Bayesian paradigm (Oaksford & Chater, 2007). In this thesis the probabilistic paradigm is represented by the remaining three models. First is the Oaksford et al.'s (2000) Probabilistic Model (from now on referred to as the "Oaksford-Chater (Probabilistic) Model"), which is one of the most influential Bayesian cognitive models (Singmann et al., 2016). Following is an extenstion of Spohn's (2009) ranking theory with logistic regression – the Logistic Regression Model (Skovgaard-Olsen, 2016). Finally, the Dual-Source Model (Klauer, Beller, & Hütter, 2010) is presented, which aims to disentangle the conditionals' form and content and builds up on the Oaksford-Chater Model. All of these theories and their variants are implemented in the benchmark, as presented in Chapter 4.

2.1 Mental Model Theory

> "To deduce is to maintain semantic information, to simplify, and to reach a new conclusion." (Johnson-Laird & Byrne, 1991, p. 22)

In simple words, the Mental Model Theory (MMT) assumes that when individuals are presented with some information, they build a mental representation of it using *mental models*. They aim to reach a conclusion based on the maintained information, and often, individuals would engage in a search for counterexamples to the conclusion. If their search is successful, they would no longer accept the conclusion (Khemlani & Johnson-Laird, 2012).

A *mental model* consists of the truth states of the propositions in the premise. Given a conditional premise "If X then Y", the initial mental model that an individual would construct is the one where both propositions are true, i.e., X Y.

The Mental Model Theory assumes that once the initial model is created it triggers the recollection of relevant facts and knowledge (Johnson-Laird & Byrne, 1991). Those facts can either serve as evidence that the initial model is correct or will stimulate a search for alternatives and lead to extending the mental model representation in a second process.

The extended mental model representation is also called a fleshed-out representation. It contains models that describe cases where X is false (written as ¬X), as shown in Table 4. The fleshed-out representation consists of all the possible combinations of truth-values for the propositions X and Y for which the conditional "If X then Y" is true. Johnson-Laird and Byrne (2002) call representing only what is true and not false the *principle of truth*. This coincides with the material implication definition which is the leading interpretation of conditionals in the deductive paradigm (Elqayam & Over, 2013). However, by claiming that conditionals are truth functional, then we accept as valid the paradoxes stemming from the material implication, which I discussed in Section 1.3. Johnson-Laird and Byrne (1991, 2002) state that they accept these paradoxes and that they are only an apparent problem for MMT (Over, 2004).

Table 4: Mental Model representation of a conditional premise "If X then Y"

Premise	**Mental Model**	**Fleshed-out Models**
If X then Y	X Y	X Y
	...	¬X ¬Y
		¬X Y

Schroyens, Schaeken, and d'Ydewalle (2001) revise the model (not the theory) presented by Johnson-Laird and Byrne (1991). They present an alternative model within the mental models approach, focusing on the stage where individuals validate the conclusion based on their mental model representation. They assume that when looking for counterexamples individuals do not construct all possible alternative models but rather aim for the model that would falsify the inference. Based on the original model's theory, when looking for counterexamples individuals only test whether alternative models are consistent with the premises. Schroyens et al. (2001), on the other hand, take background knowledge about the content into consideration and suppose that counterexamples are retrieved from long-term memory, which is where this theory stops being a theory of deductive reasoning (Oberauer, 2006). With this modification the material implication interpretation of conditionals within the theory is abandoned because now the X ¬Y model can actually be retrieved from long-term memory and become a part of the mental model representation. That leads to the possibility of refuting the logically valid MP and MT inferences.

Oberauer (2006) has provided a formalization of the Mental Model Theory as a multinomial processing tree (MPT), as shown in Fig. 1, built directly on the equations provided by Schroyens et al. (2001). The MPT can be interpreted as a binary decision diagram with parameters on the edges. In this case, the decisions are whether a human reasoner will add a model to their representation, or not, and the parameters describe the probability of a model being added.

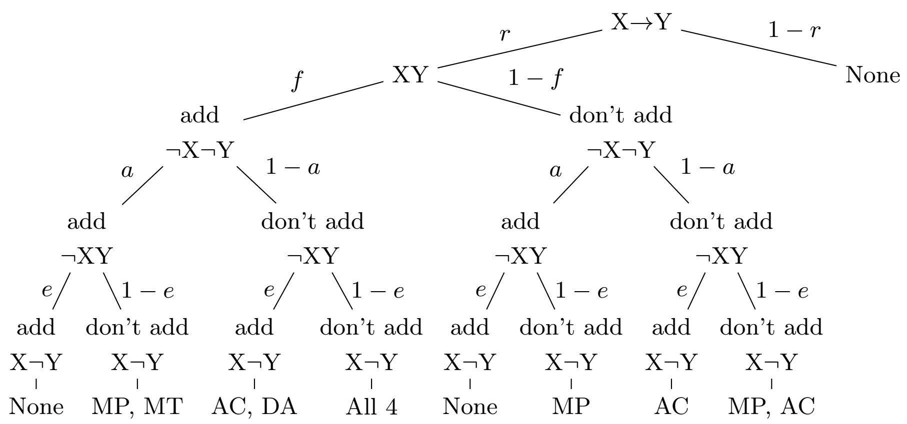


Figure 1: Oberauer's (2006) formalization of the MMT for the conditional "If X then Y". The parameters r, f, a, e take on values in the interval $[0, 1]$, indicating the probability of taking the respective decision path in the model. The leafs represent the responses.

If a special case of the MPT shown in Figure 1 is used, where the parameter e is set to 0, a *strictly deductive variant* of the MMT can be achieved, where the X ¬Y model is impossible to be added to the mental representation. From now on, this variant will be referred to as "MMT Deductive".

Evans (1993) criticized the Mental Model Theory for its lack of directionality.

Specifically, comparing the differences between reasoning problems involving "If X then Y" and the ones involving "X only if Y". Namely, people are more likely to draw backward inferences (inferences from Y or ¬Y as a minor premise) in the "only if" case and they are more likely to draw forward inferences (inferences from X or ¬X as a minor premise) in the case of "if...then". Following this observation, Oberauer (2006) provided a second formalization of the MMT including directionality. From this point onward, this variant will be referred to as "MMT with Directionality". Assuming a directionality from the antecedent variable to the consequent variable, the only reachable conclusions are the ones following that direction. For example, having only the model X Y in the mental representation supports concluding Y from X (MP), but not X from Y (AC). By expanding the previous formalization (Figure 1) with one more parameter d, this directionality limitation can be overcome. The MMT with Directionality MPT is shown in Figure 2.

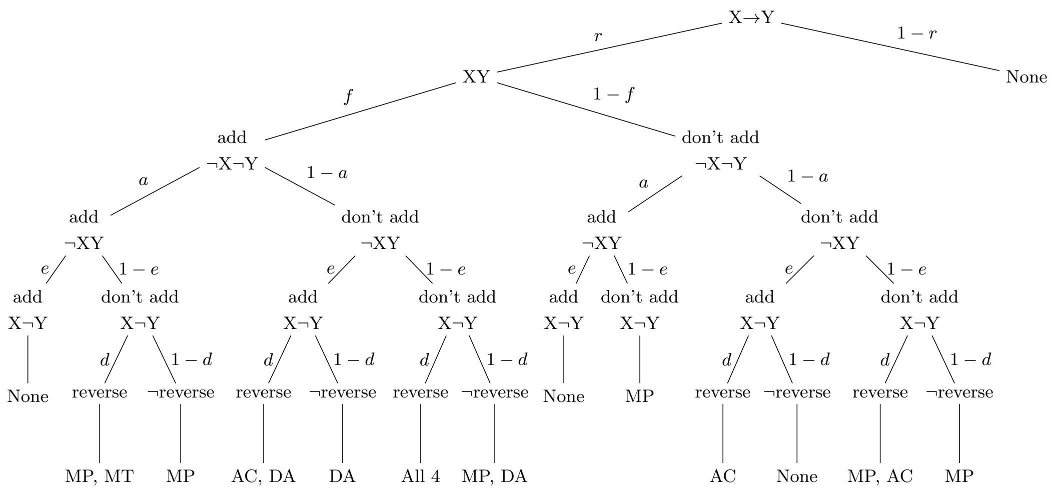


Figure 2: Oberauer's (2006) formalization of the MMT with Directionality for the conditional "If X then Y". The parameters r, f, a, e, d take on values in the interval $[0, 1]$, indicating the probability of taking the respective decision path in the model. The leafs represent the responses. "reverse": Reversing the directionality; "¬reverse": Not reversing the directionality.

2.2 Suppositional Theory

The Suppositional Theory (Evans & Over, 2004) assumes that two systems of reasoning are used in order to make a conclusion for a given reasoning problem, called System

1 and System 2. System 1 describes the heuristic or pragmatic inferences, whereas System 2 is dedicated to abstract rule-based reasoning (Evans & Over, 2004). The influence of each system depends from one individual to another in terms of belief-based vs. logic-based responses. System 1 is capable of only accepting MP inferences, unless additional conditional premises are added through pragmatic implicature. System 2, on the other hand, can accept MT as well by a *suppositional line of proof* (Oberauer, 2006).

The suppositional proof starts by supposing that X is true, which combined with the given conditional rule, “If X then Y”, leads to the conclusion that Y is true (MP inference). However, this contradicts the minor premise of MT, ¬Y, therefore the supposition must be false, so ¬X must be true. This leads to the acceptance of the MT inference form. Starting the suppositional proof with Y, combined with the converse, “If Y then X” (assuming that the converse of the conditional has been added by pragmatic implicature), leads to accepting the DA inference form.

Oberauer (2006) has also provided a formalization of the Suppositional Theory. Evans and Over (2004) do not explicitly specify how the two systems, System 1 and System 2, interact, so Oberauer (2006) provided two different variants of the formalization of the theory: Sequential and Exclusive.

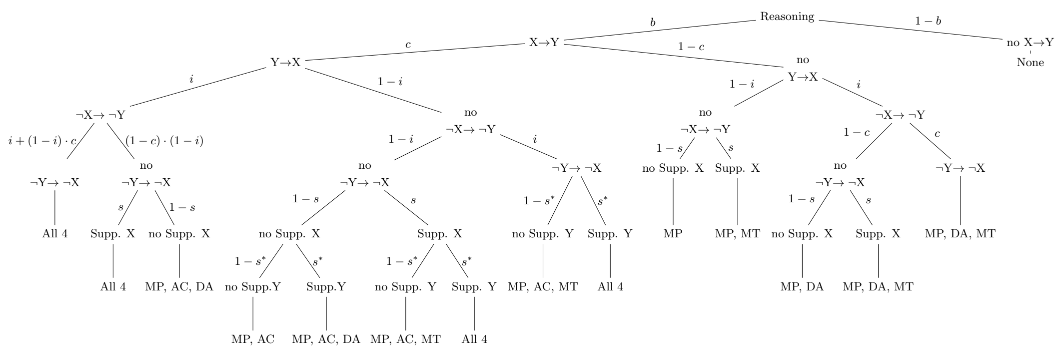


Figure 3: Oberauer’s (2006) formalization of the Suppositional Theory (Sequential Variant) (Oberauer, 2006) for the conditional “If X then Y”. The parameters b, c, i, s, s^* take on values in the interval $[0, 1]$, indicating the probability of taking the respective decision path in the model. The leafs represent the responses.

In the Sequential Variant (Figure 3), as the name suggests, it is assumed that the two systems work sequentially. After the individual determines their degree of belief in the conditional X→Y, System 1 takes over and has the possibility of adding the converse, Y→X, and inverse, ¬X→ ¬Y by pragmatic implicature. By combining the two, it can arrive to the contraposition, ¬Y→ ¬X. Once System 1 is done with adding conditionals, it derives all the conclusions following from them. Next, System 2 takes the conditionals System 1 has added and with a certain probability, applies the procedure for suppositional proof. Two different suppositional proofs can take place. One supposes the truth of X and the original conditional and the other supposes the truth of Y and the converse.

In the Exclusive Variant (Figure 4), it is assumed that the two systems work independently. After the individual determines their degree of belief in the conditional, they will reason using either System 1 or System 2. If System 1 is chosen, the conditionals are added via pragmatic implicature, and finally the conclusions are derived from them. If System 2 is chosen, the approach is strictly deductive – MP is derived from the conditional, and an attempt for deriving MT by a suppositional proof is made.

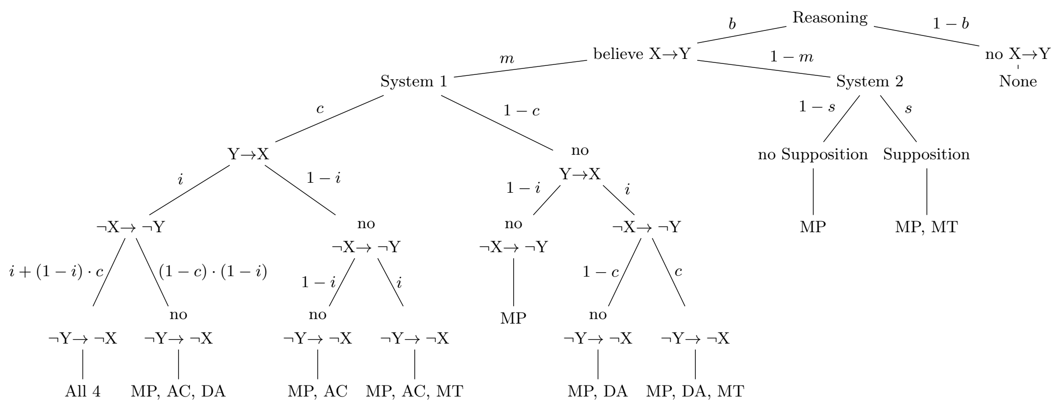


Figure 4: Oberauer's formalization of the Suppositional Theory (Exclusive Variant) (Oberauer, 2006) for the conditional "If X then Y". The parameters b, m, c, i, s take on values in the interval $[0, 1]$, indicating the probability of taking the respective decision path in the model. The leafs represent the responses.

2.3 Oaksford-Chater Probabilistic Model

Oaksford et al. (2000) propose a probabilistic computational level model (Marr, 1982) for conditional reasoning. By using a 2 × 2 contingency table, as in Table 5, they represent conditional rules, where, a = P(X) and b = P(Y), probabilities of the antecedent and consequent, respectively and $\epsilon = P(\neg Y|X)$ is the exception parameter.

Table 5: Contingency table for a conditional rule "If X then Y". $a = P(X)$ - Probability of the antecedent; $b = P(Y)$ - Probability of the consequent; $\epsilon = P(\neg Y|X)$ - Probability of the exception.

	Y	**¬Y**
X	$a(1-\epsilon)$	$a\epsilon$
¬X	$b - a(1-\epsilon)$	$(1-b) - a\epsilon$

Their model belongs to the Bayesian paradigm as it assumes that inference form endorsement is proportional to the conditional probability of the consequent given the antecedent. So, from Table 5, they derived expressions for the inference form endorsements as follows:

$$\textbf{MP:} \quad P(Y|X) = 1 - \epsilon \qquad \textbf{DA:} \quad P(\neg Y|\neg X) = \frac{1 - b - a \cdot \epsilon}{1 - a}$$

$$\textbf{AC:} \quad P(X|Y) = \frac{a(1-\epsilon)}{b} \qquad \textbf{MT:} \quad P(\neg X|\neg Y) = \frac{1 - b - a \cdot \epsilon}{1 - b}$$

2.4 Logistic Regression Model

Skovgaard-Olsen (2016) proposes a Logistic Regression Model by taking Spohn's (2009) ranking theory and extending it using logistic regression. For the theoretical background on ranking theory, logistic regression and the relation between the two, see Appendix A.2.

Following the assumption that inference form endorsements are expressed through conditional probability of the conclusion given the minor premise, regression lines can be created to which the data can be fitted. However, the regression lines would differ depending on whether X is the predictor of Y (MP, DA), or Y is the predictor of X (AC, MT). As shown in Table 6, their slopes are identical, but, their intercepts differ.

Table 6: Logistic Regression Model parameters

	X as a predictor of Y	**Y as a predictor of X**
Intercept	$e^{b_0} = \frac{P(Y\|\neg X)}{P(\neg Y\|\neg X)}$	$e^{b_0^*} = \frac{P(X\|\neg Y)}{P(\neg X\|\neg Y)}$
"Slope"	$e^{b_1} = \frac{P(Y \wedge X)}{P(\neg Y \wedge X)} \cdot \frac{P(\neg Y \wedge \neg X)}{P(Y \wedge \neg X)}$	

The following equations are formulated and used as predicting expressions for inference form endorsements:

$$\textbf{MP:} \quad \text{P(Y|X)} = \frac{1}{1+e^{-(b_0+b_1)}} \qquad \textbf{DA:} \quad \text{P}(\neg\text{Y}|\neg\text{X}) = \frac{1}{1+e^{b_0}}$$

$$\textbf{AC:} \quad \text{P(X|Y)} = \frac{1}{1+e^{-(b_0^*+b_1)}} \qquad \textbf{MT:} \quad \text{P}(\neg\text{X}|\neg\text{Y}) = \frac{1}{1+e^{b_0^*}}$$

2.5 Dual-Source Model

The Dual-Source Model (DSM) (Klauer et al., 2010; Singmann et al., 2016), is an extension of the Oaksford-Chater Probabilistic Model (Oaksford et al., 2000). It assumes that individuals integrate two different kinds of information: background knowledge about the content and information related to the logical form of the inference. Taking that into consideration, the Dual-Source Model uses three types of parameters:

- $\xi(C, x)$ - knowledge-based component, depending on the content C and inference x
- $\tau(x)$ - form-based component reflecting the subjective degree of belief in the inference x
- λ - a weight given to the form-based component (integrating $\xi(C, x)$ and $\tau(x)$ using Bayesian model averaging).

A *reduced inference* is when an individual is *not* given a conditional rule, but only a minor premise and asked for the conclusion's endorsement. The DSM expresses this inference through the knowledge-based component $\xi(C, x)$:

$$E_r(C, x) = \xi(C, x)$$

Then, the endorsement of the full inference x with content C is given by:

$$E_f(C, x) = \lambda\{\tau(x) + (1 - \tau(x)) \cdot \xi(C, x)\} + (1 - \lambda)\xi(C, x)$$

The $\xi(C, x)$ parameters are obtained by using Oaksford et al. (2000)'s equations, as follows:

$$\textbf{MP:} \quad \xi(C, \text{MP}) = 1 - \epsilon \qquad \textbf{DA:} \quad \xi(C, \text{DA}) = \frac{1 - b - a \cdot \epsilon}{1 - a}$$

$$\textbf{AC:} \quad \xi(C, \text{AC}) = \frac{a(1 - \epsilon)}{b} \qquad \textbf{MT:} \quad \xi(C, \text{MT}) = \frac{1 - b - a \cdot \epsilon}{1 - b}$$

In Singmann, Klauer, and Over (2014) an alternative approach for the DSM is proposed, along with a presentation of two experiments. One of the experiments uses abstract contents and asks participants for inference form endorsements. The form-based component $\tau(x)$ is estimated by averaging the individual answers, per participant. The second experiment uses everyday conditionals and asks participants for a rather large variety of subjective probabilities. Their answers are then used to determine the knowledge-based components $\xi(C, x)$. This leaves the model with only one free parameter, λ. However, in this approach, the components are estimated by using equations describing Law of Total Probability, e.g., for MP:

$$P(Y) = P(Y|X) \cdot P(X) + P(Y|\neg X) \cdot (1 - P(X)) \tag{2}$$

In this approach the whole idea of representing inference form endorsements with conditional probability is completely abandoned. When presenting the DSM once again in Singmann et al. (2016), they return to the original idea of the model being an extension to Oaksford et al. (2000)'s probabilistic model, using conditional probability, as explained above.

In general, unfortunately, it is highly unlikely to find other experiments that ask for all of the necessary subjective probabilities in order to be able to apply this model. As a matter of fact, to my knowledge, no other such data exists (yet). Following personal communication with Henrik Singmann about the Dual-Source Model in general and the application and validity of this alternative approach, I decided that it will not be used in this thesis due to lack of credibility in the alternative representation of the endorsements.

3 Data

In order to analyze the models' performance on various data, I have chosen experiments that exploit different effects and reasoning properties. In this chapter I first explain the manipulated properties. Afterwards, I present the motivation behind the experiments and their goals and methods. All of their data is included in the benchmark and the analysis results are presented in Chapter 6.

3.1 Reasoning Properties and Effects

3.1.1 Suppression effect

Deductive reasoning is truth-preserving – given premises that hold, the conclusion must also hold. That means if any additional information would be added, i.e., the already existing premises would be conjoined with additional premises, the conclusion would still hold. This is called 'monotonicity'. 'Non-monotonicity', on the other hand, is when additional information prevents the conclusion from still being valid.

A popular example of non-monotonicity is the *suppression effect*. Formal theorists (Markovits, 1984, 1985; Rumain, Connell, & Braine, 1983) have shown that individuals can be forced to reject the two fallacious inference forms, DA and AC, when a conditional rule is accompanied with *alternative* antecedents. As the name suggests, alternative antecedents are events different from the antecedent in the original major premise that also allow for the consequent to be true. Consider Example 2 of the DA inference form where the conclusion follows from the major and minor premise. However once an alternative antecedent is introduced (the colored sentence), the conclusion no longer holds, i.e., the DA inference has been suppressed.

Example 2. If she meets her friend then she will go to a play.
If she meets her family then she will go to a play.
She doesn't meet her friend.

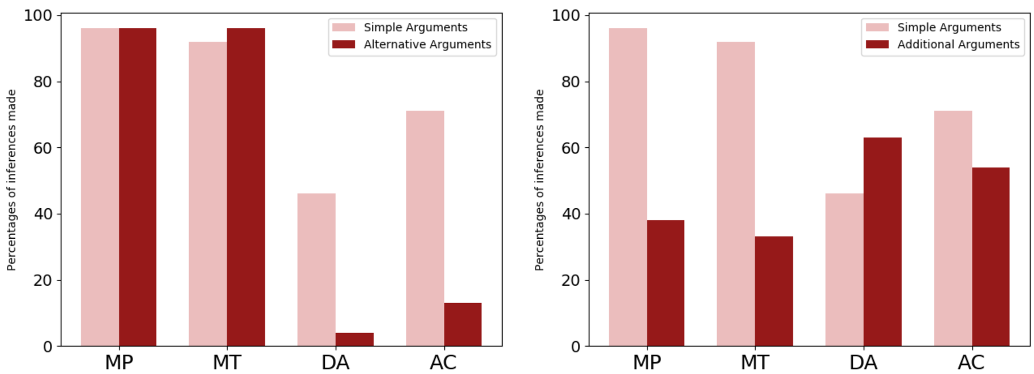
She will not go to a play.

Byrne (1989) then shows that in a similar way, the valid inferences, MP and MT, can also be suppressed by providing *additional* antecedents. They provide constraints that might influence the effect of the major premise's antecedent and would prevent the consequent from coming true. Example 3 shows the MP inference form, where the conclusion follows given the major and minor premises. However by providing an additional antecedent (the colored sentence), the conclusion does not hold anymore, i.e., the MP inference has been suppressed.

Example 3. If she has an essay to write then she will study late in the library.
If the library stays open then she will study late in the library.
She has an essay to write.

She will study late in the library.

Byrne (1989) showed the suppression effect on all inference forms in an experiment where participants were divided in three groups – in the first one they were presented with alternative antecedents, in the second one with additional antecedents, and the last group did not get any extra information. The mean percentages of inferences made by all three groups are shown in Figure 5, illustrating the suppression effect.



(a) Suppression of the fallacious inference forms (DA, AC) by providing alternative arguments.

(b) Suppression of the valid inference forms (MP, MT) by providing additional arguments.

Figure 5: Suppression Effect (Byrne, 1989)

Additionally, there are further reasoning properties closely related to the suppression effect. They are also taken into consideration in this thesis and are explained in the following sections.

Varying Amounts of Disablers and Alternatives

Given a conditional "If X then Y", *disablers* are events that prevent Y from happening, even though X has occurred (expressed through the previously called *additional antecedents*). *Alternatives*, on the other hand, are events that enable Y to happen, even though X has not occurred (expressed through the previously called *alternative antecedents*). Example 4 demonstrates an instance of both.

Example 4. *Conditional:* If the air conditioner is turned on, then you feel cold.
Disabler: You are wearing a very thick winter jacket.
Alternative: It's winter and your window is open.

Conditional contents can vary in the amount of disablers and alternatives related to them, influencing the individuals' confidence in a conclusion. In this thesis, the contrast is between 'Few' and 'Many' disablers/alternatives. This way of quantifying the amount of disablers/alternatives can be thought of as a simplified or naïve way. Additionally, it is not necessarily true that the subjective understanding of 'Many' for one conditional is the same as for another one, but it would be rather relative. Simultaneously, it can obviously not be guaranteed that how one individual perceives 'Many' or 'Few' would match the opinion of another one. However, in terms of available data and research, it is a good starting point.

In some of the experiments taken into consideration, conditionals with few disablers and many alternatives are also called *prological conditionals*. On the other hand, conditionals with many disablers and few alternatives are called *counterlogical conditionals*. Conditionals with both many alternatives and disablers are called *neutral conditionals*.

Validity and Plausibility Effect

MP and MT are logically valid inference forms, and AC and DA are not. Given a prological conditional, both MP and MT are also plausible. On the other hand, when given a counterlogical conditional, validity and plausibility are pitted against each other, i.e., the valid inference forms MP and MT become implausible, but the invalid ones, AC and DA, become plausible.

A stronger endorsement for logically valid but implausible inferences is called the *validity effect*. A stronger endorsement for logically invalid but plausible inferences is called the *plausibility effect* (Singmann & Klauer, 2011).

3.1.2 Conditional Presentation Form

The presentation form of an inference task can affect how much people would rely on their personal background knowledge, compared to being highly influenced by the content of what is presented to them. Here, three different presentation forms are relevant: *reduced inference*, *conditional* and *biconditional*.

In the reduced inference case, no rule/major premise is presented, only the minor premise, followed by the conclusion, as shown in Example 5 (Singmann et al., 2016).

Example 5. This person drinks a lot of coke.

Therefore, the person will gain weight.

The conditional case is the "classical" inference task presentation, where a conditional rule is presented as a major premise, followed by the minor premise and the conclusion, as shown in Example 6 (Singmann et al., 2016).

Example 6. If a person drinks a lot of coke, then the person will gain weight.
This person drinks a lot of coke.

Therefore, the person will gain weight.

The biconditional case, as the name suggests, is a presentation form where the major premise is a biconditional ("If and only if X then Y."), followed by the minor premise and the conclusion, as shown in Example 7 (Singmann et al., 2016).

Example 7. If and only if a person drinks a lot of coke, then the person will gain weight.
This person drinks a lot of coke.

Therefore, the person will gain weight.

3.1.3 Deductive vs. Inductive Instructions

The fact that human reasoning does not conform to the standard deductive logic rules has already been discussed extensively in Section 1.3. When individuals receive *deductive* instructions they are asked to judge the conclusion from a *logical validity* perspective ("How valid is the conclusion?"). In the case of *inductive* instructions they are asked to judge the *general likelihood* of a conclusion ("How likely is the conclusion?"). By manipulating the instructions the contrast between deductive and inductive reasoning is shown.

3.1.4 Speaker Expertise

Depending on the content of a conditional, individuals incorporate their background knowledge on the relevant topic when reasoning. Some experiments manipulate the degree to which an individual is expected to rely on their background knowledge by stating that a conditional is uttered by an expert or a non-expert, as shown in Examples 8 and 9 (Singmann et al., 2016). Generally, it is expected that when a conditional is uttered by a non-expert, individuals would rely on their background knowledge more, compared to when the conditional would be uttered by an expert.

Example 8. A nutrition scientist says: If Anne eats a lot of parsley then the level of iron in her blood will increase.

Example 9. A drugstore clerk says: If Anne eats a lot of parsley then the level of iron in her blood will increase.

3.2 Experiments

3.2.1 Experiments 1–4

These four experiments are presented in Singmann et al. (2016). The goal of their work was to provide validation of the psychological interpretation of the Dual-Source Model's parameters. Additionally they aim to show that while the usage of background knowledge when reasoning is supported, the conditional rule's form also plays a considerable role, irrespective of the content.

The motivation behind the experiments was to implement manipulations that would influence the Dual-Source Model's parameter values in a way that would aid in the psychological interpretation validation.

For each experiment there are also two control groups (knowledge control group and rule control group), to assess possible biases. The answers provided in the control experiments were not taken into consideration in this thesis.

The experiments' data files along with the authors' analysis scripts and the online appendix can be accessed at `https://osf.io/zcdfq/`.

Experiment 1

The main goal of this experiment was to contrast conditionals in their standard form, "If X then Y", with biconditionals, "If X then and only then Y", and demonstrate this

Table 7: Task example for Experiment 1 (Singmann et al., 2016)

Conditional Presentation Form	Content
Reduced Inference	A balloon is pricked with a needle. How likely is it that it will pop?
Conditional	If a balloon is pricked with a needle then it will pop. A balloon is pricked with a needle. How likely is it that it will pop?
Biconditional	If a balloon is pricked with a needle then and only then it will pop. A balloon is pricked with a needle. How likely is it that it will pop?

manipulation's effect on the Dual-Source Model's form-based component, the $\tau(x)$ parameter.

The varying factors in this experiment are the *conditional presentation form* and *varying amounts of disablers and alternatives.*

A total of 31 participants took part in this experiment. They gave estimates to four inference forms and their converses for four different contents and three presentation forms, making it a total of 96 answers per participant.

Table 7 provides an example of the question form that participants received in all three conditional presentation forms. The content has few disablers and many alternatives.

The contents of the conditionals used in this experiment can be found in Appendix A.1.1.

Experiment 2

The main goal of this experiment was to manipulate the expertise with which a conditional has been uttered. Singmann et al. (2016) demonstrate its effect on the Dual-Source Model's parameter λ, which is a weight given to the form-based component versus the knowledge-based one.

The varying factors in this experiment are *conditional presentation form* and *speaker expertise.*

A total of 47 participants took part in this experiment. They gave estimates to four inference forms and their converses for six different contents (randomly selected out

Table 8: Task example for Experiment 2 (Singmann et al., 2016)

Speaker Expertise	Content
Non-Expert	A drugstore clerk says: If Anne eats a lot of parsley then the level of iron in her blood will increase. Anne eats a lot of parsley. How likely is it that the level of iron in her blood will increase?
Expert	A nutrition scientist says: If Anne eats a lot of parsley then the level of iron in her blood will increase. Anne eats a lot of parsley. How likely is it that the level of iron in her blood will increase?

of the pool of seven) and two presentation forms, adding up to a total of 96 answers per participant.

Table 8 provides an example of the question form that participants received in both speaker expertise types.

The contents of the conditionals used in this experiment can be found in Appendix A.1.2.

Experiments 3 & 4

Experiments 3 and 4 are two independent replications of the same experiment. The main goal of these experiments is to suppress the inference form endorsements, i.e. to provoke the suppression effect, by providing additional information in the form of disablers and alternatives.

The varying factors in this experiment are *conditional presentation form*, *varying amount of disablers and alternatives* and *inducing the suppression effect*.

A total of 77 participants took part in Experiment 3 and 91 in Experiment 4. In both experiments, participants were divided in three groups: 'Baseline', where they responded to regular conditional tasks, 'Disablers' and 'Alternatives' where they were provided with additional disablers or alternatives, respectively. They all gave estimates to four inference forms and their converses for four different contents and two presentation forms, making it a total of 64 answers per participant.

Table 9 provides an example of the question form that participants received in the 'Disablers' and 'Alternatives' groups.

The contents of the conditionals used in these experiments can be found in Appendices A.1.3 and A.1.4.

Table 9: Task example for Experiment 3 & 4 (Singmann et al., 2016)

Group	Content
Disablers	If a person drinks a lot of coke then the person will gain weight. A person drinks a lot of coke. How likely is it that the person will gain weight? Please note: A person *only* gains weight if: - the metabolism of the person permits it, - the person does not exercise as a compensation, - the person does not only drink diet coke.
Alternatives	If a person drinks a lot of coke then the person will gain weight. A person drinks a lot of coke. How likely is it that the person will gain weight? Please note: A person *also* gains weight if: - the person eats a lot, - the person has metabolic problems, - the person hardly exercises.

3.2.2 Experiments 5–6

These two experiments are presented in Singmann and Klauer (2011). The goal of their work was to determine whether responses under deductive and inductive instructions can be explained by a single process or rather reflect two modes of conditional reasoning.

The motivation behind the experiments was to establish double dissociation between deductive and inductive instructions when validity and plausibility were pitted against each other, simultaneously aiming to induce the validity and plausibility effects.

In order to test whether the validity and plausibility effects hold, Singmann and Klauer (2011) examined the affirmation inferences (MP, AC) separately from the denial ones (MT, DA). Only counterlogical conditionals are taken into consideration. Under deductive instructions they expected that MP is endorsed more strongly than AC and MT is endorsed more strongly than DA (validity effect). Under inductive instructions, they expected that AC is endorsed more strongly than MP, and DA is endorsed more strongly than MT (plausibility effect).

When looking at the affirmation problems (MP, AC) they found that under deductive instructions participants showed stronger endorsements for the valid but implausible MP than the invalid but plausible AC. So, the validity effect holds. However, for the denial problems (MT, DA), the validity effect did not seem to hold, i.e. participants showed stronger endorsements for the invalid but plausible DA than the valid but implausible MT.

Under inductive instructions, the plausibility effect takes place in both affirmation and denial problems. Participants showed higher endorsements for the invalid but plausible AC than for the valid but implausible MP and for the invalid but plausible DA than for the valid but implausible MT.

In both experiments, some of the participants received deductive instructions and were asked to assume the truth of the premises and disregard background knowledge when judging the validity of the conclusion. The others received inductive instructions and were encouraged to use background knowledge when judging the probability of the conclusion. Additionally, it has been explicitly stated that the conditional rules have to be perceived as unidirectional.

Experiment 5

In this experiment participants were provided with two prological conditionals (few disablers and many alternatives). In order to pit validity and plausibility against each other, the authors constructed counterlogical conditionals by reversing the antecedent and consequent in the prological conditionals.

The varying factors in this experiment are *prological vs. counterlogical conditionals* and *deductive vs. inductive instructions.*

A total of 40 participants took part in this experiment. They gave estimates to four inference forms for four different contents, adding up to a total of 16 answers per participant. Half of the participants (20) received deductive instructions and the other half (20) inductive.

Table 10 provides an example of the question form that participants received under deductive and inductive instructions.

The contents of the conditionals used in this experiment can be found in Appendix A.1.5.

Table 10: Task example for Experiment 5 & 6 (Singmann & Klauer, 2011)

Instructions	Content
Deductive	If a campfire goes out, then water has been poured on it. A campfire goes out. How valid is the conclusion that water has been poured on it?
Inductive	If a campfire goes out, then water has been poured on it. A campfire goes out. How likely is it that water has been poured on it?

Experiment 6

The goal of this experiment is to replicate the results of the previous experiment, and to rule out possible alternative explanations for the findings. This time, the contents for the prological and counterlogical conditionals are different, and furthermore, they also introduced neutral conditionals.

The varying factors in this experiment are *prological vs. counterlogical vs. neutral conditionals* and *deductive vs. inductive instructions.*

A total of 55 participants took part in this experiment. They gave estimates to four inference forms for nine different contents, adding up to a total of 36 answers per participant. 27 of the participants received deductive instructions and 28 inductive.

Table 10 provides an example of the question form that participants received under deductive and inductive instructions.

The contents of the conditionals used in this experiment can be found in Appendix A.1.6.

3.2.3 Experiments 7–8

These two experiments are presented in Singmann et al. (2014). The goal of their work was to present the results of an empirical test of normative standards in the new paradigm and examine which of the following three: conditional probability P(Y|X), conjunctive probability P(X∧Y) or probability of the material conditional P(¬X∨Y), provides unique variance to predicting the probability of the conditional P("If X then Y") .

The motivation behind the first experiment is to develop a novel probabilized conditional task which provides more insight into individuals' subjective beliefs compared to tasks focusing on inference form endorsements. The participants' answers are then used to obtain the knowledge-based components of the Dual-Source Model when using the alternative approach. The second experiment, on the other hand, presents a deductive task and is used for the form-based components of the Dual-Source Model in the alternative approach.

Experiment 7

This is the *probabilized conditional inference task* where participants were presented with highly-believable everyday conditionals and asked for various subjective probabilities. In total there were 16 conditionals, each participant worked on four randomly selected ones, and performed only one inference per conditional.

A total of 29 participants took part in this experiment. They gave estimates to 8 different probabilities for four different contents, making it a total of 32 answers per participant.

In this experiment, the correlations between parameter values and subjective probability estimates of the *conditional* and *material conditional* are examined.

Table 11 provides an example of the questions that participants had to answer for one task.

It is important to emphasize that in this article the authors differentiate between inference form endorsement and conditional probability and ask for both of them in a different way, as shown in Table 11. The approach I am taking in this thesis does not follow their assumptions, but rather follows the supposition that inference form endorsements are expressed as conditional probability of the conclusion given the minor premise, as it is most common in the Bayesian paradigm.

The contents of the conditionals used in this experiment can be found in Appendix A.1.7.

Experiment 8

This is the *deductive conditional inference task*. Participants were presented with a conditional with abstract content and then instructed to judge the logical validity of the arguments.

A total of 29 participants took part in this experiment. They gave estimates to four inference forms for two different contents, adding up to a total of 8 answers per participant.

Table 12 provides an example of the question form that participants received.

The contents of the conditionals used in this experiment can be found in Appendix A.1.8.

Table 11: Task example for Experiment 7 (Singmann et al., 2014)

Probability	Content
"If X then Y"	If Greece leaves the Euro then Italy will too. In your opinion, how probable is the above statement?
Minor Premise	Greece will leave the Euro. In your opinion, how probable is it that the above event occurs?
Inference Form	If Greece leaves the Euro then Italy will too. Greece will leave the Euro. Under these premises, how probable is that Italy will leave the Euro, too?
Conditional	How probable is that Italy will leave the Euro should Greece leave the Euro?
Conjunctive	Greece will leave the Euro and simultaneously Italy will leave the Euro. In your opinion, how probable is it that the above event occurs?
Material Conditional	Greece will NOT leave the Euro or Italy will leave the Euro. In your opinion, how probable is it that the above event occurs?
Alternatives	How probable is that Italy will leave the Euro should Greece NOT leave the Euro?
Conclusion	Italy will leave the Euro. In your opinion, how probable is it that the above event occurs?

Table 12: Task example for Experiment 8 (Singmann et al., 2014)

Content
If the letter is a B then the number is a 7. The letter is a B. How valid is the conclusion that the number is a 7 from a logical perspective?

4 Benchmark

This Chapter presents the benchmark I developed for this thesis. First, I give an overview of the implemented models' formalizations and the equations used for data fitting. All of the models' theoretical background was provided in Chapter 2. Following is a description of the evaluation methods the benchmark uses in different settings and how they were implemented. Finally, the chapter is concluded with a brief overview of the benchmark program functionality.

The implementation can be found at `https://gkigit.informatik.uni-freiburg.de/todoroviks/conditional-reasoning-benchmark`.

4.1 Implemented Models

All of the models presented in Chapter 2 are implemented in the benchmark. They are fit to data by optimizing their endorsement equations, to have a minimal RMSE (definition and equation in the following Section 4.2). The optimization is done with Python's `scipy.optimize.minimize`[1] using the SLSQP method which allows for constraints and bounds.

The first two theories are the Mental Model Theory and the Suppositional Theory. The formalizations used in this benchmark are provided by Oberauer (2006). All three MMT variants and the two Suppositional Theory ones are included. Originally, he applied those models to aggregated data, but they are easily applicable to individual data as well, as done in this benchmark.

Endorsement expressions can be easily derived from the MPTs provided by Oberauer (2006), which were depicted in Chapter 2. In the following I will provide an example of the derivations and then the final equations.

[1] `https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.minimize.html`

For example, in the case of **MMT**, let's focus on the following possible mental representation:

$$\begin{array}{rr} X & Y \\ \neg X & \neg Y \\ \neg X & Y \\ X & \neg Y \end{array}$$

Based on the MPT, this representation supports the acceptance of the MP and MT inference forms. The probability with which this decision path can be taken is $r \cdot f \cdot a \cdot (1 - e)$. The parameters describe the probabilities of adding models to the representation. r is the probability of adding the model X Y, f for the model ¬X ¬Y, a for ¬X Y, and finally, e for X ¬Y.

In order to derive endorsement expressions for each inference form separately, the probabilities of all possible paths leading to accepting that inference form will be summed up and simplified. Example 10 shows a detailed derivation of the expression representing the endorsement of the Modus Ponens (MP) inference form. Similarly, expressions for the other inference forms are derived. The complete derivations of endorsement expressions for this and all following models formalized using MPTs can be found in Appendix A.3.

Example 10. Deriving an expression representing endorsement of the Modus Ponens (MP) inference form based on MMT.

$$\begin{aligned} E(MP) =& [r \cdot f \cdot a \cdot (1-e)] + [r \cdot f \cdot (1-a) \cdot (1-e)] + \\ & [r \cdot (1-f) \cdot a \cdot (1-e)] + [r \cdot (1-f) \cdot (1-a) \cdot (1-e)] \\ =& r \cdot (1-e) \cdot [f \cdot a + f \cdot (1-a) + (1-f) \cdot a + (1-f) \cdot (1-a)] \\ =& r \cdot (1-e) \cdot [f \cdot (a + 1 - a) + (1-f) \cdot (a + 1 - a)] \\ =& r \cdot (1-e) \cdot [f \cdot 1 + (1-f) \cdot 1] \\ =& r \cdot (1-e) \cdot [f + 1 - f] \\ =& r \cdot (1-e) \cdot 1 \\ =& r \cdot (1-e) \end{aligned}$$

The final forms of the inference form endorsement expressions, according to **MMT**, as used in the implementation, are:

MP:	$r \cdot (1 - e)$	**DA:**	$r \cdot f \cdot (1 - a)$
AC:	$r \cdot (1 - a)$	**MT:**	$r \cdot f \cdot (1 - e)$

By setting the parameter e to 0, the strictly deductive variant (**MMT Deductive**) is obtained, leading to the following inference form endorsement expressions:

MP:	r	**DA:**	$r \cdot f \cdot (1 - a)$
AC:	$r \cdot (1 - a)$	**MT:**	$r \cdot f$

Finally, the last MMT variant is **MMT with Directionality** which uses the following derived inference form endorsement expressions:

MP:	$r \cdot (1 - e)$	**DA:**	$r \cdot f \cdot (1 - a)$
AC:	$r \cdot (1 - a) \cdot d$	**MT:**	$r \cdot f \cdot (1 - e) \cdot d$

The new parameter d describes the probability of reversing the directionality.

I derived the endorsement expressions for the Suppositional Theory variants using the same approach as for the MMT variants. For **Suppositional Sequential** the following expressions are used:

MP:	b	**DA:**	$b \cdot (c \cdot s^* \cdot (1 - i) + i)$
AC:	$b \cdot c$	**MT:**	$b \cdot ((1 - s) \cdot (2 \cdot c \cdot i - c^2 \cdot i^2) + s)$

The parameter b describes the belief in the conditional X→Y, c is the probability of the converse to be added (Y→X) and i, the inverse (¬X→ ¬Y). s is the probability of the suppositional proof using X and the original conditional to succeed and s^* for the proof using Y and the converse.

For **Suppositional Exclusive**, on the other hand, I derived the following endorsement expressions:

$$\textbf{MP:}\quad b \qquad \textbf{DA:}\quad b \cdot m \cdot i$$

$$\textbf{AC:}\quad b \cdot m \cdot c \qquad \textbf{MT:}\quad b \cdot (m \cdot c \cdot i \cdot (2 - c \cdot i) + (1 - m) \cdot s)$$

The new parameter m describes the probability of System 1 being used. The rest of the parameters are the same as in the Sequential variant, without the possibility for a suppositional proof using Y and the converse.

Following are the three models from the Bayesian paradigm. First, there is the Oaksford-Chater Probabilistic Model, followed by the Logistic Regression Model, and finally, the extension of Oaksford-Chater – the Dual-Source Model.

The endorsement expressions that the **Oaksford-Chater** model uses are the ones presented in Chapter 2:

$$\textbf{MP:}\quad \mathrm{P(Y|X)} = 1 - \epsilon \qquad \textbf{DA:}\quad \mathrm{P(\neg Y|\neg X)} = \frac{1 - b - a \cdot \epsilon}{1 - a}$$

$$\textbf{AC:}\quad \mathrm{P(X|Y)} = \frac{a(1 - \epsilon)}{b} \qquad \textbf{MT:}\quad \mathrm{P(\neg X|\neg Y)} = \frac{1 - b - a \cdot \epsilon}{1 - b}$$

The parameter a describes the probability of the antecedent, P(X), b is the probability of the consequent, P(Y) and ϵ is the probability of the exception, P(¬Y|X).

Additionally, based on the parameters' definitions, I extended the model to represent participants' answers that are not only inference form endorsements. The other subjective probabilities that participants in some of the experiments are asked for are modeled by **Oaksford-Chater** in the following way:

Conjunctive Probability

$$\mathrm{P(X \wedge Y)} = a \cdot (1 - \epsilon)$$

Probability of Alternatives

$$\mathrm{P(Y|\neg X)} = \frac{b - a \cdot (1 - \epsilon)}{1 - a}$$

Probability of the Material Conditional

$$\mathrm{P(\neg X \vee Y)} = 1 - a \cdot \epsilon$$

$$\begin{array}{cccc} & \multicolumn{2}{c}{\text{P(minor premise)}} & \\ \text{MP:} & \text{P(X)} = a & \text{DA:} & \text{P}(\neg\text{X}) = 1 - a \\ \\ \text{AC:} & \text{P(Y)} = b & \text{MT:} & \text{P}(\neg\text{Y}) = 1 - b \\ & \multicolumn{2}{c}{\text{P(conclusion)}} & \\ \text{MP:} & \text{P(Y)} = b & \text{DA:} & \text{P}(\neg\text{Y}) = 1 - b \\ \\ \text{AC:} & \text{P(X)} = a & \text{MT:} & \text{P}(\neg\text{X}) = 1 - a \end{array}$$

Logistic Regression uses the following endorsement expressions, as presented in Chapter 2:

$$\textbf{MP:} \quad \text{P(Y|X)} = \frac{1}{1 + e^{-(b_0 + b_1)}} \quad \textbf{DA:} \quad \text{P}(\neg\text{Y}|\neg\text{X}) = \frac{1}{1 + e^{b_0}}$$

$$\textbf{AC:} \quad \text{P(X|Y)} = \frac{1}{1 + e^{-(b_0^* + b_1)}} \quad \textbf{MT:} \quad \text{P}(\neg\text{X}|\neg\text{Y}) = \frac{1}{1 + e^{b_0^*}}$$

e^{b_0} and $e^{b_0^*}$ are the intercepts of the logistic regression line when X is a predictor of Y and Y is a predictor of X respectively and e^{b_1} is the slope in both cases.

Finally, the **Dual-Source Model** is applied as explained in Chapter 2. Endorsement of reduced inference x with content C is represented with the following expression:

$$E_r(C, x) = \xi(C, x)$$

Then, endorsement of the full inference x with content C is represented as:

$$E_f(C, x) = \lambda\{\tau(x) + (1 - \tau(x)) \cdot \xi(C, x)\} + (1 - \lambda)\xi(C, x)$$

The parameters $\xi(C, x)$ are the knowledge-based parameters for content C and inference form x, $\tau(x)$ are the form-based parameters for inference form x and λ is a weight given to the form-based parameter.

The $\xi(C, x)$ parameters are obtained by using Oaksford et al. (2000)'s equations, as follows:

$$\textbf{MP:} \quad \xi(C, \text{MP}) = 1 - \epsilon \qquad \textbf{DA:} \quad \xi(C, \text{DA}) = \frac{1 - b - a \cdot \epsilon}{1 - a}$$

$$\textbf{AC:} \quad \xi(C, \text{AC}) = \frac{a(1 - \epsilon)}{b} \qquad \textbf{MT:} \quad \xi(C, \text{MT}) = \frac{1 - b - a \cdot \epsilon}{1 - b}$$

The parameters $\tau(x)$ and λ cannot be uniquely estimated, only their products are obtained. When fitting, the largest product $\lambda \cdot \tau(x)$ is set to be λ, and then by dividing with it the $\tau(x)$ values are obtained, and the largest $\tau(x)$ is set to 1.

Table 13 shows the number of free parameters each model needs to model one *task*. The DSM is not included since it is not able to simply model one singular *task*, but rather models *individuals* directly. For the rest of the models, the number of free parameters to model one *individual* depends on the type of experiment and to how many tasks participants had to provide answers. That number is reported for each model for each data set separately in the analysis (Chapter 6).

Table 13: Number of free parameters each model needs to fit one task

Model	# Free Parameters
MMT	4
MMT Deductive	3
MMT with Directionality	5
Suppositional Sequential	5
Suppositional Exclusive	5
Oaksford-Chater	3
Logistic Regression	3

From all of these implemented models, only **Oaksford-Chater** is able to represent probabilities other than the inference form endorsements. The formalizations based on MPTs (the **MMT** and **Suppositional** variants) are specifically made to model inference forms only. Due to the fact that logistic regression in general is used to predict a dependent variable using one or more independent variables, the **Logistic Regression** model is only applicable to the conditional inference task, i.e. it can be used to model only inference form endorsements, as confirmed via personal communication with Skovgaard-Olsen. The **Dual-Source Model** has a very attractive theoretical assumption on how people reason by disentangling form and content. However, a significant limitation is that it can only be fit to data that provides *independent* information for the knowledge-based and form-based components, as confirmed via personal communication with Singmann. Unfortunately, the only such data in the benchmark is from the experiments presented in Singmann et al. (2016), making it difficult to fairly compare it to other models.

4.2 Evaluation Methods

The benchmark performs two different kinds of evaluation. Firstly, we are interested in the models' performance on the data. In order to assess that, three different goodness-of-fit measures are calculated and compared. Furthermore, ϵ-MMT's parameters are statistically analyzed in order to determine whether the model can successfully account for the manipulated properties and effects in the experiments and additionally to provide insight into how humans reason with conditionals.

4.2.1 Goodness-of-Fit Measures

In order to be able to compare the performance of the different models, their goodness of fit is determined through various measures. In this thesis, the following are used: Root Mean Square Error (RMSE), the Adjusted Root Mean Square Error (RMSE_{adj}) and the Coefficient of Determination (R^2). Their definitions are provided below.

In this thesis, the fitting of the models is done per *individual*, in contrast to aggregate data fitting. Riesterer, Brand, and Ragni (2020) propose a model evaluation setting focusing on the individual, i.e. determines the model's ability to "cover" the response of individual reasoners. They argue that if models are capable of accurately reflecting cognitive processing, the individuals' behavior should be captured with the model's parameterization. That said, the goodness-of-fit measures in this benchmark are used for *coverage analysis*, in order to reveal whether the models manage to provide parameters that can capture the differences between individuals.

Root Mean Square Error (RMSE)

The Root Mean Square Error (RMSE) is a common measure of accuracy that considers the differences between observed true values and values predicted by a model. The optimization of the models in the benchmark is done by minimizing the RMSE for every individual. A lower RMSE value indicates better predictions and therefore better model performance.

The RMSE measure is implemented as shown in Eq. 3, where N is the number of data points, $true_i$ is the i-th true answer, $pred_i$ is the i-th predicted answer.

$$RMSE = \sqrt{\frac{\sum_{i=0}^{N}(true_i - pred_i)^2}{N}} \tag{3}$$

Adjusted Root Mean Square Error (RMSE$_{adj}$)

Given that all of the models in this benchmark have different number of free parameters, the Adjusted Root Mean Square Error (RMSE$_{adj}$) is also considered here. In contrast to RMSE, RMSE$_{adj}$ also takes the model's number of parameters into account, and punishes models with a larger number of parameters. A lower RMSE$_{adj}$ value indicates better predictions and therefore better model performance.

The RMSE$_{adj}$ measure is implemented as shown in Eq. 4, where N is the number of data points, p is the number of parameters, $true_i$ is the i-th true answer, $pred_i$ is the i-th predicted answer.

$$RMSE_{adj} = \sqrt{\frac{\sum_{i=0}^{N}(true_i - pred_i)^2}{N - p}} \tag{4}$$

In some of the experiments the RMSE$_{adj}$ measure provides either ∞ or nan values for a number of models. ∞ means that the model needs as many free parameters as data points that it needs to fit (division by 0). nan means that it needs more free parameters than provided data points (negative value under square root). Taking into consideration that in most cases these models' RMSE values are not competitive to begin with, the RMSE$_{adj}$ measure confirms that the substantial amount of necessary free parameters is absolutely not justified.

Coefficient of Determination (R^2)

Used by Singmann et al. (2016) to determine the goodness-of-fit of the Dual-Source Model, the Coefficient of Determination (R^2) is part of this benchmark as well. The data in the benchmark manipulates various reasoning properties and exploits effects, often leading to large variance in the participants' answers. It is valuable to find out whether the models can successfully account for that variance. The statistic R^2 provides the percentage of variance in the true values that is predictable by the model. Depending on the definition, the range interval of R^2 differs. Here, the interval is $[-\infty, 1]$. An R^2 value of 1 means that the model can account for 100% of the variance, an R^2 value of 0 means that the model predictions are as accurate as using the mean of the observed values as a predictor. A value below 0 means that the predictions are worse than just taking the mean of the data. A higher R^2 value indicates a better grasp of the data variance and therefore better model performance.

The R^2 measure is implemented as shown in Eq. 5, where N is the number of data points, $true_i$ is the i-th true answer, $pred_i$ is the i-th predicted answer, $\overline{true}$ is the average of the true answers.

$$R^2 = 1 - \frac{\sum_{i=0}^{N}(true_i - pred_i)^2}{\sum_{i=0}^{N}(true_i - \overline{true})^2} \tag{5}$$

Removal of Participants Unfortunately, in some cases participants provided the same exact answer to every single question, which obviously means that there was no variance to be accounted for in their data. In such a case, the denominator in Eq. 5 evaluates to 0, making the division impossible, and the final R^2 value, *nan*. In order to properly analyze the models' abilities to account for variance in the participants' answers, with *complete* data, one participant (Participant ID: U18U) has been completely removed from Singmann et al. (2014), Exp. 1 & 2 (in this thesis Exp. 7 & 8).

4.2.2 Parameter Analysis Methods

Two kinds of statistical analysis are performed: significance of *parameter value changes* and *correlation* between parameters and various perceived factor values. In both cases p-values are calculated, significant values are $p < .05$.

Parameter Value Changes Analysis

By analyzing the statistical significance of the differences in the parameter values for different tasks, the ability of the new model ϵ-MMT to account for certain effects is shown.

Based on how the different tasks in the experiment are distributed among participants, the parameter value analysis is done in two ways:

1. Within participants
2. Between participants

When analyzing the parameter values within participants, the Wilcoxon signed-rank test (Wilcoxon, 1945) is used, as implemented in `Python`'s `scipy.stats` library[2].

[2] `https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.wilcoxon.html`

Since the parameter values are paired, the reported difference is the mean of the differences between each pair.

When analyzing the parameter values between participants, the Mann-Whitney U test (Mann & Whitney, 1947) is used, as implemented in `Python`'s `scipy.stats` library[3]. The two sets of parameter values are independent and of different sizes. The reported difference is obtained by bootstrapping – 10 000 iterations of picking as many random samples from the larger of the two sets as the smaller set has, calculating their mean difference, and finally reporting the mean of all obtained differences.

Correlation Analysis

Taking into consideration the definition of ϵ-MMT's parameters provided in the following Chapter, the correlation between their values and various subjective probability estimates given by participants can be used to get a better insight into how humans interpret conditionals.

When calculating correlation, the Kendall rank correlation coefficient (Kendall, 1938) is used, as implemented in `Python`'s `scipy.stats` library[4].

4.3 Benchmark Program Functionality

The Benchmark Program is an interactive program that allows the user to fit models on different data and analyze the results, specifically ϵ-MMT's parameters behavior. The program provides the following options:

Data Information: This part of the program can give the user information about the data: the names of the datasets, reasoning properties, number of free parameters each model needs to fit an individual in a specific experiment, number of participants and the experiment's varying factors.

Fit: The user can choose the model(s) they would like to fit on all of the data or one specific dataset of their own choosing. When a model is done fitting a dataset, the corresponding mean RMSE, RMSE_{adj} and R^2 are then printed on the screen.

Plot: The program can produce different plots. The user can choose between plotting RMSE/RMSE_{adj}/R^2 values overall, or per dataset. Additionally, it can provide plots of the parameter values for different varying factors in experiments.

Goodness of Fit: By running this part of the program, a .CSV file is created with every model's mean RMSE, RMSE_{adj} and R^2 for every data set that it can fit.

[3] https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.mannwhitneyu.html
[4] https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kendalltau.html

Analysis: The user can choose the experiment they'd like to focus on, whether they would like to perform parameter value analysis or correlation analysis (if applicable). For the former, the Wilcoxon signed-rank test is done for within participant analysis, and, the Mann-Whitney U test for between participant analysis. For the latter, the Kendall rank correlation coefficient is calculated.

5 ϵ-MMT

ϵ-semantics is described by Pearl (1991) as a 'formal framework for belief revision', where belief statements are interpreted as statements of high probability and belief revision shapes current beliefs on newly available evidence.

ϵ-semantics differentiates between two types of sentences:

1. Sentences describing truths and general tendencies (e.g., 'Birds fly.')

2. Sentences describing observations specific to a given situation (e.g., 'All blocks on this table are green.')

This distinction is highly relevant because it is reflected in the natural language when using the word 'If'. For example, saying 'If it is a bird then it flies' is legitimate, whereas 'If this block were on this table, it would be green' is not.

The following is an important definition provided by Pearl (1991, pg. 5) that lays basis for the probabilistic aspect of the new model:

> Let L be the language of propositional formulas, and let a *truth-valuation* for L be a function t, that maps the sentences in L to the set $\{1, 0\}$, (1 for 'true', 0 for 'false'). To define a probability assignment over the sentences in L, we regard each truth valuation t as a world w and define $\mathrm{P}(w)$ such that $\sum_w \mathrm{P}(w) = 1$. This assigns a probability measure to each sentence l of L.

The development of ϵ-MMT is inspired by the shift to a new paradigm of reasoning – making inferences depends on our *subjective degrees of belief* and there is a certain degree of *uncertainty* in the premises, rather than assuming that inferences are simply either true or false. Simultaneously, uncertainty also varies from person to person, based on their individual experiences and background knowledge that could lead to potential biases. Therefore ϵ-MMT will be used to model people *individually*, and with that offer a better insight in certain patterns or differences between individuals.

Schroyens et al.'s (2001) revision of the MMT is not a theory of deductive reasoning because it allows for the X $\neg$Y model to be retrieved from long-term memory. Since the interest is to model realistic data where individuals integrate their background knowledge when reasoning, similarly to Schroyens et al. (2001)'s revision, the ϵ-MMT model takes the mental model representation of *all* the conditional's propositions' truth state combinations, or as they will be called from now on – *possible worlds*. Finally, ϵ-MMT applies Pearl's (1991) definition to them, i.e. defines a probability distribution over all worlds.

The concept of possible worlds has already been explored in philosophy by Lewis (1973) and Stalnaker (1968) for the counterfactual conditional ("If X had happened, then Y would have happened"). Based on their account, the conditional does not depend on the way the world is, but rather on possible ways the world might be, or, in philosophical terms, it is not *extensional*, but *intensional* (Oaksford & Chater, 2007). Stalnaker's (1968) idea of possible worlds and closeness between them was discussed in Section 1.3 when introducing the history of characterization of conditionals. ϵ-MMT's approach does not explicitly consider closeness between worlds. However, it does use the notion of representing a conditional with possible worlds and parameterizes it. Thus, a variety of subjective beliefs considering the conditional's world can be expressed in order to gain more insight into how humans interpret conditionals and ultimately reason.

Given a premise containing two propositions, X and Y, all possible worlds described by the premise along with the corresponding probability values are shown in Table 14.

Table 14: The possible worlds described by a premise containing two prepositions, X and Y, the probability distribution P and probability values p_i, $i \in (1, 2, 3, 4)$.

X	**Y**	***P***
0	0	p_1
0	1	p_2
1	0	p_3
1	1	p_4

For example, given a conditional "If it is a bird, then it flies", the probability value assigned to the world where it is a bird, and it is not flying (X = 1, Y = 0) is p_3. This specific world is generally not considered in theories that adhere to the material implication interpretation of conditionals, like the original version of the Mental Model Theory (Johnson-Laird & Byrne, 1991) itself.

Previous accounts in the Bayesian paradigm (Chan & Chua, 1994; Stevenson &

Over, 2001; Liu, Lo, & Wu, 1996; Oaksford et al., 2000) assume that an individual's endorsement of an inference form can be expressed as a conditional probability of the conclusion given the minor premise. ϵ-MMT follows the same approach.

$$\mathrm{P}(\beta|\alpha) = \frac{\mathrm{P}(\alpha \wedge \beta)}{\mathrm{P}(\alpha)} \tag{6}$$

Equation 6 shows the definition of conditional probability, from which Bayes' theorem may be derived. Following Eq. 6, the four equations shown below are obtained. They describe the endorsement of the four inference forms through the probability distribution P of the conditional's worlds (Table 14):

$$\textbf{MP:}\quad \mathrm{P(Y|X)} = \frac{p_4}{p_3 + p_4} \qquad \textbf{DA:}\quad \mathrm{P(\neg Y|\neg X)} = \frac{p_1}{p_1 + p_2}$$

$$\textbf{AC:}\quad \mathrm{P(X|Y)} = \frac{p_4}{p_4 + p_2} \qquad \textbf{MT:}\quad \mathrm{P(\neg X|\neg Y)} = \frac{p_1}{p_1 + p_3}$$

Hadjichristidis et al. (2001) conducted a study whose results suggest that when reasoning, individuals construct an imaginary world where the antecedent of the conditional holds and then consider the likelihood of the consequent also holding in the same world. Even though they are analyzing the interpretation of P("If X then Y") through conditional probability, their conclusions are still relevant here. They point out the need for the world X ¬Y to be (at least implicitly) considered due to the fact that it is necessary to formulate the conditional probability. Here, that is applicable specifically to the MP and MT inference forms.

Besides inference form endorsements, participants of some experiments have been asked for different subjective probabilities which ϵ-MMT is able to model with the following equations:

Conjunctive Probability

$$\mathrm{P(X{\wedge}Y)} = p_4$$

Probability of Alternatives

$$\mathrm{P(Y|\neg X)} = \frac{p_2}{p_1 + p_2}$$

Probability of the Material Conditional

$$\mathrm{P(\neg X{\vee}Y)} = p_1 + p_2 + p_4$$

P(minor premise)

$$\text{MP:}\quad \text{P(X)} = p_3 + p_4 \quad \text{DA:}\quad \text{P}(\neg\text{X}) = p_1 + p_2$$

$$\text{AC:}\quad \text{P(Y)} = p_4 + p_2 \quad \text{MT:}\quad \text{P}(\neg\text{Y}) = p_1 + p_3$$

P(conclusion)

$$\text{MP:}\quad \text{P(Y)} = p_2 + p_4 \quad \text{DA:}\quad \text{P}(\neg\text{Y}) = p_1 + p_3$$

$$\text{AC:}\quad \text{P(X)} = p_4 + p_3 \quad \text{MT:}\quad \text{P}(\neg\text{X}) = p_1 + p_2$$

With these equations participants can be modeled individually, by fitting the parameters p_1, p_2, p_3 and p_4 to their answers, i.e. obtaining their subjective probability distribution that describes the conditional's world, according to the individual's beliefs.

The parameters are bound by their sum, $\sum_i p_i = 1$, meaning that the number of free parameters for modeling one task with this model is three. The total number of parameters to model one *individual* therefore depends on the number of tasks that they have to complete.

5.1 Related Work

Pearl (1988) explored the use of probability theory as a semantic basis for conditional reasoning using ϵ-semantics. He is following the interpretation of the probability of a conditional rule, P("If X then Y"), to be equal to the conditional probability P(C|A), when being infinitesimally distant from 0 or from 1. Similarly, he applies the probability distribution definition stated above to possible worlds. However, his focus is on whether a conclusion is *plausible* enough to be *accepted*. For example, given a conditional "If X then Y", a conclusion would be accepted if $\text{P(Y|X)} \geq 1 - \epsilon$, where ϵ is an arbitrarily small value, short of being zero. As Pearl (1988) points out, extreme probabilities, i.e. probabilities that are infinitesimally close to 0 and 1 are rare in the real world. The proposed model ϵ-MMT differs from Pearl's (1988) approach as it is not concerned with extreme probabilities nor with acceptance of conclusions, but their likelihood.

6 Analysis

This Chapter presents the benchmark's analysis results.

First, we will evaluate the models' fit on all data. A point of interest is to also look at the theories that had multiple variants implemented and make a comparison between them.

Afterwards, the focus is on ϵ-MMT. First it is established how ϵ-MMT competes with other models in the benchmark for each experiment separately. Then, an analysis is performed on how ϵ-MMT's parameters are influenced by manipulations of different reasoning properties and effects, along with the psychological interpretation of the value changes.

6.1 Benchmark Model Performance

In order to fairly compare the general models' performances across all datasets, only a reduced set of seven experiments is considered. The eliminated one is Experiment 7 (the probabilized conditional task presented in Singmann et al. (2014)) because it asks for probabilities that only ϵ-MMT and the Oaksford-Chater Probabilistic Model can represent with their parameters. Additionally, the Dual-Source Model is not included in this comparison, due to the fact that it is only applied to experiments 1 – 4, whose fitting results are discussed in their respective analysis sections.

The results are presented in Table 15. The models are ordered by their RMSE values, from best to worst. Overall, the best performing models are ϵ-MMT and Oaksford-Chater with an equal RMSE of 0.054 and RMSE_{adj} of 0.108. Oaksford-Chater's R^2 is only slightly better, 0.925, compared to ϵ-MMT's 0.924. However, given that both values are considerably high, the minor difference is negligible – both models account for more than an impressive 92% of data variance, on average. Both of the models need only three free parameters to model one task. The Logistic Regression Model is a close competitor to the two models, taking into consideration that it also needs only three parameters to model one task, and its RMSE value is only slightly worse, 0.073.

Out of the three MMT variants, the MMT with Directionality has the best performance with an RMSE of 0.091 and R^2 of 0.834. The deductive variant has the worst performance not only within these three models, but overall, with an RMSE value of 0.151 which is almost three times higher than the one of ϵ-MMT and Oaksford-Chater, and quite a low R^2 of 0.600.

When looking at the variants of the Suppositional Theory models, the Sequential variant performs a lot better than the Exclusive one, with an RMSE of 0.082 and R^2 of 0.819 compared to 0.124 and 0.694. However, the Sequential model needs 5 parameters to model one task, which is too many compared to the leading models ϵ-MMT and Oaksford-Chater that need only 3.

Table 15: Fitting results for each model on all datasets (fair comparison). 'Free' - Number of free parameters needed to model one *task*. The missing RMSE_{adj} values mean that the model gets either *nan* or ∞ values for some datasets.

Model	**RMSE**	**RMSE_{adj}**	**R^2**	**Free**
ϵ-MMT	0.054	0.108	0.924	3
Oaksford-Chater	0.054	0.108	0.925	3
Logistic Regression	0.073	0.146	0.800	3
Suppositional Sequential	0.082	–	0.819	5
MMT with Directionality	0.091	–	0.834	5
MMT	0.116	–	0.754	4
Suppositional Exclusive	0.124	–	0.694	5
MMT Deductive	0.151	0.301	0.600	3

6.2 Experiments Analysis

The main goal in the following is to analyze ϵ-MMT's abilities to account for the various properties and effects that are manipulated in the chosen experiments and present how they affect its parameters while providing an explanation from a psychological point of view.

Where applicable, if other models have been applied to this data and their results have been published elsewhere, they will also be reported here.

6.2.1 Experiments 1–4

These are the experiments presented in Singmann et al. (2016) and Section 3.2.1. Their goal was to show the importance of background knowledge as well as the conditional rule's form when reasoning.

Opposed to the fitting approach used in this benchmark, Singmann et al. (2016) used weighted least-squares in order to estimate the Dual-Source Model's parameters, where the weight is 1 - $|s$ - 1$|$, s being the sum of the inference form endorsement and its converse in a [0, 1] interval. They combined the two responses into one estimate: $\frac{E(inf)+(1-E(conv(inf)))}{2}$. Given the parameter optimization approach in my benchmark, combining the responses is not applicable. Ultimately, we are interested in the *exact* individual's answers and the models' capabilities to capture them, therefore this benchmark takes only the original inference form endorsements into consideration.

Singmann et al. (2016) also provide a goodness-of-fit meta-analysis where they compare the performance of the Dual-Source Model to other probabilistic models, including Oaksford et al.'s (2000) model, for which they adopted a different approach than the one used in this thesis. Moreover, they excluded some conditions from their experiments due to the fact that it has not been entirely clear how to model certain manipulations with the competitor models. The modifications for their meta-analysis diverge from the approach taken in this benchmark, so their results will not be reported or discussed any further as they are not relevant.

Experiment 1

Models Fits All of the models are able to fit this experiment's data. The goodness of fit measures values for each model are presented in Table 16.

In Table 16, the number of free parameters to model one *individual* is reported. ϵ-MMT needs three free parameters to model one task (4 data points per task). Given 12 tasks, the total number of free parameters to model an individual is 3 $\times$ 12 = 36. The DSM on the other hand fits 12 $\xi(C, x)$ parameters and 2 $\times$ 4 $\lambda\tau$ parameters $\rightarrow$ 20 parameters in total per individual.

The best performing models are ϵ-MMT and the Oaksford-Chater Probabilistic Model with an RMSE of 0.035, an RMSE$_{adj}$ of 0.071, and a significantly high R^2 value of 0.967 (ϵ-MMT) and 0.966 (Oaksford-Chater). Both models need the same amount of free parameters, which is three per task, or, 36 per individual.

Using the different fitting approach, Singmann et al. (2016) reported a R^2 value of 0.900, compared to the value of 0.795 obtained here, but no other measure values are provided. The DSM needs only 20 parameters, which is the least out of all models. Howbeit, its performance is not on a level with the two best-performing models, as shown by a RMSE$_{adj}$ value of 0.127, compared to 0.071 for both ϵ-MMT and Oaksford-Chater.

Table 16: Fitting results for each model on experiment 1. 'Free' - Number of free parameters needed to model one individual.

Model	RMSE	$RMSE_{adj}$	R^2	Free
ϵ-MMT	0.035	0.071	0.967	36
Oaksford-Chater	0.035	0.071	0.966	36
Logistic Regression	0.083	0.166	0.769	36
Dual-Source Model	0.097	0.127	0.795	20
MMT	0.084	∞	0.853	48
MMT Deductive	0.146	0.291	0.586	36
MMT with Directionality	0.072	nan	0.889	60
Suppositional Sequential	0.107	nan	0.770	60
Suppositional Exclusive	0.138	nan	0.625	60

Conditional Presentation Form In this experiment the conditional presentation form varies between reduced inference, conditional and biconditional. It is expected that when individuals are presented with a conditional or a biconditional their confidence in the rule will increase, compared to a reduced inference. That leads to an assumption that the value of the p_4 parameter will be higher in the conditional and biconditional cases. In order to determine whether the differences between parameter values are significant, a Wilcoxon signed-rank test was performed. Table 17 presents the results of p_4's value changes analysis. When comparing reduced inference to conditional, a significant increase of p_4 is present in the 'Many/Few' and 'Many/Many' tasks (p < .001 for both). On the other hand, when comparing reduced inference to biconditional, there is also a significant increase of p_4 in the 'Few/Many' task (p < .001) along with the 'Many/Few' (p = .001) and 'Many/Many' (p = .004) tasks. However, no significant change is encountered in the p_4 values when comparing the conditional and biconditional cases.

Singmann et al. (2016) report a significant (p < .001) effect on DSM's form-parameters $\tau(x)$ when contrasting conditional and biconditional inferences.

Varying Amount of Disablers and Alternatives When presented with a reduced inference individuals rely more on their background knowledge. Therefore, that is the only conditional presentation form taken into consideration here. Given a conditional "If X then Y", the p_2 parameter describes the probability of the world ¬X Y, which is interpreted as the impact of *alternatives*. On the other hand, p_3 describes the probability of the world X ¬Y, which is interpreted as the impact of *disablers*. Taking that into account, an increase in the amount of alternatives would

Table 17: Experiment 1, Conditional Presentation Form: Mean percentages of the individuals' values for p_4. Means of the differences between individuals' values for p_4. ('D' - Disablers, 'A' - Alternatives, 'F' - Few, 'M' - Many, 'Form' - Conditional Presentation Form, 'Red' - Reduced Inference, 'Cnd' - Conditional, 'Bcnd' - Biconditional, 'p_i' - Parameter)

D/A	Form	p_i	Mean
F/F	Red	p_4	51.39
	Cnd		60.01
	Bcnd		51.53
F/M	Red	p_4	30.28
	Cnd		42.61
	Bcnd		52.30
M/F	Red	p_4	20.22
	Cnd		52.70
	Bcnd		42.82
M/M	Red	p_4	29.69
	Cnd		45.72
	Bcnd		47.38

(a) Mean values of p_4.

D/A	Form 1	Form 2	p_i	Mean Δ	p-value
F/F	Red	Cnd	p_4	-8.61	.075
	Red	Bcnd		-0.14	.891
	Cnd	Bcnd		8.47	.060
F/M	Red	Cnd	p_4	-12.33	.117
	Red	Bcnd		-22.03	< **.001**
	Cnd	Bcnd		-9.69	.327
M/F	Red	Cnd	p_4	-32.47	< **.001**
	Red	Bcnd		-22.60	**.001**
	Cnd	Bcnd		9.87	.117
M/M	Red	Cnd	p_4	-16.03	< **.001**
	Red	Bcnd		-17.69	**.004**
	Cnd	Bcnd		-1.66	.570

(b) Means of the differences between values of p_4. Significant p-values are marked in bold.

lead to an increase of p_2's value and similarly an increase in the amount of disablers would lead to an increase of p_3's value. Therefore, it is expected that in the case of 'Many' alternatives/disablers the values of p_2/p_3 would be larger in comparison to 'Few'.

Firstly, it should be noted that the task with 'Many' disablers and 'Few' alternatives, whose content is "If a girl has sexual intercourse then she will be pregnant" can be thought of as slightly controversial. There are some discrepancies and unexpected results in the analysis when looking at this specific task only.

In order to determine whether the differences between parameter values are significant, a Wilcoxon signed-rank test was performed. The results can be found in Table 18. It is immediately noticeable that all of the expected parameter value changes are present and statistically significant.

Previous Work Todorovikj, Friemann, and Ragni (2019) took the first step towards establishing this new cognitive model for reasoning with conditionals, ϵ-MMT. The goal was to show that this model can account for the effect of various amounts of disablers and alternatives, and for the conditional presentation form effect. That

Table 18: Experiment 1, Varying Amounts of Disablers and Alternatives: Mean percentages of the individuals' values for p_2 and p_3 in the reduced inference case. Means of the differences between individuals' values for p_2 and p_3. ('D' - Disablers, 'A' - Alternatives, 'F' - Few, 'M' - Many, p_i - Parameter)

D/A	p_i	**Mean**
F/F	p_2	5.10
	p_3	3.42
F/M	p_2	15.41
	p_3	4.11
M/F	p_2	0.50
	p_3	46.70
M/M	p_2	19.24
	p_3	18.15

(a) Mean values of p_2 and p_3.

D/A 1	**D/A 2**	p_i	**Mean Δ**	**p-value**
F/F	F/M	p_2	-10.31	**.001**
		p_3	-0.69	**.012**
F/F	M/F	p_2	4.60	< **.001**
		p_3	-43.28	< **.001**
F/F	M/M	p_2	-14.14	< **.001**
		p_3	-14.74	< **.001**
F/M	M/F	p_2	14.91	< **.001**
		p_3	-42.59	< **.001**
F/M	M/M	p_2	-3.83	.100
		p_3	-14.05	< **.001**
M/F	M/M	p_2	-18.74	< **.001**
		p_3	28.55	< **.001**

(b) Means of the differences between values of p_2 and p_3. Significant p-values are marked in bold.

was done by modeling this exact experiment. However, the fitting of the model was done *per task*. Here, the focus is on *individual fitting*, i.e. fitting on all the tasks that an individual has answered at once. Therefore, this experiment is taken into consideration once again. Either way, the conclusions made based on the analysis results do not differ.

Experiment 2

Models Fits All of the models are able to fit this experiment's data. The goodness of fit measures values for each model are presented in Table 19.

In Table 19, the number of free parameters to model one *individual* is reported. ϵ-MMT needs three free parameters to model one task (4 data points per task). Given 12 tasks, the total number of free parameters to model an individual is $3 \times 12 = 36$. The DSM on the other hand fits 18 $\xi(C, x)$ parameters and 2×4 $\lambda\tau$ parameters $\rightarrow$ 26 parameters in total per individual.

The best performing models are ϵ-MMT and the Oaksford-Chater Probabilistic Model with an RMSE of 0.050 and 0.049 respectively, an RMSE_{adj} of 0.101 and 0.098 respectively, and high R^2 values of 0.928 and 0.930 respectively. Both models need the same amount of free parameters, which is three per task, or, 36 per individual.

Using a different fitting approach, Singmann et al. (2016) reported a R^2 value of 0.820, compared to the value of 0.726 obtained here, but, again, no other measure values are provided. The DSM needs 26 parameters, which is the least out of all models. Even so, its performance is not as good as the two best-performing models, based on its $RMSE_{adj}$ value of 0.155 (compared to 0.101 and 0.098 for ϵ-MMT and Oaksford-Chater).

Table 19: Fitting results for each model on experiment 2. 'Free' - Number of free parameters needed to model one individual.

Model	RMSE	$RMSE_{adj}$	R^2	Free
ϵ-MMT	0.050	0.101	0.928	36
Oaksford-Chater	0.049	0.098	0.930	36
Logistic Regression	0.057	0.113	0.908	36
Dual-Source Model	0.105	0.155	0.726	26
MMT	0.107	∞	0.729	48
MMT Deductive	0.125	0.251	0.638	36
MMT with Directionality	0.091	nan	0.804	60
Suppositional Sequential	0.079	nan	0.839	60
Suppositional Exclusive	0.109	nan	0.724	60

Conditional Presentation Form In this experiment the conditional presentation form varies between reduced inference and conditional. It is expected that when individuals are presented with a reduced inference they will rely more on their background knowledge and think of scenarios that are alternative to the one described by a rule, compared to when they are presented with a conditional. Therefore, the hypothesis is that there will be an increase in the value of p_4 in the conditional case.

In order to determine whether the differences between parameter values are significant, a Wilcoxon signed-rank test was performed. The results in Table 20 show that the increase of the p_4 parameter, in both speaker expertise cases, is significant ($p < .001$), which confirms the hypothesis.

Speaker Expertise Stevenson and Over (2001) have shown that individuals tend to perceive a conclusion's likelihood as greater when it has been stated that a major or a minor premise has been uttered by an expert, in contrast to a non-expert. Singmann et al. (2016) only do the speaker expertise manipulation on the major premise, i.e. the conditional rule. However, they report that they could not find the effect on all

Table 20: Experiment 2, Conditional Presentation Form: Mean percentages of the individuals' values for p_4. Means of the differences between individuals' values for p_4. ('Speak' - Speaker Expertise, 'Non' - Non-expert, 'Exp' - Expert, 'Form' - Conditional Presentation Form, 'Red' - Reduced Inference, 'Cnd' - Conditional, 'p_i' - Parameter)

Speak	**Form**	p_i	**Mean**
Non	Red	p_4	31.49
	Cnd		41.18
Exp	Red	p_4	31.20
	Cnd		42.11

(a) Mean values of p_4.

Speak	**Form 1**	**Form 2**	p_i	**Mean Δ**	**p-value**
Non	Red	Cnd	p_4	-9.69	< **.001**
Exp	Red	Cnd	p_4	-10.91	< **.001**

(b) Means of the differences between values of p_4. Significant p-values are marked in bold.

inference forms. In general, here, it is expected that the belief in the X Y world will become stronger when the rule has been uttered by an expert, meaning an increase in the p_4 parameter in the 'Expert' case.

Table 21: Experiment 2, Speaker Expertise: Mean percentages of the individuals' values for p_3 and p_4 in the conditional case. Means of the differences between individuals' values for p_3 and p_4. ('Speak' - Speaker Expertise, 'Non' - Non-expert, 'Exp' - Expert, p_i - Parameter)

Speak	p_i	**Mean**
Non	p_3	11.40
	p_4	41.18
Exp	p_3	9.04
	p_4	42.11

(a) Mean values of p_3 and p_4.

Speak 1	**Speak 2**	p_i	**Mean Δ**	**p-value**
Non	Exp	p_3	2.36	**.004**
		p_4	-0.93	.628

(b) Means of the differences between values of p_3 and p_4. Significant p-values are marked in bold.

In order to determine whether the differences between parameter values are significant, a Wilcoxon signed-rank test was performed. There is a very small increase in the p_4 parameter by only 0.04, however it is not statistically significant (p = .628). On the other hand, as the results in Table 21 show, a significant decrease in the p_3 parameter is present (p = .004).

Singmann et al. (2016) report a significant (p = .010) effect on DSM's weight parameter λ, more specifically, its mean value is higher in the 'Expert' case, meaning that more weight is given to the form-based component, compared to the knowledge-based one.

Experiments 3 & 4

Models Fits All of the models are able to fit these experiments' data. The goodness of fit measures values for each model are presented in Table 22 (Exp. 3) and 23 (Exp. 4).

In Table 22 and 23, the number of free parameters to model one *individual* is reported. ϵ-MMT needs three free parameters to model one task (4 data points per task). Given eight tasks, the total number of free parameters to model an individual is $3 \times 8 = 24$. The DSM on the other hand fits 12 $\xi(C, x)$ parameters and 4 $\lambda\tau$ parameters $\rightarrow$ 16 parameters in total per individual.

The best performing models are ϵ-MMT and the Oaksford-Chater Probabilistic Model with an RMSE of 0.052 (Exp. 3) and 0.051 (Exp. 4), an RMSE$_{adj}$ of 0.103 (Exp. 3) and 0.102 (Exp. 4), and high R^2 values of 0.943 (Exp. 3) and 0.947 (Exp. 4). Both models need the same amount of free parameters, which is three per task, or, 24 per individual.

Since the pattern of the results between the two experiments was not distinguishable, Singmann et al. (2016) did not provide separate analysis, but combined the two in one. Using a different fitting approach, they reported a *combined* R^2 value of 0.890, compared to a combined value of 0.663 obtained in this benchmark – 0.670 for exp. 3 and 0.657 for exp. 4. Again, no other measure values are provided. To make the comparison complete, the combined R^2 for each ϵ-MMT and Oaksford-Chater is 0.945. The DSM needs 16 parameters, which is the least out of all models. However, its performance is not as good as the two best-performing models, based on its RMSE$_{adj}$ value of 0.191 for Exp. 3, compared to 0.103 for ϵ-MMT and Oaksford-Chater), and 0.200 for Exp. 4, compared to 0.102 for ϵ-MMT and Oaksford-Chater.

Table 22: Fitting results for each model on experiment 3. 'Free' - Number of free parameters needed to model one individual.

Model	**RMSE**	**RMSE$_{adj}$**	**R^2**	**Free**
ϵ-MMT	0.052	0.103	0.943	24
Oaksford-Chater	0.052	0.103	0.943	24
Logistic Regression	0.070	0.141	0.816	24
Dual-Source Model	0.135	0.191	0.670	16
MMT	0.099	∞	0.825	32
MMT Deductive	0.164	0.328	0.566	24
MMT with Directionality	0.084	nan	0.871	40
Suppositional Sequential	0.119	nan	0.755	40
Suppositional Exclusive	0.152	nan	0.616	40

Table 23: Fitting results for each model on experiment 4. 'Free' - Number of free parameters needed to model one individual.

Model	RMSE	RMSE$_{adj}$	R^2	Free
ϵ-MMT	0.051	0.102	0.947	24
Oaksford-Chater	0.051	0.102	0.947	24
Logistic Regression	0.068	0.136	0.896	24
Dual-Source Model	0.142	0.200	0.657	16
MMT	0.109	∞	0.799	32
MMT Deductive	0.161	0.321	0.579	24
MMT with Directionality	0.094	nan	0.845	40
Suppositional Sequential	0.148	nan	0.638	40
Suppositional Exclusive	0.147	nan	0.637	40

Conditional Presentation Form The participants of these experiments were divided in three groups: 'Baseline', 'Disablers' and 'Alternatives', where the last two were provided with additional information in the form of disablers and alternatives, respectively, in order to provoke the Suppression Effect. There is no expected impact of the conditional presentation form in those two groups, therefore in the following only the 'Baseline' group will be taken into consideration.

In these experiments the conditional presentation form varies between reduced inference and conditional. The assumption is that when individuals are presented with a reduced inference they use their background knowledge more and will find alternative scenarios to the one described by a rule easier than when they are presented with the conditional. That leads to a hypothesis stating that the p_4 value should increase in the conditional case.

In order to determine whether the differences between parameter values are significant, a Wilcoxon signed-rank test was performed. The results for both experiments can be found in Table 24 and it can be easily seen that they confirm the findings in the analysis of Experiment 1. First, we have the 'Few/Few' task, which again lacks significant changes. Furthermore, in both experiments the 'Many/Few' and 'Many/Many' tasks have a significant increase in the conditional case (exp. 3: $p = .002$ and $p = .034$; exp.4: $p = .001$, $p = .016$). Experiment 3 also repeats the lack of significance in the parameter change for the 'Few/Many' task as it was the case in Experiment 1, however in Experiment 4 a significant increase ($p = .028$) can be observed.

Table 24: Experiment 3 & 4, Conditional Presentation Form: Mean percentages of the individuals' values for p_4 in the 'Baseline' group. Means of the differences between individuals' values for p_4. ('D' - Disablers, 'A' - Alternatives, 'F' - Few, 'M' - Many, 'Form' - Conditional Presentation Form, 'Red' - Reduced Inference, 'Cnd' - Conditional, 'p_i' - Parameter)

D/A	**Form**	p_i	**Mean**
F/F	Red	p_4	55.50
	Cnd		54.23
F/M	Red	p_4	51.87
	Cnd		48.31
M/F	Red	p_4	18.86
	Cnd		40.03
M/M	Red	p_4	27.54
	Cnd		39.06

(a) Exp. 3: Mean values of p_4.

D/A	**Form 1**	**Form 2**	p_i	**Mean Δ**	**p-value**
F/F	Red	Cnd	p_4	1.27	.770
F/M	Red	Cnd	p_4	3.56	.469
M/F	Red	Cnd	p_4	-21.17	**.002**
M/M	Red	Cnd	p_4	-11.53	**.034**

(b) Exp. 3: Means of the differences between values of p_4. Significant p-values are marked in bold.

D/A	**Form**	p_i	**Mean**
F/F	Red	p_4	57.85
	Cnd		53.81
F/M	Red	p_4	33.75
	Cnd		46.74
M/F	Red	p_4	18.45
	Cnd		49.16
M/M	Red	p_4	37.72
	Cnd		47.88

(c) Exp. 4: Mean values of p_4.

D/A	**Form 1**	**Form 2**	p_i	**Mean Δ**	**p-value**
F/F	Red	Cnd	p_4	4.04	.482
F/M	Red	Cnd	p_4	-12.99	**.028**
M/F	Red	Cnd	p_4	-30.71	**.001**
M/M	Red	Cnd	p_4	-10.16	**.016**

(d) Exp. 4: Means of the differences between values of p_4. Significant p-values are marked in bold.

Varying Amount of Disablers and Alternatives In this part of the analysis all three groups of participants ('Baseline', 'Disablers' and 'Alternatives') will be taken into consideration. The idea behind including all of them is to examine whether providing additional information overpowers the influence of the knowledge that individuals already possess, i.e. whether the expected parameter changes take place in all three groups.

As in Experiment 1, the only conditional presentation form taken into consideration is the reduced inference. Given a conditional "If X then Y" the probability of the world ¬X Y (*alternatives*) is described by p_2. The probability of the world X ¬Y (*disablers*) is described by p_3. It is expected that when the amount of alternatives increases, so will the value of p_2, and similarly, when the amount of disablers increases, so will the value of p_3. Therefore the values of p_2/p_3 in the case of 'Many' alternatives/disablers should be larger in comparison to 'Few'.

A reminder that the task with 'Many' disablers and 'Few' alternatives, whose content is "If a girl has sexual intercourse then she will be pregnant", can be thought of as slightly controversial, which leads to some discrepancies and unexpected results in the analysis related to this task.

In order to determine whether the differences between parameter values are significant, a Wilcoxon signed-rank test was performed. For the 'Baseline' groups of both experiments, the results can be found in Table 25. With the exception of the unexpected lack of an increase in p_2's value when comparing the task 'Few/Few' with 'Few/Many', the rest of the results in both experiments coincide with what was concluded in Experiment 1 – the expected parameter changes are statistically significant. The results of the 'Disablers' groups of both experiments are presented in Table 26. All of the expected parameter changes can be observed as statistically significant. Finally, the results of the 'Alternatives' groups of both experiments are presented in Table 27. In Experiment 3, all of the expected parameter changes are present and statistically significant. In Experiment 4, on the other hand, a significant increase of p_2 when comparing 'Few/Few' and 'Few/Many' is once again lacking. A special point of interest is the comparison between 'Few/Few' and 'Many/Many' where a significant increase in both parameters is expected, however one is lacking for the disablers parameter, p_2. The rest of the expected parameter changes, though, still do take place and are significant.

Table 25: Experiment 3 & 4, Varying Amounts of Disablers and Alternatives: Mean percentages of the individuals' values for p_2 and p_3 in the reduced inference case in the 'Baseline' group. Means of the differences between individuals' values for p_2 and p_3. ('D' - Disablers, 'A' - Alternatives, 'F' - Few, 'M' - Many, p_i - Parameter)

D/A	$\boldsymbol{p_i}$	**Exp. 3**	**Exp. 4**
		Mean	
F/F	p_2	5.24	4.93
	p_3	4.09	3.40
F/M	p_2	11.81	9.57
	p_3	3.16	4.83
M/F	p_2	2.31	2.20
	p_3	46.56	32.40
M/M	p_2	24.79	13.26
	p_3	19.18	37.72

(a) Mean values of p_2 and p_3.

D/A 1	**D/A 2**	$\boldsymbol{p_i}$	**Exp. 3**		**Exp. 4**	
			Mean Δ	**p-value**	**Mean Δ**	**p-value**
F/F	F/M	p_2	-6.57	.066	-4.64	.206
		p_3	0.93	.485	-1.43	.991
F/F	M/F	p_2	2.93	**.025**	2.73	**.001**
		p_3	-42.47	< **.001**	-29.00	< **.001**
F/F	M/M	p_2	-19.55	< **.001**	-19.18	< **.001**
		p_3	-15.10	< **.001**	-9.86	< **.001**
F/M	M/F	p_2	9.50	**.005**	7.37	**.005**
		p_3	-43.40	< **.001**	-27.57	< **.001**
F/M	M/M	p_2	-12.99	**.013**	-14.54	< **.001**
		p_3	-16.02	< **.001**	-8.43	**.002**
M/F	M/M	p_2	-22.49	< **.001**	-21.91	< **.001**
		p_3	27.38	< **.001**	19.13	< **.001**

(b) Means of the differences between values of p_2 and p_3. Significant p-values are marked in bold.

Table 26: Experiment 3 & 4, Varying Amounts of Disablers and Alternatives: Mean percentages of the individuals' values for p_2 and p_3 in the reduced inference case in the 'Disablers' group. Means of the differences between individuals' values for p_2 and p_3. ('D' - Disablers, 'A' - Alternatives, 'F' - Few, 'M' - Many, p_i - Parameter)

D/A	p_i	**Exp. 3**	**Exp. 4**
		Mean	
F/F	p_2	6.17	3.70
	p_3	8.58	8.48
F/M	p_2	13.67	8.48
	p_3	5.67	7.79
M/F	p_2	2.96	3.11
	p_3	54.42	41.0
M/M	p_2	19.78	21.69
	p_3	16.09	17.94

(a) Mean values of p_2 and p_3.

D/A 1	**D/A 2**	p_i	**Exp. 3**		**Exp. 4**	
			Mean Δ	**p-value**	**Mean Δ**	**p-value**
F/F	F/M	p_2	-7.51	**.009**	-4.78	**.006**
		p_3	2.92	.136	0.69	.557
F/F	M/F	p_2	3.20	**.021**	0.59	.096
		p_3	-45.83	< **.001**	-32.57	< **.001**
F/F	M/M	p_2	-13.61	< **.001**	-17.99	< **.001**
		p_3	-7.51	**.012**	-9.46	**.001**
F/M	M/F	p_2	10.71	< **.001**	5.37	**.002**
		p_3	-48.75	< **.001**	-33.26	< **.001**
F/M	M/M	p_2	-6.11	**.044**	-13.21	< **.001**
		p_3	-10.42	**.001**	-10.15	< **.001**
M/F	M/M	p_2	-16.82	< **.001**	-18.58	< **.001**
		p_3	38.33	< **.001**	23.11	< **.001**

(b) Means of the differences between values of p_2 and p_3. Significant p-values are marked in bold.

Table 27: Experiment 3 & 4, Varying Amounts of Disablers and Alternatives: Mean percentages of the individuals' values for p_2 and p_3 in the reduced inference case in the 'Alternatives' group. Means of the differences between individuals' values for p_2 and p_3. ('D' - Disablers, 'A' - Alternatives, 'F' - Few, 'M' - Many, p_i - Parameter)

D/A	p_i	**Exp. 3**	**Exp. 4**
		Mean	
F/F	p_2	15.58	23.15
	p_3	5.31	5.05
F/M	p_2	25.01	21.61
	p_3	4.44	4.12
M/F	p_2	8.47	9.55
	p_3	43.38	37.68
M/M	p_2	23.69	31.18
	p_3	17.25	18.90

(a) Mean values of p_2 and p_3.

D/A 1	**D/A 2**	p_i	**Exp. 3**		**Exp. 4**	
			Mean Δ	**p-value**	**Mean Δ**	**p-value**
F/F	F/M	p_2	-9.44	**.006**	1.54	.845
		p_3	0.87	.391	0.94	.088
F/F	M/F	p_2	7.10	**.006**	13.60	**.002**
		p_3	-38.06	< **.001**	-32.63	< **.001**
F/F	M/M	p_2	-8.11	**.019**	-8.03	.055
		p_3	-11.94	< **.001**	-13.85	< **.001**
F/M	M/F	p_2	16.54	< **.001**	12.06	**.001**
		p_3	-38.94	< **.001**	-33.57	< **.001**
F/M	M/M	p_2	1.33	.475	-9.56	**.014**
		p_3	-12.81	< **.001**	-14.78	< **.001**
M/F	M/M	p_2	-15.22	< **.001**	-21.63	< **.001**
		p_3	26.12	< **.001**	18.78	**.007**

(b) Means of the differences between values of p_2 and p_3. Significant p-values are marked in bold.

Suppression Effect by Providing Disablers and Alternatives Participants in these experiments are divided in three groups, similarly to Byrne's (1989) experiment where it was examined whether both fallacies *and* valid inference forms can be suppressed. It should be noted that the presentation form in these experiments differs slightly from the one that Byrne (1989) used. Alternative and additional arguments (*alternatives* and *disablers*) were not given as conditional rules equivalent to the major premise's form, but rather as a list of multiple alternatives or disablers. Nevertheless, the suppression effect was provoked successfully.

Participants in the 'Alternatives' group are provided with alternatives or alternative arguments that suppress the endorsements of the fallacies, DA and AC. The ones in the 'Disablers' group are provided with disablers or additional arguments that suppress the endorsements of the valid inference forms, MP and MT.

All four of ϵ-MMT's equations for inference form endorsement are of the form $\frac{x}{x+y}$:

$$\textbf{MP:}\quad \mathrm{P(Y|X)} = \frac{p_4}{p_3 + p_4} \qquad \textbf{DA:}\quad \mathrm{P(\neg Y|\neg X)} = \frac{p_1}{p_1 + p_2}$$

$$\textbf{AC:}\quad \mathrm{P(X|Y)} = \frac{p_4}{p_4 + p_2} \qquad \textbf{MT:}\quad \mathrm{P(\neg X|\neg Y)} = \frac{p_1}{p_1 + p_3}$$

Suppressing the endorsements means $\frac{x}{x+y}$ should decrease. The value of that fraction drops as y increases. In the case of MP and MT y is p_3. Therefore, a significant increase in p_3 is expected in the 'Disablers' group where the aim is to suppress MP and MT. Similarly, in previous disablers and alternatives analysis, p_3 was the disablers parameter since it describes the probability of the world X ¬Y. In the case of DA and AC y is p_2. That means a significant increase in p_2 is expected in the 'Alternatives' group where DA and AC are suppressed. Additionally, p_2 was the alternatives parameter since it describes the probability of the world ¬X Y.

Participants in these experiments were provided with both reduced inference and conditional. In the reduced inference case individuals rely more on their background knowledge and are able to think of and integrate any possible disablers and alternatives when giving an endorsement. Moreover, the general idea behind provoking the suppression effect is to provide *additional* information to the one in the rule. Taking that into consideration, this analysis is performed only for the conditional case.

In order to determine whether the differences between parameter values are significant, a Mann-Whitney U test was performed. The analysis results of Experiment 3 are shown in Table 28. In the 'Disablers' group there is a significant increase in the p_3 parameter, as expected, in three out of four cases, the exception being the 'Many/Many' task, where there is an increase but it is not significant ($p = .153$). In the 'Alternatives' group there is a significant increase in the p_2 parameter for every task. The analysis results of Experiment 4 are presented in Table 29. It can be immediately seen that for both groups, 'Disablers' and 'Alternatives', the expected increases of p_3 and p_2 respectively are found and are statistically significant. Therefore it can be concluded that the suppression of the fallacious and valid inference forms can be accounted for by ϵ-MMT's parameters.

6.2.2 Experiments 5–6

These are the experiments presented in Singmann and Klauer (2011) and Section 3.2.2. Their goal was to establish a double dissociation between deductive and inductive instructions when validity and plausibility are pitted against each other.

Experiment 5

Models Fits All of the models, except for the Dual-Source Model, are able to fit this experiment's data. The goodness of fit measures values for each model are presented in Table 30.

In Table 30, the number of free parameters to model one *individual* is reported. ϵ-MMT needs three free parameters to model one task (4 data points per task). Given 4 tasks, the total number of free parameters to model an individual is $3 \times 4 = 12$.

The best performing models are ϵ-MMT and the Oaksford-Chater Probabilistic Model with an RMSE of 0.068 and 0.067, respectively, an RMSE_{adj} of 0.136 and 0.135, respectively, and R^2 values of 0.860 and 0.862, respectively. Both models need the same amount of free parameters, which is three per task, or, 12 per individual.

Out of the models that have invalid RMSE_{adj} values, the performance of the Suppositional Sequential model is worth noting, with a RMSE value that matches ϵ-MMT and Oaksford-Chater. However, the Suppositional model needs eight more parameters to fit an individual (or two more per task).

Prological vs. Counterlogical Conditionals As it has been mentioned when analyzing parameter values for varying amounts of alternatives and disablers, the main focus are the parameters p_2 and p_3. p_2 describes the probability of the world ¬X Y (*alternatives*) and p_3 describes the probability of the world X ¬Y (*disablers*). That said, in the case of prological conditionals, which have many alternatives, an increase in the value of p_2 is expected and for counterlogical conditionals, which have many

Table 28: Experiment 3, Suppression Effect: Mean percentages of the individuals' values for p_2 and p_3 in the conditional case. Means of the differences between individuals' values for p_2 and p_3. ('D'/'Dis' - Disablers, 'A'/'Alt' - Alternatives, 'F' - Few, 'M' - Many, 'Supp' - Suppression Effect Condition, 'Base' - Baseline, 'p_i' - Parameter)

D/A	Supp	p_i	Mean
F/F	Base	p_2	4.68
		p_3	2.49
	Dis	p_2	5.21
		p_3	7.50
	Alt	p_2	13.64
		p_3	4.07
F/M	Base	p_2	8.12
		p_3	2.65
	Dis	p_2	10.12
		p_3	5.04
	Alt	p_2	22.19
		p_3	2.99
M/F	Base	p_2	1.90
		p_3	20.69
	Dis	p_2	40.62
		p_3	19.86
	Alt	p_2	23.99
		p_3	38.54
M/M	Base	p_2	15.16
		p_3	10.43
	Dis	p_2	18.78
		p_3	13.44
	Alt	p_2	28.37
		p_3	9.66

(a) Mean values of p_2 and p_3.

D/A	Supp 1	Supp 2	p_i	Mean Δ	p-value
F/F	Base	Dis	p_2	-0.53	.279
			p_3	-5.01	**.005**
	Base	Alt	p_2	-8.96	**.001**
			p_3	-1.57	.283
F/M	Base	Dis	p_2	-2.00	**.048**
			p_3	-2.39	**.019**
	Base	Alt	p_2	-14.07	< **.001**
			p_3	-0.34	.367
M/F	Base	Dis	p_2	-2.28	.475
			p_3	-19.92	**.003**
	Base	Alt	p_2	- 10.02	< **.001**
			p_3	-3.27	.297
M/M	Base	Dis	p_2	-3.62	.084
			p_3	-3.01	.153
	Base	Alt	p_2	-13.21	**.002**
			p_3	0.77	.419

(b) Means of the differences between values of p_2 and p_3. Significant p-values are marked in bold.

Table 29: Experiment 4, Suppression Effect: Mean percentages of the individuals' values for p_2 and p_3 in the conditional case. Means of the differences between individuals' values for p_2 and p_3. ('D'/'Dis' - Disablers, 'A'/'Alt' - Alternatives, 'F' - Few, 'M' - Many, 'Supp' - Suppression Effect Condition, 'Base' - Baseline, 'p_i' - Parameter)

D/A	Supp	p_i	Mean
F/F	Base	p_2	4.93
		p_3	3.40
	Dis	p_2	3.70
		p_3	8.48
	Alt	p_2	23.15
		p_3	5.05
F/M	Base	p_2	9.57
		p_3	4.83
	Dis	p_2	8.48
		p_3	7.79
	Alt	p_2	21.61
		p_3	4.12
M/F	Base	p_2	2.20
		p_3	32.40
	Dis	p_2	3.11
		p_3	41.05
	Alt	p_2	9.55
		p_3	37.68
M/M	Base	p_2	24.10
		p_3	13.26
	Dis	p_2	21.69
		p_3	17.94
	Alt	p_2	31.18
		p_3	18.90

(a) Mean values of p_2 and p_3.

D/A	Supp 1	Supp 2	p_i	Mean Δ	p-value
F/F	Base	Dis	p_2	-0.69	.066
			p_3	-6.83	< **.001**
	Base	Alt	p_2	-15.90	< **.001**
			p_3	-3.77	**.004**
F/M	Base	Dis	p_2	-1.20	.176
			p_3	-3.02	**.001**
	Base	Alt	p_2	-8.67	**.002**
			p_3	0.57	.071
M/F	Base	Dis	p_2	-1.08	.176
			p_3	-22.38	**.001**
	Base	Alt	p_2	-9.89	**.001**
			p_3	-11.74	.078
M/M	Base	Dis	p_2	-4.93	**.015**
			p_3	-10.44	**.001**
	Base	Alt	p_2	-14.80	< **.001**
			p_3	-2.03	.127

(b) Means of the differences between values of p_2 and p_3. Significant p-values are marked in bold.

Table 30: Fitting results for each model on experiment 5. 'Free' - Number of free parameters needed to model one individual.

Model	RMSE	$RMSE_{adj}$	R^2	Free
ϵ-MMT	0.068	0.136	0.860	12
Oaksford-Chater	0.067	0.135	0.862	12
Logistic Regression	0.086	0.172	0.582	12
MMT	0.126	∞	0.634	16
MMT Deductive	0.151	0.302	0.483	16
MMT with Directionality	0.084	nan	0.776	20
Suppositional Sequential	0.068	nan	0.759	20
Suppositional Exclusive	0.102	nan	0.652	20

disablers, an increase in the value of p_3 is expected. In this experiment, participants were divided in two groups, one of them received instructions to reason deductively, and the other one - inductively. Only the latter group will be taken into consideration here, given that deductive instructions suggest ignoring any background knowledge, which is highly relevant for analyzing the effect of prological and counterlogical conditionals.

Table 31: Experiment 5, Prological vs. Counterlogical Conditionals: Mean percentages of the individuals' values for p_2 and p_3 under inductive instructions. Means of the differences between individuals' values for p_2 and p_3. ('Type' - Conditional Type, 'Pro' - Prological, 'Count' - Counterlogical, 'p_i' - Parameter)

Type	p_i	Mean
Pro	p_2	11.20
	p_3	1.79
Count	p_2	4.40
	p_3	16.04

(a) Mean values of p_2 and p_3.

Type 1	Type 2	p_i	Mean Δ	p-value
Pro	Count	p_2	6.80	**.001**
		p_3	-14.24	< **.001**

(b) Means of the differences between values of p_2 and p_3. Significant p-values are marked in bold.

In order to determine whether the differences between parameter values are significant, a Wilcoxon signed-rank test was performed. As Table 31 shows, p_2 has a significantly ($p = .001$) higher value in the prological case and p_3 has a significantly ($p < .001$) higher value in the counterlogical case, which confirms the hypothesis.

Deductive vs. Inductive Instructions The characterization of conditionals was discussed in Section 1.3 and the *material implication* was introduced which focuses on logical validity. Based on that interpretation of conditionals the world X ¬Y violates

Table 32: Experiment 5, Deductive vs. Inductive Instructions: Mean percentages of the individuals' values for p_3. Means of the differences between individuals' values for p_3. ('Type' - Conditional Type, 'Pro' - Prological, 'Count' - Counterlogical, 'Instr' - Instructions, 'Ded' - Deductive, 'Ind' - Inductive, 'p_i' - Parameter)

Type	**Instr**	p_i	**Mean**
Pro	Ded	p_3	3.70
	Ind		5.96
Count	Ded	p_3	1.79
	Ind		16.04

(a) Mean values of p_3.

Type	**Instr 1**	**Instr 2**	p_i	**Mean Δ**	**p-value**
Pro	Ded	Ind	p_3	1.91	**.042**
Count	Ded	Ind	p_3	-10.07	< **.001**

(b) Means of the differences between values of p_3. Significant p-values are marked in bold.

the conditional and is therefore regarded as invalid. The probability of that world is described by ϵ-MMT with the parameter p_3. Under deductive instructions, individuals are encouraged to prioritize logical validity over their background knowledge, therefore a decrease in p_3 is expected in contrast to inductive instructions.

In order to determine whether the differences in the parameter values are statistically significant, a Mann-Whitney U test was performed. The results, as presented in Table 32 show that for prological conditionals, the expected change in p_3 is not present, instead there is a small but significant ($p = .042$) *decrease*. However, for counterlogical conditionals, p_3 is significantly ($p < .001$) smaller under deductive instructions, as expected.

Experiment 6

Models Fits All of the models, except for the Dual-Source Model, are able to fit this experiment's data. The goodness of fit measures values for each model are presented in Table 33.

In Table 33 the number of free parameters to model one *individual* is reported. ϵ-MMT needs three free parameters to model one task (4 data points per task). Given 9 tasks, the total number of free parameters to model an individual is $3 \times 9 = 27$.

The best performing model is the Suppositional Model (Sequential Variant) with an RMSE of 0.057, and R^2 value of 0.922. However, the model needs five free parameters per task, or, 45 per individual, which is more than the number of data points per individual (thus the invalid $RMSE_{adj}$ value).

Out of the models that have a reasonable number of parameters, the two best performing ones are ϵ-MMT and the Oaksford-Chater Probabilistic Model with an

RMSE of 0.075 and 0.072, respectively, an $RMSE_{adj}$ of 0.150 and 0.143 respectively and R^2 values of 0.860 and 0.913 respectively. Both models need the same amount of free parameters, which is three per task, or 27 per individual.

Table 33: Fitting results for each model on experiment 6. 'Free' - Number of free parameters needed to model one individual.

Model	RMSE	$RMSE_{adj}$	R^2	Free
ϵ-MMT	0.075	0.150	0.860	27
Oaksford-Chater	0.072	0.143	0.913	27
Logistic Regression	0.089	0.178	0.884	27
MMT	0.151	∞	0.769	36
MMT Deductive	0.170	0.340	0.696	27
MMT with Directionality	0.113	nan	0.864	45
Suppositional Sequential	0.057	nan	0.922	45
Suppositional Exclusive	0.124	nan	0.806	45

Prological vs. Counterlogical vs. Neutral Conditionals Similarly to the analysis in Experiment 5, the effect of prological and counterlogical conditionals is analyzed here. In this experiment a third type of conditionals is introduced, the neutral conditional, which has both many disablers and alternatives. The same assumption is made in this experiment – an increased value of p_2 for the prological conditionals is expected and an increased value of p_3 for the counterlogical ones. Additionally, an increased value for both p_2 and p_3 is expected for the neutral conditionals. As in Experiment 5, participants were divided in two groups, one received instructions to reason deductively and the other one inductively and only participants from the latter group are taken into consideration, given that integration of background knowledge is "allowed" only in inductive reasoning.

In order to determine whether the differences between parameter values are significant, a Wilcoxon signed-rank test was performed, whose results are presented in Table 34. When comparing prological to counterlogical conditionals the expected decrease of p_2 in the latter type is encountered ($p = .034$) along with an increase in p_3 ($p < .001$). In the comparison between prological and neutral conditionals an increase only in p_3 would be expected, which is found ($p < .001$). However a decrease in p_2 is also present ($p < .001$). This is, however, not the case when comparing counterlogical and netural conditionals, where the only expected change was an increase in p_2 in the neutral case, but what can be observed is an unexpected significant *decrease* ($p < .001$). When looking at the mean value of p_2 for neutral conditionals in Table

34a it can be seen that it has a surprisingly low value. Simultaneously, p_2 has an inexplicably high mean value for counterlogical conditionals.

Table 34: Experiment 6, Prological vs. Counterlogical vs. Neutral Conditionals: Mean percentages of the individuals' values for p_2 and p_3 under inductive instructions. Means of the differences between individuals' values for p_2 and p_3. ('Type' - Conditional Type, 'Pro' - Prological, 'Count' - Counterlogical, 'Neut' - Neutral, 'p_i' - Parameter)

Type	p_i	**Mean**
Pro	p_2	29.20
	p_3	2.16
Count	p_2	19.86
	p_3	13.58
Neut	p_2	4.98
	p_3	17.02

(a) Mean values of p_2 and p_3.

Type 1	**Type 2**	p_i	**Mean Δ**	**p-value**
Pro	Count	p_2	9.34	**.034**
		p_3	-11.42	< **.001**
Pro	Neut	p_2	24.22	< **.001**
		p_3	-14.87	< **.001**
Count	Neut	p_2	14.87	< **.001**
		p_3	-3.44	.672

(b) Means of the differences between values of p_2 and p_3. Significant p-values are marked in bold.

Deductive vs. Inductive Instructions Similarly to experiment 5, based on the material implication interpretation of conditionals, which focuses on logical validity, it is expected that the parameter p_3 will be smaller for participants taking part in the group that received deductive instructions, in contrast to the one that received inductive instructions.

In order to determine whether the differences in the parameter values are statistically significant, a Mann-Whitney U test was performed. The results in Table 35 show that for each conditional type, the decrease of p_3 in the deductive case is indeed present and significant ($p = .009$ for prological conditionals and $p < .001$ for the other two types).

6.2.3 Experiments 7–8

These are the experiments presented in Singmann et al. (2014) and Section 3.2.3. Their goal was to perform an empirical test of normative standards in the new paradigm and analyze the interpretation of P("If X then Y") through a novel probabilized conditional task.

As mentioned in Section 4.2.1, one participant has been removed from the data (Participant ID: U18U), due to the lack of variance in their answers that leads to NaN R^2 values.

Table 35: Experiment 6, Deductive vs. Inductive Instructions: Mean percentages of the individuals' values for p_3. Means of the differences between individuals' values for p_3. ('Type' - Conditional Type, 'Pro' - Prological, 'Count' - Counterlogical, 'Neut' - Neutral, 'Instr' - Instructions, 'Ded' - Deductive, 'Ind' - Inductive, 'p_i' - Parameter)

Type	Instr	p_i	Mean
Pro	Ded	p_3	1.78
	Ind		2.16
Count	Ded	p_3	4.57
	Ind		13.58
Neut	Ded	p_3	3.83
	Ind		17.02

(a) Mean values of p_3.

Type	Instr 1	Instr 2	p_i	Mean Δ	p-value
Pro	Ded	Ind	p_3	-0.37	**.009**
Count	Ded	Ind	p_3	-9.01	< **.001**
Neut	Ded	Ind	p_3	-13.19	< **.001**

(b) Means of the differences between values of p_3. Significant p-values are marked in bold.

Experiment 7

Model Fits The only two models that can fit the experiment's data are ϵ-MMT and the Oaksford-Chater Probabilistic Model. The goodness of fit measures values for both of them are presented in Table 36.

In Table 36 the number of free parameters to model one *individual* is reported. ϵ-MMT needs three free parameters to model one task (participants provided eight answers per task, however only six of them are used when fitting the models). Given 4 tasks, the total number of free parameters to model an individual is $3 \times 4 = 12$.

The best performing model is ϵ-MMT with an RMSE of 0.189, an RMSE_{adj} of 0.268 and a R^2 value of 0.497. It needs three free parameters per task, as its competitor, the Oaksford-Chater Probabilistic Model.

Table 36: Fitting results for each model on experiment 7. 'Free' - Number of free parameters needed to model one individual.

Model	RMSE	RMSE_{adj}	R^2	Free
ϵ-MMT	0.189	0.268	0.497	12
Oaksford-Chater	0.260	0.367	0.045	12

P(Conditional) and P(Material Conditional) The participants in this experiment provided estimates to the probability of the conditional P("If X then Y") and to the probability of the material conditional, in its logically equivalent form $P(\neg X \vee Y)$.

Evans et al. (2003) conducted an experimental study in order to determine whether individuals equate the probability of a conditional, P("If X then Y") with: (1) that one of the material conditional, $1 - P(X \wedge \neg Y)$; (2) conditional probability, $P(Y|X)$; and (3) conjunctive probability, $P(X \wedge Y)$. They concluded that the first hypothesis can be rejected, whereas half of the participants provided support for the second one and the other half for the third one. Based on that, the following is taken into consideration: the conditional probability, expressed with ϵ-MMT parameters as $\frac{p_4}{p_3+p_4}$, and, the conjunctive probability, expressed with ϵ-MMT's p_4 parameter. Their correlation with the individuals' answers to P("If X then Y") is analyzed.

On the other hand, given the definition of the material conditional, where the world X $\neg$Y, falsifies the rule. That world is described with the parameter p_3, and a negative correlation between it and the individuals' answers to P(Material Conditional) is expected.

The Kendall rank correlation coefficients can be found in Table 37. It can be immediately observed that every single correlation coefficient is significant. In the case of P(Conditional), both the p_4 parameter and the conditional probability, $\frac{p_4}{p_3+p_4}$ have a positive correlation ($p < .001$ for both) which confirms the assumptions above. In the case of the material conditional, the p_3 parameter has a negative correlation which also confirms the assumption.

Table 37: Analysis of correlation between parameter values and subjective estimates of P(Conditional) and P(Material Conditional) in experiment 7. Significant p-values are marked in bold. ('p_i' - Parameter)

Probability	p_i	τ	**p-value**
Conditional	p_4	0.465	< **.001**
	$\frac{p_4}{p_3+p_4}$	0.578	< **.001**
Material Conditional	p_3	-0.525	< **.001**

Experiment 8

Model Fits All of the models, except for the Dual-Source Model, are able to fit this experiment's data. The goodness of fit measures values for each model are presented in Table 38.

In Table 38 the number of free parameters to model one *individual* is reported. ϵ-MMT needs three free parameters to model one task (4 data points per task). Given 2 tasks, the total number of free parameters to model an individual is $3 \times 2 = 6$.

The best performing model is the Suppositional Model (Sequential Variant) with an RMSE of 0.031 and R^2 value of 0.917. However, the model needs five free parameters per task, or ten per individual, which is more than the number of data points per individual (thus the invalid $RMSE_{adj}$ value).

Out of the models that have a reasonable number of parameters, the two best performing ones are ϵ-MMT and the Oaksford-Chater Probabilistic Model with an RMSE of 0.054 and 0.051, an $RMSE_{adj}$ of 0.107 and 0.101 and R^2 of 0.910 and 0.911, respectively. Both models need the same amount of free parameters, which is three per task, or six per individual.

Table 38: Fitting results for each model on experiment 8. 'Free' - Number of free parameters needed to model one individual.

Model	RMSE	$RMSE_{adj}$	R^2	Free
ϵ-MMT	0.054	0.107	0.910	6
Oaksford-Chater	0.051	0.101	0.911	6
Logistic Regression	0.059	0.118	0.743	6
MMT	0.136	nan	0.669	8
MMT Deductive	0.138	0.277	0.654	6
MMT with Directionality	0.100	nan	0.787	10
Suppositional Sequential	0.031	nan	0.917	10
Suppositional Exclusive	0.095	nan	0.794	10

Inductive vs. Deductive Reasoning

It is expected when individuals are instructed to judge the logical validity of a rule that their interpretation of the conditional would match the material implication definition. The fact that the world X $\neg$Y makes the conditional invalid, leads to a hypothesis that the p_3 parameter should have a lower value in the deductive case, in contrast when individuals are reasoning inductively.

In order to determine whether the differences between parameter values are significant, a Mann-Whitney U test was performed. The results presented in Table 39 show that p_3 is significantly lower ($p < .001$) under deductive instructions, which confirms the hypothesis.

Table 39: Experiment 7 and 8, Deductive vs. Inductive Instructions: Mean percentages of the individuals' values for p_3. Means of the differences between individuals' values for p_3. ('Instr' - Instructions, 'Ded' - Deductive, 'Ind' - Inductive, p_i - Parameter)

Instr	p_i	**Mean**
Ded	p_3	1.68
Ind	p_3	32.00

(a) Mean values of p_3.

Instr 1	**Instr 2**	p_i	**Mean Δ**	**p-value**
Ded	Ind	p_3	-30.28	< **.001**

(b) Means of the differences between values of p_3. Significant p-values are marked in bold.

7 Discussion

To summarize, in this thesis, a conditional reasoning benchmark was developed, the performance of various cognitive models was evaluated, a new model was proposed and its performance, as well as its parameters' psychological interpretation and capability of accounting for effects, was analyzed.

Benchmark Form The benchmark I developed aims to test the models' limits in terms of adapting to many reasoning properties while maintaining a satisfying performance. In order to achieve that, I included data from experiments that manipulate the degree of background knowledge integration and induce various reasoning effects, leading to non-uniformity in the individuals' answers. Apart from that, experiments that focus on deductive reasoning were also included which shed some light on the limitations of the normally best-performing probabilistic models. Simultaneously, the benchmark also provides performance analysis for each experiment separately, showing how a model's performance can vary given different experiment scenarios. Now, focusing on the models – including theories from both the deductive and the Bayesian paradigm is also essential. Even though probabilistic models are considered state-of-the-art and often outperform their deductive rivals, that does not make all assumptions and approaches from the deductive paradigm invalid. Aside from that, some of the theories had multiple variants that were included which offers the possibility to investigate how models can be improved and how important or valid some of their assumptions are. Finally, this benchmark uses the same, very simple fitting approach for *all* models. It optimizes the models' prediction equations so that they achieve a minimal RMSE. This way all of the models are compared in a fair way and not one model is favored because of the way it has been fit.

Model Performance In the benchmark six different theories were included, out of which one had three model variants (MMT) and one had two (Suppositional Theory), totaling up to nine models. When focusing on the overall performance, the Dual-Source Model is unfortunately excluded, as it can be fit to only half of the available data, leaving us with eight models to focus on. The best performing models in the benchmark are ϵ-MMT and Oaksford-Chater. When inspecting only models from the deductive paradigm, the best performance is achieved by the Suppositional Sequential model. The goodness-of-fit was also analyzed for each experiment separately. All eight models can be applied to 7 out of 8 experiments. In 4 out of those 7, ϵ-MMT and Oaksford-Chater had the best performance. The Suppositional Sequential model tied with them in Experiment 5, and in other two (Experiment 6 and 8) it actually outperformed the probabilistic models. Interestingly, the experiments where the Suppositional Sequential model dominates are related to deductive reasoning. In Experiment 5 and 6 (around) half of the participants were instructed to determine logical validity and ignore background knowledge. The same instructions to judge logical validity were provided to all participants in Experiment 8 and additionally the conditional contents were abstract, i.e. no possible background knowledge would be integrated. While ϵ-MMT's and Oaksford-Chater's performance in those experiments is still satisfactory, it is quite valuable to learn how they do not manage to represent and fit data from a deductive experiment as good as a model from the deductive paradigm. Participants in Experiment 7 were asked to provide subjective probabilities *other* than inference form endorsements and only ϵ-MMT and Oaksford-Chater were able to represent them. In that scenario, ϵ-MMT actually surpasses its main competitor. Both models have the same number of *free* parameters (3), however ϵ-MMT has one more dependent parameter which seems to aid when expressing a larger number of various types of probabilities. Given that conditional reasoning research has been largely focused on inference form endorsements, representing other subjective probabilities has not been a major focus or requirement when developing cognitive models. However, seeing as those probabilities offer yet another perspective on how individuals interpret conditionals it is certainly a useful model feature. The Dual-Source Model could be applied in Experiment 1 – 4, however it did not perform as well as ϵ-MMT and Oaksford-Chater. It should be pointed out that Singmann et al. (2016) report a more satisfying performance of the model, however they used a different fitting approach. In this benchmark, a uniform and very simple way of fitting was used that should not give advantage to any model. It is likely that different fitting approaches would benefit some models more than others and would maybe expose some undiscovered potential, however, right now, that was not a goal when developing this benchmark.

Model Variants Comparison Another objective was to analyze the difference between variants of the same theory. In the case of MMT, the Directionality variant had the best performance, meaning that directionality from the antecedent to the consequent is indeed something that should be taken into consideration. As a matter of fact, the mean value of the parameter d that determines whether directionality will be reversed or not, across all data is 0.90 (median: 0.99) – in 90% of the cases, individuals actually reverse the directionality. The deductive variant has the worst performance out of all three variants and out of all models in the benchmark, demonstrating the real necessity of the X ¬Y model. It is noteworthy that even in experiments where deductive reasoning played a role, MMT Deductive was still the worst performing model. On the other hand, we have the Suppositional Theory variants, where the Sequential model's performance definitely exceeds the Exclusive one. Therefore, it can be inferred that given the assumption that humans use System 1 and System 2 when reasoning, it is more likely that those two systems interact, rather than individuals using either one or the other.

ϵ-MMT Performance In this thesis I proposed a new cognitive computational model, ϵ-MMT. I took the intuitive representations of conditional's propositions of the Mental Model Theory, interpreted them as possible worlds and applied a probability distribution to them, following Pearl's (1991) ϵ-semantics framework. Interpreting inference form endorsement as the conditional probability of the consequent given the antecedent makes this model belong to the state-of-the-art Bayesian paradigm. Its performance is competent, having the best results in the benchmark together with the Oaksford-Chater Probabilistic Model and outperforming the other two included state-of-the-art models, the Logistic Regression Model and DSM. As already discussed above, ϵ-MMT is also capable of representing subjective probabilities other than inference form endorsements, giving it the possibility to fit a larger variety of data that provides insight into how humans reason and interpret conditionals.

ϵ-MMT: Suppression Effect The benchmark contains data from experiments that manipulate a variety of reasoning properties and effects. Starting with the *suppression effect* – it can be confidently said that through its parameters, ϵ-MMT is able to represent the presence of this effect from various perspectives. There are two experiments, Experiment 3 and 4, which explicitly aim to induce the suppression effect by providing additional information to the participants. The parameters p_2 and p_3, which describe the probabilities of the worlds ¬X Y and X ¬Y respectively, account for the presence of alternative and additional information.

ϵ-MMT: Varying Amounts of Disablers and Alternatives In five experiments (Exp. 1, 3, 4, 5 and 6) the contents presented to the participants had varying amounts of disablers and alternatives. Similarly, through p_2 and p_3 the amount of alternatives and disablers can be shown. A high p_2 value indicates the presence of many alternatives, whereas a high p_3 value indicates many disablers. An interesting question in this situation is what does 'Few' or 'Many' mean for different contents. It is possible that an individual can think of 20 alternatives to one conditional rule and 40 to another yet both would be considered as 'Many' alternatives. However, in the latter case it is possible that the individual's subjective probability of the world ¬X Y would be much higher compared to the first one. When the results were analyzed, it was noted that in some cases when a significant change was expected in only one of these two parameters, one was found in both. For example, when going from a task with 'Few' disablers and 'Few' alternatives to one with 'Few' disablers and 'Many' alternatives, only p_2 is expected to increase. However, a change is found in p_3 as well. That leads to a possible difference in interpretation of what 'Few' disablers means for these two different contents. An alternative perspective is a potential relationship between the two. Following the same example, by introducing content with many alternatives, the confidence in those few disablers decreases, causing a significant decrease in p_2, where no change was expected.

In Experiment 3 and 4 participants were divided into three groups: 'Baseline', 'Disablers' and 'Alternatives', where the last two groups were provided with additional information. The influence of varying amounts of disablers and alternatives was investigated in all three groups. In the 'Disablers' group, it can be concluded that providing additional disablers does not affect the influence of the individuals' background knowledge, every expected change in the alternatives parameter, p_2 still takes place. Interestingly, there is one case, in Experiment 4 when comparing the task 'Few/Few' with 'Many/Few' where providing additional information led to having only an increase in p_3 and no change in p_2, as it actually is expected, contrary to every single result presented before where both parameters would have a significant change. A prevention of an unexpected parameter to have a significant change by providing additional information can be observed in the 'Alternatives' group as well when comparing the task 'Few/Many' with 'Many/Many' and the only expected and encountered change is an increase in p_3. Additionally, in the 'Alternatives' group of Experiment 4, when comparing 'Few/Few' with 'Many/Many' and both p_2 and p_3 are expected to increase, only p_3 does, which is logical due to the fact that the additional information presented to this group are alternatives. Similarly, when comparing 'Few/Few' with 'Few/Many' in the same group, the p_2 parameter did not have the expected significant change. Explicitly providing alternative information might impact the influence of the 'Few' alternatives in the first task, leading to lack of significant changes.

ϵ-MMT: Conditional Presentation Form Following is the *conditional presentation form*, which was manipulated in four experiments (Exp. 1, 2, 3, 4). When presented with a reduced inference individuals are expected to rely more on their background knowledge compared to when they are presented with a rule. In terms of ϵ-MMT's parameters, a stronger belief in the world X Y is expected when a rule is present. Experiment 2 was the only one that did not have contents with varying amounts of disablers and alternatives and a significant change in p_4 was found. In the other three experiments, an interesting finding is that for contents with 'Few' disablers and alternatives p_4 did not change significantly. The lack of significant changes is logical given that in such a task there would not be a great deal of contradicting background knowledge that would be either integrated when reasoning with a reduced inference or suppressed when presented with a conditional rule. That leads to a high belief in the world X Y, already in the reduced inference case. Additionally, in Experiment 1 and 3, p_4 lacked a significant change also for tasks with 'Few' disablers and 'Many' alternatives. Previously it was discussed how the presence of disablers and alternatives is successfully accounted for through the parameters p_2 and p_3, including this specific task. That means the high amount of alternatives is recognized by the individuals, yet the belief in the world X Y does not change significantly once they are presented with a rule. On one hand, this can be interpreted as a lack of influence of the alternatives. Their large amount is acknowledged, however, they are not powerful enough to significantly lower the belief in the world X Y. On the other hand, it is possible that individuals think of the specific antecedent as an event with a high probability and even if there are many possible alternatives, the world X Y is still very likely to happen.

In Experiment 1 a third conditional presentation form was included – the biconditional. Significant increases in p_4 were encountered between reduced inference and biconditional for every task except, once again, the one with 'Few' disablers and alternatives. The explanation for the lack of significant change in this task when analyzing reduced inference and conditional is applicable now as well. Interestingly, a significant change of p_4 when comparing conditional and biconditional forms was *not* found. That means once an individual has been presented with a conditional rule the belief in the world X Y becomes as strong as possible and therefore is not strengthened when presented with a biconditional.

ϵ-MMT: Deductive vs. Inductive Instructions Next is the analysis of the effect of a particularly interesting manipulation – *deductive vs. inductive instructions* (Exp. 5, 6, 7, 8). In the experiments, participants under deductive instructions are directed to judge logical validity of a conclusion and ignore any background knowledge whereas under inductive instructions they are asked for the likelihood of conclusions and are encouraged to integrate their background knowledge. Under deductive instructions an indication of a material conditional implication was found – the p_3 parameter, describing the probability of the world X ¬Y (which falsifies a conditional), is significantly smaller in the deductive cases. This discovery provides support for the material characterization of conditionals in the *deductive* paradigm. On the other hand, it simultaneously supports the abandonment of the characterization in the *Bayesian* paradigm when focusing on everyday content and taking integration of background knowledge into consideration. Ultimately, it shows that in this scenario there is no 'right' or 'wrong', but rather things are relative. In Experiment 5, for prological conditionals ('Many' alternatives and 'Few' disablers), the increase in p_3 was not found. As a matter of fact, a very small but significant *decrease* was encountered. It is likely that contents with 'Few' disablers induce very low p_3 values under inductive instructions and that could overshadow the effect of deductive instructions.

ϵ-MMT: Speaker Expertise The final property whose effect is analyzed is *speaker expertise*. In Experiment 2 participants received contents either uttered by an expert or a non-expert. The original assumption was that the belief in the world X Y will be stronger in the expert case, i.e. higher values of p_4 were expected. That was not the case, however, a very interesting outcome was a significant *decrease* of p_3 (X ¬Y) in the expert case. As discussed previously, low p_3 values are associated with material implication and logical validity. So, this result suggests that when contents have said to be uttered by experts individuals reason more *deductively*. Simultaneously, the conditionals' contents should also be taken into consideration. Singmann et al. (2016) pointed out: "Note that we nevertheless did not use abstract materials, but everyday contents for which participants prior knowledge was assumed to be rather vague. This ensured that the manipulation of speaker expertise could overshadow participants' prior knowledge.". It can be easily assumed that whether a non-expert or an expert has uttered a certain rule, given the lack of background knowledge about the topic, individuals' belief in the world X Y (described by the parameter p_4) would be on the same level, as the results have indeed indicated.

ϵ-MMT: Correlation Analysis Additionally, the correlation between parameter values and individuals' subjective probabilities for the conditional and material conditional was examined in Experiment 7. For P("If X then Y"), the focus was on the conjunction P(X∧Y) expressed through p_4 and on the conditional probability P(Y|X), expressed as $\frac{p_4}{p_3+p_4}$, both of which showed a significant positive correlation. A crucial fact here is that when fitting the parameters, the values of P("If X then Y") were not used. The probability distribution of the conditional's possible worlds was obtained through numerous other subjective probabilities. Nonetheless, ϵ-MMT still provides insight into the individuals' interpretation of the conditional from other perspectives, specifically – its probability. For the material conditional, the focus is on the parameter p_3, describing the probability of the world X ¬Y, and the expected negative correlation was found – the stronger the belief in the material conditional, the lower the probability of the logically incorrect world X ¬Y.

Number of Free Parameters When analyzing the models' performances in some cases invalid RMSE_{adj} values were obtained, either ∞ or *nan*. In the case of ∞ the models needed as many free parameters as provided data points in a task. *nan* was obtained when the models needed *more* parameters than available data points. This is not entirely an issue solely because of the RMSE_{adj} measure. Rather, the fact that a model would need as many parameters as data points, or even more, to represent one single task is questionable. This concern is relevant when fitting per *individual*, in contrast to fitting on aggregate data. Most of the available experimental data provides endorsements to the four inference forms. That automatically gives models needing at least four parameters to fit one task a disadvantage. The insistence on individual fitting is due to the differences between individuals. Factors like background knowledge or education level can contribute to variation in their answers.

Data The experiments whose data is used in this thesis have 50 participants on average. While a decent basis for conditional reasoning research is provided, having a larger and more diverse pool of people would allow for better insight into response patterns and variation. Generally, the amount of available data that can be used in such benchmarks is rather limited. Some experiments ask very few questions per task and cannot be properly modeled by the implemented models. An example of that are the experiments presented by Skovgaard-Olsen, Singmann, and Klauer (2016). Participants only gave three answers per task and therefore an accurate parameter value analysis was unattainable.

Finally, an important question is – does manipulating effects always produce the expected results? Consider, for instance, how Singmann and Klauer (2011) showed that the validity effect was *not* encountered in denial problems, even though it was expected. But, what does that tell us? Are higher endorsements of valid inference forms under deductive instructions not producible by human reasoning processes? Or, did the experiment's contents fail to trigger the effect? Increasing the amount of data with varying content could help answer questions of this nature.

7.1 Future Work

The benchmark I implemented provided very useful analysis for this thesis, but it is still in its early stages of development. The first expansion step is implementing more models from both, the deductive and the Bayesian paradigms. Additionally, including even more variants of the same theory is also beneficial in order to determine the strengths and weaknesses of a model – improvement being the ultimate goal. Then, the next step is to find and add more data. Including experiments that manipulate other reasoning properties and effects will give us even more insight into how humans interpret conditionals. Another interesting option could be to look into experimental data from the deductive paradigm. In those experiments participants either accept or refute a conclusion, in contrast to giving endorsements. Finally, augmenting the analysis options would definitely be a helpful feature. For now, the analysis is focused on the models' performances and ϵ-MMT's parameters. The next step would be to also provide an option to analyze the data.

In terms of fitting capabilities the only true competitor to ϵ-MMT is the Oaksford-Chater Probabilistic Model, both of them achieving the same results. However, the Oaksford-Chater model is rather mathematical and does not necessarily provide an account of how this theory would be implemented in the mind. As Oaksford and Chater (2003a) state themselves, they doubt that people actually perform those calculations when reasoning, however they are asking the question whether such a probabilistic, computational account provides a better computational level theory of the reasoning performance compared to logic. ϵ-MMT, on the other hand, provides a finer basis for psychological interpretation. However, given the mathematical definition of ϵ-MMT's and Oaksford-Chater's parameters their representation of subjective probabilities can actually be brought to the same level (see Appendix A.4). An extensive comparison between these two models is an interesting direction for further research.

Todorovikj et al. (2019) explored the possibility to reduce the number of ϵ-MMT's parameters. Focusing on varying amounts of disablers and alternatives and conditional presentation form, constant changes of parameter values between two tasks were established and used to *predict* parameter values when modeling one task after having fitted another. In this thesis a significant change of parameter values was discovered for various reasoning properties which lays basis for a deeper investigation of ϵ-MMT's predictive capabilities. Starting by reducing the number of parameters within single experiments, it would be of interest to achieve a universal predictive ability in the long run.

When introducing Experiment 5 and 6 it was mentioned that Singmann and Klauer (2011) examined the presence of the validity and plausibility effects. With the exception of the validity effect in denial problems, their expectations were met, the effects were found. An interesting next step would be to explore and get more insight into why the validity effect was not encountered using ϵ-MMT.

In this thesis, the effect of contents with varying amounts of alternatives and disablers was analyzed in multiple experiments. The main focus was on ϵ-MMT's parameters p_2 and p_3 which represent the probabilities of the worlds ¬X Y and X ¬Y, respectively. On multiple occasions, the results showed that when changing the quantifier of only one, disablers or alternatives, a significant change in both parameters was present, instead of one. I discussed the disadvantages of using only the quantifiers 'Few' and 'Many' as they can be perceived differently for different contents. It would be beneficial if this measure is done in a more precise way. Obviously it is not possible to state an exact number. However, perhaps conducting a pre-study where individuals would rank the amounts for specific contents would aid with refining this quantification. Another perspective that I offered was that changing the amount of one impacts the influence of the other. The validity of this suggestion can be examined with more experimental data that uses even more different contents. Additionally, it is very curious how the occasional lack of expected significant parameter value changes only occurred when alternatives were relevant, in contrast to disablers. Further research on how disablers and alternatives are perceived by individuals and additionally the difference in their influences on the belief in the X Y world would be of great interest.

8 Conclusion

In this final chapter, I will use the main points of the discussion to answer my research questions.

RQ 1: What would be a good form of a benchmark?

While working on this thesis, my idea of a good benchmark was one that would be able to answer a lot more questions rather than just "Which model performs the best overall?". I achieved that by:

- including data whose content is not uniform – exploiting various reasoning properties creates a bigger challenge for the models to adapt to atypical answers and we learn more about their potential and limitations.

- including models from both paradigms – even though in present day research the focus is on probabilistic models, deductive theories and their psychological assumptions are still relevant and should not be dismissed so swiftly.

- analyzing experiments separately – determining which model performs the best overall does not imply that it truly manages to capture every aspect of reasoning, it is useful to see where certain models fall short to learn how to potentially improve them.

- including variants of models that focus on different theoretical assumptions and/or improve the theories – by comparing the performance of these variants we learn whether certain assumptions actually help explain the reasoning processes and discover ways in which the theories can be improved.

- using the same fitting approach for all models – this way all models get the same treatment and are compared in a fair manner. Certainly a fitting approach can be adapted in a way that one model can benefit from it in contrast to others. That would be useful when trying to unveil the full potential of a theory alone. However, that idea does not coincide with the objective of my benchmark, which is a fair and equal comparison of models.

RQ 2: How do various cognitive models perform when evaluated on sensible data?

Out of four probabilistic and five deductive models, the two best performing ones are ϵ-MMT and Oaksford-Chater, both belonging to the Bayesian paradigm. The Suppositional Sequential model is a close competitor, sometimes even outpeforming the two in experiments related to deductive reasoning. So, while the probabilistic models excel at representing reasoning with everyday content, their performance does not stand out when dealing with abstract materials and/or deductive reasoning instructions. Ultimately, the conclusion is that no single theory or approach explored here was able to explain all reasoning processes in all scenarios.

RQ 3: Can ϵ-MMT compete with state-of-the-art models?

Yes. The benchmark results show that its fitting capabilities match the ones of the Oaksford-Chater Probabilistic Model. Moreover, its parameter definitions and psychological interpretations provide a valuable opportunity for investigating and learning about the human reasoning processes. Given that this is the initial theoretical basis for ϵ-MMT, with further research, improvements that would enhance its performance can be achieved, making it a valuable competitor among state-of-the-art models.

RQ 4: Can ϵ-MMT account for various reasoning effects and properties, and, provide insight into human reasoning and interpretation of conditionals?

Yes. The suppression effect was accounted for from various points of view in five different experiments. The goal in two of them was specifically to provoke the effect by providing additional information. In all five the effect of contents with varying amount of disablers and alternatives was shown with ϵ-MMT's parameters. In one the validity and plausibility effect (or lack thereof) were also accounted for through parameter value rankings. The effect of background knowledge dominance when providing individuals with a reduced inference in contrast to a rule is also expressed through ϵ-MMT parameters. Also, we gained insight into how people interpret conditional tasks when speaker expertise was manipulated. Most interestingly, with parameter value analysis it was also shown that when humans reason deductively the material implication interpretation of conditionals is actually valid compared to inductive reasoning. Additionally ϵ-MMT's parameters correlate to the subjective probabilities of the conditional and material conditional.

Acknowledgements

Firstly, I would like to extend my sincere gratitude to Apl. Prof. Dr. Dr. Marco Ragni for providing support, advice and giving me the opportunity to dive deep in a field that I became incredibly passionate about. I am extremely thankful for all of the trust that has been put in me and the freedom that I have been given, both of which inspired me to not see absolutely any limits to the things that I can do and learn. Ultimately, that enabled me to always keep trying to satisfy my scientific curiosity. I am also very grateful to Prof. Dr. Gabriele Kern-Isberner for showing interest in my work and being one of my examiners.

I would like to express my immeasurable appreciation and sincerely thank the people that took the time to proofread this thesis and gave me invaluable comments, advice and suggestions: Oleksandra Boychenko, Vasco Brazão, Elira Daja, Paulina Friemann, Konstantina Galani and Maria Volkova. I promise that I will stop putting so many commas everywhere.

Well, these were strange and often difficult times in *many* ways[1]. However, with the selfless love and support of my friends it all went well and there are no words that can properly express my gratitude.
Elira Daja and Konstantina Galani for the endless conversations, advice and support, for their will to be there for me beyond limits, at any time, no matter what the situation is and for always reminding me that I am never alone.
Ammar Nayal and Sweetin Paul for all of the patience that they have for me, and for the late card playing nights and early coffee-and-deep-talk mornings that kept me sane and made me feel safe and cared for.
Mariya Georgieva for all the Saturday distractions that teach me how to enjoy life even more, for brightening up my days with loud laughter and for always making me feel special and loved.
Désirée Schwindenhammer for supporting me all the way since my bachelors through

[1]Special shout-out to the global pandemic.

everything that led up to this moment and for always being right there next to me when I need her through thick and thin.
Meri Popovska for continuously showing me that love and care know no distance, for making me laugh with silly conversations, but also for picking up the phone and listening patiently every single time I call and for believing in me *all* of the time, сорѝ детé.
Daniel Sauter, the other half of the best thesis (and beyond) support group, for all of the late night working sessions and (often too long) coffee breaks, for always providing me with more comfort, moral support and encouragement than I could ever imagine receiving and generally for making life easy-peasy-lemon-squeezy.

Finally, my biggest thanks goes to my parents, Renata and Vlastimir Todorovikj. Thank you for doing absolutely everything in your power to be there for me in every possible way. Thank you for always believing in me and constantly showing me that I can achieve everything that I set my mind to. Thank you for bearing with me through every single thing that is happening in my life and for always having my back no matter what. I would've never gotten here without you, your love and support. *"Ohana means family and family means nobody gets left behind, or forgotten."*

Well, one last *very* important thank you goes to coffee. I think this one is quite self-explanatory.

References

Baratgin, J., Over, D. E., & Politzer, G. (2013). Uncertainty and the de finetti tables. *Thinking & Reasoning*, *19*(3 – 4), 308 – 328.

Byrne, R. M. (1989). Suppressing valid inferences with conditionals. *Cognition*, *31*(1), 61–83.

Chan, D., & Chua, F. (1994). Suppression of valid inferences: Syntactic views, mental models, and relative salience. *Cognition*, *53*(3), 217 – 238.

de Finetti, B. (1936/1995). The logic of probability [translation of 1936 original]. *Philosophical Studies*, *77*, 181 – 190.

Elqayam, S., & Over, D. E. (2013). New paradigm psychology of reasoning. *Thinking & Reasoning*, *19*(3–4), 249–265.

Evans, J. (1972). Interpretation and matching bias in a reasoning task. *Quarterly Journal of Experimental Psychology*, *24*(2), 193–199.

Evans, J. (1993). The mental model theory of conditional reasoning: Critical appraisal and revision. *Cognition*, *48*(1), 1–20.

Evans, J., Handley, S., & Over, D. (2003). Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(2), 321–335.

Evans, J., & Over, D. E. (2004). *If.* Oxford University Press.

Grice, H. P. (1989). *Studies in the way of words.* Harvard University Press.

Hadjichristidis, C., Stevenson, R. J., Over, D. E., Sloman, S. A., Evans, J., & Feeney, A. (2001). On the evaluation of 'if p then q' conditionals. In *Proceedings of the twenty-third annual conference of the cognitive science society* (pp. 409–414). Lawrence Erlbaum Associates.

Johnson-Laird, P. N., & Byrne, R. M. (1991). *Deduction.* Lawrence Erlbaum Associates, Inc.

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

Johnson-Laird, P. N., & Tagart, J. (1969). How implication is understood. *The American Journal of Psychology*, *82*(3), 367–373.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, *30*(1/2), 81–93.

Khemlani, S., & Johnson-Laird, P. N. (2012). Theories of the syllogism: A meta-analysis. *Psychological bulletin*, *138*(3), 427–458.

Klauer, K. C., Beller, S., & Hütter, M. (2010). Conditional reasoning in context: A dual-source model of probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 298–323.

Lewis, D. (1973). *Counterfactuals.* Oxford University Press.

Liu, I. M., Lo, K. C., & Wu, J. T. (1996). A probabilistic interpretation of "if-then". *The quarterly journal of experimental psychology section A*, *49*(3), 828 – 844.

Manktelow, K. I. (1999). *Reasoning and thinking.* Psychology Press.

Manktelow, K. I. (2012). *Thinking and reasoning: An introduction to the psychology of reason, judgment and decision making.* Psychology Press.

Mann, H., & Whitney, D. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, *18*(1), 50–60.

Markovits, H. (1984). Awareness of the 'possible' as a mediator of formal thinking in conditional reasoning problems. *British Journal of Psychology*, *75*(3), 367–376.

Markovits, H. (1985). Incorrect conditional reasoning among adults: Competence or peformance? *British Journal of Psychology*, *76*(2), 241–247.

Marr, D. (1982). *Vision.* San Francisco: W.H. Freeman.

Oaksford, M., & Chater, N. (1995). Theories of reasoning and the computational explanation of everyday inference. *Thinking & Reasoning*, *1*(2), 121 – 152.

Oaksford, M., & Chater, N. (2003a). Conditional probability and the cognitive science of conditional reasoning. *Mind & Language*, *18*(4), 359 – 379.

Oaksford, M., & Chater, N. (2003b). Probabilities and pragmatics in conditional inference: Suppression and order effects. In *Thinking: Psychological perspectives on reasoning, judgement and decision making* (pp. 95–122). John Wiley & Sons Ltd.

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

Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(4), 883.

Oberauer, K. (2006). Reasoning with conditionals: A test of formal models of four theories. *Cognitive Psychology*, *53*(3), 238–283.

Over, D. (2004). The psychology of conditionals. In *Psychology of reasoning: Theoretical and historical perspectives* (pp. 75–94). Psychology Press.

Over, D., & Evans, J. (2003). The probability of conditionals: The psychological evidence.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann Publishers, Inc.

Pearl, J. (1991). *Epsilon-semantics* (Tech. Rep.). Computer Science Department, University of California.

Ramsey, F. (1931/2013). *Foundations of mathematics and other logical essays*. Martino Fine Books.

Riesterer, N., Brand, D., & Ragni, M. (2020). Do models capture individuals? evaluating parameterized models for syllogistic reasoning. In *Proceedings of the 42nd annual conference of the cognitive science society* (pp. 3377 – 3383).

Rips, L. J., & Marcus, S. L. (1977). Suppositions and the analysis of conditional sentences. In *Cognitive processes in comprehension* (pp. 185–220).

Rumain, H., Connell, J., & Braine, M. (1983). Conversational comprehension processes are responsible for reasoning fallacies in children as well as adults: If is not the biconditional. *Developmental psychology*, *19*(4), 471.

Russel, S. J., & Norvig, P. (2010). *Artificial intelligence – a modern approach* (3rd ed.). Pearson Education.

Schroyens, W. J., Schaeken, W., & d'Ydewalle, G. (2001). The processing of negations in conditional reasoning: A meta-analytic case study in mental model and/or mental logic theory. *Thinking & Reasoning*, *7*(2), 121 – 172.

Singmann, H., & Klauer, K. C. (2011). Deductive and inductive conditional inferences: Two modes of reasoning. *Thinking & reasoning*, *17*(3), 247–281.

Singmann, H., Klauer, K. C., & Beller, S. (2016). Probabilistic conditional reasoning: Disentangling form and content with the dual-source model. *Cognitive Psychology*, *88*, 61–87.

Singmann, H., Klauer, K. C., & Over, D. (2014). New normative standards of conditional reasoning and the dual-source model. *Frontiers in psychology*, *5*, 316.

Skovgaard-Olsen, N. (2016). Ranking theory and conditional reasoning. *Cognitive Science*, *40*(4), 848–880.

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

Spohn, W. (2009). A survey of ranking theory. *Degrees of belief*, 185–228.

Stalnaker, R. (1968). A theory of conditionals. In *Ifs* (pp. 41–55). Springer.

Stalnaker, R. (1975). Indicative conditionals. *Philosophia*, *5*, 269–286.

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

Todorovikj, S., Friemann, P., & Ragni, M. (2019). Combining mental models and probabilities: A new computational cognitive approach for conditional reasoning.

Wason, P. (1966). Reasoning. In *New horizons in psychology* (pp. 135–151). Penguin Books.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, *1*(6), 80–83.

Appendices

A.1 Contents in Experiments

A.1.1 Experiment 1

The relevant details of this experiment are provided in Section 3.2.1. For more, see Singmann et al. (2016): Experiment 1, and, `https://osf.io/zcdfq/`.

In the experiment four different contents were presented in German. Both the original German version and the corresponding translations in English follow:

1. Few Disablers/Few Alternatives: If a predator is hungry then it will search for prey. (Wenn ein Raubtier Hunger hat, dann geht es auf Suche nach Beute.)

2. Few Disablers/Many Alternatives: If a balloon is pricked with a needle then it will pop. (Wenn ein Ballon mit einer Nadel gestochen wurde, dann platzt er.)

3. Many Disablers/Few Alternatives: If a girl has sexual intercourse then she will be pregnant. (Wenn ein Mädchen Geschlechtsverkehr hat, dann ist es schwanger.)

4. Many Disablers/Many Alternatives: If a person drinks a lot of coke then the person will gain weight. (Wenn eine Person viel Cola trinkt, dann nimmt sie an Gewicht zu.)

A.1.2 Experiment 2

The relevant details of this experiment are provided in Section 3.2.1. For more, see Singmann et al. (2016): Experiment 2, and, `https://osf.io/zcdfq/`.

In the experiment six different contents (out of total seven) were presented in German. Both the original German version and the corresponding translations in English follow:

1. A nutrition scientist *(expert)* / drugstore clerk *(non-expert)* says: If Anne eats a lot of parsley then the level of iron in her blood will increase. (Ein Ernährungswissenschaftler *(expert)* / eine Drogerieangestellte *(non-expert)* sagt: Wenn Anne viel Petersilie isst, dann verbessert sich ihr Eisenwert im Blut.)

2. A ballet dancer *(expert)* / ballet audience member *(non-expert)* says: If Katherina grows 10 cm then her ballet partner will fail to perform the lifting routine with her. (Ein Balletttänzer *(expert)* / Ballettbesucher *(non-expert)* sagt: Wenn Katharina 10 cm wächst, dann scheitert ihr Balletpartner an der Hebefigur mit ihr.)

3. A manager *(expert)* / cashier *(non-expert)* says: If the shampoo "Fresh and Soft" is removed from the assortment then the yearly volume of sales will decrease. (Ein Filialleiter *(expert)* / ein Kassierer *(non-expert)* sagt: Wenn das Haarshampoo "Fresh and Soft" aus dem Sortiment der Drogerie genommen wird, dann sinken die Jahresumsatzzahlen.)

4. A pilot *(expert)* / passenger *(non-expert)* says: If the Airbus A380 flies through turbulences then its voltage level will fluctuate. (Ein Pilot *(expert)* / Fluggast *(non-expert)* sagt: Wenn der Airbus A380 durch Turbulenzen fliegt, dann schwankt die Spannung im Stromnetz der Maschine.)

5. A(n) environmental scientist *(expert)* / newspaper reader *(non-expert)* says: If individuals in industrialized nations continue to emit similar amounts of CO2 then the Gulf stream will stop. (Ein Umweltwissenschaftler *(expert)* / Zeitungsleser *(non-expert)* sagt: Wenn die Menschen in den Industrienationen weiterhin so viel CO2 ausstoßen, dann wird der Golfstrom zum Erliegen kommen.)

6. A(n) asset consultant *(expert)* / bank teller *(non-expert)* says: If Lisa invests in hedge funds then she will lose all the invested money. (Ein Vermögensberater *(expert)* / Banklehrling *(non-expert)* sagt: Wenn Lisa in Hedgefonds investiert hat, dann verliert sie ihr investiertes Geld.)

7. A medical attendant of the German national football team *(expert)* / journalist *(non-expert)* says: If Miroslav Klose's lactat level is within the norm then he plays all games for the German national football team. (Der ärztliche Betreuer der Nationalmannschaft sagt *(expert)* / ein Journalist *(non-expert)* sagt: Wenn Miroslav Kloses Lactatwert im Normbereich ist, dann spielt er jedes Spiel der Fußballnationalmannschaft.)

A.1.3 Experiment 3

The relevant details of this experiment are provided in Section 3.2.1. For more, see Singmann et al. (2016): Experiment 3a, and, `https://osf.io/zcdfq/`.

In the experiment four different contents were presented in German, whether by themselves, or with additional alternatives or disablers. Both the original German version and the corresponding translations in English follow:

1. If a predator is hungry then it will search for prey. (Wenn ein Raubtier Hunger hat, dann geht es auf Suche nach Beute.)

Disablers: A predator can search for prey only when (Ein Raubtier kann nur dann auf Suche nach Beute gehen, wenn):

- it lives in the wild. (es in freier Wildbahn lebt.)
- it is not heavily injured. (es nicht zu schwer verletzt ist.)
- it does not have to protect its cub. (es nicht gerade Junge verteidigen muss.)

Alternatives: A predator can search for prey also when (Ein Raubtier kann auch auf Suche nach Beute gehen, wenn):

- it has to supply food to its cub. (es Junge versorgen muss.)
- it wants to stock up. (es einen Vorrat anlegen will.)
- it just wants to hunt. (es gerade Lust am Jagen hat.)

2. If a balloon is pricked with a needle then it will quickly lose air. (Wenn ein Ballon mit einer Nadel gestochen wurde, dann verliert er schnell an Luft.)

   Disablers: A balloon can quickly lose air only when (Ein Ballon kann nur dann schnell an Luft verlieren, wenn):

   - it is not made out of tear-resistant material. (er nicht aus reißfestem Material besteht.)
   - it is not already empty to begin with. (er nicht schon vorher leer ist.)
   - it is not filled with something else than air. (er nicht mit etwas anderem als Luft gefüllt ist.)

   Alternatives: A balloon can quickly lose air also when (Ein Ballon kann auch dann schnell an Luft verlieren, wenn):

   - it is pricked with something else than a needle. (er mit etwas anderem als einer Nadel gestochen wird.)
   - it is not knotted and it is released. (er nicht zugeknotet ist und losgelassen wird.)
   - it is inflated too much and bursts. (er zu fest aufgeblasen ist und platzt.)

3. If a girl has sexual intercourse with her boyfriend then she will be pregnant. (Wenn ein Mädchen Geschlechtsverkehr mit ihrem Freund vollzogen hat, dann ist es schwanger.)

   Disablers: A girl can be pregnant only when (Ein Mädchen kann nur dann schwanger sein, wenn):

- she doesn't use contraceptives. (es keine Verhütungsmittel benutzt hat.)
- she is not infertile. (es nicht unfruchtbar ist.)
- her partner is not infertile. (sein Partner nicht unfruchtbar ist.)

Alternatives: A girl can be pregnant also when (Ein Mädchen kann auch dann schwanger sein, wenn):

- she is impregnated by another man. (es von einem anderen Mann geschwängert wurde.)
- she is artificially fertilized. (es künstlich befruchtet wurde.)
- she carries a child as a surrogate mother. (es als Leihmutter ein fremdes Kind austrägt)

4. If a person drinks a lot of coke then the person will gain weight. (Wenn eine Person viel Cola trinkt, dann nimmt sie an Gewicht zu.)

   Disablers: A person can gain weight only when (Eine Person kann nur dann an Gewicht zunehmen, wenn):

   - they have a genetic predisposition to it. (sie eine genetische Veranlagung dazu hat.)
   - they do not move enough. (sie sich nicht genug bewegt.)
   - the coke is not sugar-free. (die Cola nicht zuckerfrei ist.)

   Alternatives: A person can gain weight also when (Eine Person kann auch dann an Gewicht zunehmen, wenn):

   - they eat a lot. (sie viel isst.)
   - they are still growing. (sie sich noch in Wachstum befindet.)
   - they drink a lot of other drinks with sugar. (sie viel andere zuckerhaltige Getränke trinkt.)

A.1.4 Experiment 4

The relevant details of this experiment are provided in Section 3.2.1. For more, see Singmann et al. (2016): Experiment 3b, and, `https://osf.io/zcdfq/`.

In the experiment four different contents were presented in German, whether by themselves, or with additional alternatives or disablers. Both the original German version and the corresponding translations in English follow:

1. If a predator is hungry then it will search for prey. (Wenn ein Raubtier Hunger hat, dann geht es auf Suche nach Beute.)

   Disablers: A predator can search for prey only when (Ein Raubtier kann nur dann auf Suche nach Beute gehen, wenn):

   - it is physically capable. (es körperlich dazu in der Lage ist.)
   - it lives in the wild. (es in freier Wildbahn lebt.)
   - it does not have any more prey. (es keine Beute mehr hat.)

   Alternatives: A predator can search for prey also when (Ein Raubtier kann auch auf Suche nach Beute gehen, wenn):

   - it has to supply food to its cub. (es Junge versorgen muss.)
   - it wants to stock up. (es einen Vorrat anlegen will.)
   - it is driven by its hunting instinct. (es der Jagdinstinkt dazu treibt.)

2. If a balloon is pricked with a needle then it will quickly lose air. (Wenn ein Ballon mit einer Nadel gestochen wurde, dann verliert er schnell an Luft.)

   Disablers: A balloon can quickly lose air only when (Ein Ballon kann nur dann schnell an Luft verlieren, wenn):

   - it was inflated before. (er vorher aufgeblasen war.)
   - the balloon bursts at the puncture site, e.g. there is no adhesive tape over it. (die Ballonhaut an der Stichstelle platzt, z.B. kein Tesa darüber)
   - it is not filled with something else than air. (er nicht mit etwas anderem als Luft gefüllt ist.)

   Alternatives: A balloon can quickly lose air also when (Ein Ballon kann auch dann schnell an Luft verlieren, wenn):

   - it has a hole because of another reason, e.g. a sharp stone. (er aus einem anderen Grund, z.B. spitzer Stein, ein Loch hat.)
   - it is not knotted properly. (er nicht richtig zugemacht ist.)
   - it comes into contact with a hot object and bursts. (er mit einem heißen Gegenstand in Berührung kommt und platzt.)

3. If a girl has sexual intercourse with her boyfriend then she will be pregnant. (Wenn ein Mädchen Geschlechtsverkehr mit ihrem Freund vollzogen hat, dann ist es schwanger.)

Disablers: A girl can be pregnant only when (Ein Mädchen kann nur dann schwanger sein, wenn):

- she doesn't use contraceptives. (es keine Verhütungsmittel benutzt hat.)
- neither of them is not infertile. (keiner von beiden unfruchtbar ist.)
- she is not in an infertile phase of her cycle. (sie sich nicht in einer unfruchtbaren Phase ihre Zyklus befand.)

Alternatives: A girl can be pregnant also when (Ein Mädchen kann auch dann schwanger sein, wenn):

- she is impregnated by another man. (es von einem anderen Mann geschwängert wurde.)
- she is artificially fertilized. (es künstlich befruchtet wurde.)
- she carries a child as a surrogate mother. (es als Leihmutter ein fremdes Kind austrägt)

4. If a person drinks a lot of coke then the person will gain weight. (Wenn eine Person viel Cola trinkt, dann nimmt sie an Gewicht zu.)

   Disablers: A person can gain weight only when (Eine Person kann nur dann an Gewicht zunehmen, wenn):

   - their metabolism allows it. (es ihr Stoffwechsel erlaubt.)
   - they don't do sports to compensate for drinking cola. (sie als Ausgleich zum Colatrinken keinen Sport treibt.)
   - they don't drink only diet cola. (sie nicht ausschließlich Diätcola trinkt.)

   Alternatives: A person can gain weight also when (Eine Person kann auch dann an Gewicht zunehmen, wenn):

   - they eat a lot. (sie viel isst.)
   - they have metabolism issues. (sie Stoffwechselprobleme hat.)
   - they are barely moving. (sie sich kaum bewegt.)

A.1.5 Experiment 5

The relevant details of this experiment are provided in Section 3.2.2. For more, see Singmann and Klauer (2011): Experiment 1.

In the experiment there were four different contents for the prological conditionals. The counterlogical conditionals have the same contents, but the antecedent and the consequent are reversed.

Prological conditionals:

1. If a person's food went down the wrong way, then the person has to cough.
2. If a person fell into a swimming pool, then the person is wet.
3. If water was poured on a campfire, then the fire goes out.
4. If a person ate a lot of salt, then the person is thirsty.

Counterlogical conditionals:

1. If a person has to cough, then the person's food went down the wrong way.
2. If a person is wet, then the person fell into a swimming pool.
3. If a campfire goes out, then water was poured on this campfire.
4. If a person is thirsty, then the person ate a lot of salt.

A.1.6 Experiment 6

The relevant details of this experiment are provided in Section 3.2.2. For more, see Singmann and Klauer (2011): Experiment 2.

In the experiment nine different contents were presented, for three prological, three neutral, and, three counterlogical conditionals.

Prological conditionals:

1. If a person fell into a swimming pool, then the person is wet.
2. If a dog has fleas, then it will scratch itself from time to time.
3. If you prick a soap-bubble, then it will pop.

Neutral conditionals:

1. If a person studies hard, then the person will get a good grade in the test.
2. If a person has turned on the air conditioner, then the person feels cool.
3. If a person drinks a lot of coke, then the person will gain weight.

Counterlogical conditionals:

1. If you water a plant well, then the plant stays green.
2. If a person brushes his/her teeth, then the person will not get cavities.
3. If a girl had sexual intercourse, then she is pregnant.

A.1.7 Experiment 7

The relevant details of this experiment are provided in Section 3.2.3. For more, see Singmann et al. (2014): Probabilized conditional inference task.

In the experiment four different contents out of the following 16 were presented in German. Both the original German version and the corresponding translations in English follow:

1. If oil prices continue to rise then German petrol prices will rise. (Wenn der Ölpreis weiter steigt, dann wird der Sprit in Deutschland teurer werden.)
2. If car ownership increases then traffic congestion will get worse. (Wenn die Anzahl der Autobesitzer steigt, dann nimmt die Anzahl der Staus zu.)
3. If more people use protective sun cream then cases of skin cancer will be reduced. (Wenn mehr Menschen Sonnencreme benutzen, dann gibt es weniger Fälle von Hautkrebs.)
4. If kindergarten teachers' salaries are improved then the recruitment of kindergarten teachers will increase. (Wenn die Löhne in Kindergärten steigen, dann werden mehr Erzieher und Erzieherinnen ausgebildet werden.)
5. If jungle deforestation continues then Gorillas will become extinct. (Wenn die Abholzung der Regenwälder fortschreitet, dann werden Gorillas aussterben.)
6. If student grants are raised then university entries will increase. (Wenn der BAföG Satz erhöht wird, wird es mehr Studienbewerber geben.)
7. If the industrialized nations reduce their CO2 emissions then global warming will be reduced. (Wenn die Industrienationen den Ausstoss von CO2 reduzieren, dann wird die globale Erwärmung abgeschwächt).
8. If student fees are brought back then the number of students will drop. (Wenn Studiengebühren wieder eingeführt werden, dann wird die Anzahl der Studierenden sinken.)

9. If primary school class sizes are reduced then the national education level will improve. (Wenn die Klassengrösse in der Grundschule reduziert wird, dann wird das allgemeine Bildungsniveau ansteigen.)

10. If immigration laws are made stricter then the number of immigrants in Germany will decrease. (Wenn die Asylgesetze verscärft werden, dann wird die Anzahl der Imigranten in Deutschland abnehmen.)

11. If the cost of fruit and vegetables is subsidised then people will eat more healthily. (Wenn die Preise für Obst un Gemüse subventioniert werden, dann werden die Menschen sich gesünder ernähren.)

12. If German troops remain in Afghanistan then acts of terrorism in Germany will increase. (Wenn deutsche Soldaten in Afghanistan blieben dann wird es in Deutschland mehr terroristische Anschläge geben.)

13. If genetic research continues then a cure for any cancer will be found. (Wenn die Genforschung fortgesetzt wird, dann wird es eine Behandlung für jeden Krebs geben.)

14. If the cost of fuel increases then more people in Freiburg will use bicycles. (Wenn di Spritpreise weiter steigenm dann werden mehr Menschen in Freiburg Fahrrad fahren.)

15. If global warning continues then Hamburg will be flooded. (Wenn die globale Erwärmung weiter anhält, dann wird Hamburg überschwemmt werden.)

16. If Greece leaves the Euro then Italy will too. (Wenn Griechenland den Euro verlässt, dann wird Italien den Euro verlassen.)

A.1.8 Experiment 8

The relevant details of this experiment are provided in Section 3.2.3. For more, see Singmann et al. (2014): Deductive conditional inference task.

In the experiment two different contents were presented in German. Both the original German version and the corresponding translations in English follow:

1. If the letter is a B then the number is a 7. (Wenn der Buchstabe ein B ist, dann ist die Zahl eine 7.)

2. If the number is a 4 then the letter is an E. (Wenn die Zahl eine 4 ist, dann ist der Buchstabe ein E.)

A.2 Logistic Regression Model Theory

A.2.1 Ranking Theory

Ranking theory quantifies *disbelief* expressed by *negative ranking functions* κ.

Assume a non-empty set W containing mutually exclusive and jointly exhaustive possibilities, and an algebra $\mathcal{A}$ over W. Then κ is a *negative ranking function* for $\mathcal{A}$, iff $\kappa(A) \to \mathbb{N} \cup \{\infty\}$, such that for all $A, B \in \mathcal{A}$:

$$\kappa(W) = 0 \text{ and } \kappa(\emptyset) = \infty$$

$$\kappa(A \cup B) = \min\{\kappa(A), \kappa(B)\}$$

where $\kappa(A)$ is called the *negative rank* of A.

Conditional rank of B, given A, representing conditional beliefs:

$$\kappa(B|A) = \kappa(A \cap B) - \kappa(A)$$

$\kappa(A) = 0$ represents that A is *not* disbelieved. $\kappa(A) = n, n > 0$ represents that A is disbelieved to the n-th degree.

Positive ranking functions β expressing degrees of *belief*:

$$\beta(A) = \kappa(\bar{A})$$

such that for all $A, B \in \mathcal{A}$:

$$\beta(W) = \infty \text{ and } \beta(\emptyset) = 0$$

$$\beta(A \cup B) = \min\{\beta(A), \beta(B)\}$$

$$\beta(B|A) = \beta(\bar{A} \cup B) - \beta(\bar{A})$$

The conditional degree of belief in B, $\beta(B|A)$, represents the degree of belief in the material implication, where $\bar{A}$ represents the false antecedent case, where the material implication is satisfied trivially.

Two-sided ranking functions:

$$\tau(A) = \beta(A) - \kappa(A) = \kappa(\bar{A}) - \kappa(A)$$

$$\tau(B|A) = \beta(B|A) - \kappa(B|A) = \kappa(\bar{B}|A) - \kappa(B|A)$$

Translation from probabilities to negative ranks is possible (Spohn, 2009; Skovgaard-Olsen, 2016).

A.2.2 Logistic Regression

Probability that the dependent variable Y will take the value 1, i.e. 'True':

$$P(Y = 1|X_1, \ldots, X_n) = \frac{z}{1+z}$$

where $z = e^{b_0 + b_1 \cdot X_1 + \ldots b_n \cdot X_n}$, and b_i are indexed weights which express how much the indexed predictor X_i contributes to reducing the variance in the depending variable.

Simplification:

$$P(Y = 1|X_1, \ldots, X_n) = \frac{1}{1+\frac{1}{z}}$$

Transformations, in order to summarize the effect of X_i by a single coefficient:

Conditional Odds:

$$O_i = \frac{P(Y = 1|X_1, \ldots, X_n)}{1 - P(Y = 1|X_1, \ldots, X_n)}$$

$$O_i = z = e^{b_0 + b_1 \cdot X_1 + \ldots b_n \cdot X_n}$$

Logged Odds:

$$\ln(O_i) = b_0 + b_1 \cdot X_1 + \ldots b_n \cdot X_n$$

(parallels multiple linear regression)

A.2.3 Logistic Regression and Ranking Theory

Two-sided ranking functions are the logged odds of a proposition *not* taking the value 'true'.

$$\tau(A) = \beta(A) - \beta(\bar{A}) = \kappa(\bar{A}) - \kappa(A)$$

$$\kappa(A) \approx \log_a(P(X = 1)), \text{ where } 0 < a < 1$$

$$\tau(A) \approx \log_a(P(X = 0)) - \log_a(P(X = 1)) = \log_a\left(\frac{P(X = 0)}{P(X = 1)}\right)$$

Table 40: Skovgaard-Olsen's derived expressions for determining individuals' inference endorsements

$$\textbf{MP:}\quad P(Y=1|X=1)=\frac{1}{1+e^{-(b_0+b_1)}}\qquad \textbf{DA:}\quad P(Y=0|X=0)=\frac{1}{1+e^{b_0}}$$

$$\textbf{AC:}\quad P(X=1|Y=1)=\frac{1}{1+e^{-(b_0^*+b_1)}}\qquad \textbf{MT:}\quad P(X=0|Y=0)=\frac{1}{1+e^{b_0^*}}$$

The parallel between ranking theory and logistic regression can be made closer by considering the logged odds format of the equations in Table 40, as shown in Table 41.

Table 41: Logged odds format of Skovgaard-Olsen's derived expressions for determining individuals' inference endorsements

$$\textbf{MP:}\quad \ln\left(\frac{P(Y=1|X=1)}{P(Y=0|X=1)}\right)=b_0+b_1\qquad \textbf{DA:}\quad \ln\left(\frac{P(Y=0|X=0)}{P(Y=1|X=0)}\right)=-b_0$$

$$\textbf{AC:}\quad \ln\left(\frac{P(X=1|Y=1)}{P(X=0|Y=1)}\right)=b_0^*+b_1\qquad \textbf{MT:}\quad \ln\left(\frac{P(X=0|Y=0)}{P(X=1|Y=0)}\right)=-b_0^*$$

The two-sided ranking function τ for conditional probability (Appendix A.2.1) can be obtained by inverting the logarithm's base in the expressions in Table 41, as shown in Table 42.

Table 42: Two-sided ranking function τ expressed through logged odds format of the derived expressions.

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

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

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

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

From Tables 41 and 42, simple inference endorsements expressions using the two-sided ranking function can be derived, as shown in Table 43.

Table 43: Expressions for determining individuals' inference endorsements through the two-sided ranking function

MP:	$\tau(C\|A) = b_0 + b_1$	**DA:**	$\tau(\bar{C}\|\bar{A}) = -b_0$
AC:	$\tau(A\|C) = b_0^* + b_1$	**MT:**	$\tau(\bar{A}\|\bar{C}) = -b_0^*$

A.3 Derivation of Endorsement Expressions

In the following, the complete derivations of the endorsement expressions for the models, as formalized by Oberauer (2006), presented in Section **??**, Mental Model Theory (Original and Directionality Variants), and, Suppositional Theory (Sequential and Exclusive Variants), are shown.

A.3.1 Mental Model Theory

Original Variant

$$\begin{aligned}
\mathbf{MP}: & \ r \cdot f \cdot a \cdot (1-e) + r \cdot f \cdot (1-a) \cdot (1-e) \\
& + r \cdot (1-f) \cdot a \cdot (1-e) + r \cdot (1-f) \cdot (1-a) \cdot (1-e) \\
& = r \cdot f \cdot (1-e) \cdot (a+1-a) + r \cdot (1-f) \cdot (1-e) \cdot (a+1-a) \\
& = r \cdot f \cdot (1-e) + r \cdot (1-f) \cdot (1-e) \\
& = r \cdot (1-e) \cdot (f+1-f) \\
& = \boxed{r \cdot (1-e)}
\end{aligned} \tag{7}$$

$$\begin{aligned}
\mathbf{AC}: & \ r \cdot f \cdot (1-a) \cdot e + r \cdot f \cdot (1-a) \cdot (1-e) \\
& + r \cdot (1-f) \cdot (1-a) \cdot e + r \cdot (1-f) \cdot (1-a) \cdot (1-e) \\
& = r \cdot (1-a) \cdot [f \cdot e + f \cdot (1-e) + (1-f) \cdot e + (1-f) \cdot (1-e)] \\
& = r \cdot (1-a) \cdot [f \cdot (e+1-e) + (1-f) \cdot (e+1-e)] \\
& = \boxed{r \cdot (1-a)}
\end{aligned} \tag{8}$$

$$\begin{aligned}\textbf{DA}: r\cdot f\cdot(1-a)\cdot e + r\cdot f\cdot(1-a)\cdot(1-e) &= r\cdot f\cdot(1-a)\cdot(e+1-e)\\ &= \boxed{r\cdot f\cdot(1-a)}\end{aligned} \tag{9}$$

$$\begin{aligned}\textbf{MT}: r\cdot f\cdot a\cdot(1-e) + r\cdot f\cdot(1-a)\cdot(1-e) &= r\cdot f\cdot(1-e)\cdot(a+1-a)\\ &= \boxed{r\cdot f\cdot(1-e)}\end{aligned} \tag{10}$$

Directionality Variant

$$\begin{aligned}\textbf{MP}: &\ r\cdot f\cdot a\cdot(1-e)\cdot d + r\cdot f\cdot a\cdot(1-e)\cdot(1-d)\\ &+ r\cdot f\cdot(1-a)\cdot(1-e)\cdot d + r\cdot f\cdot(1-a)\cdot(1-e)\cdot(1-d)\\ &+ r\cdot(1-f)\cdot a\cdot(1-e) + r\cdot(1-f)\cdot(1-a)\cdot(1-e)\cdot d\\ &+ r\cdot(1-f)\cdot(1-a)\cdot(1-e)\cdot(1-d)\\ &= r\cdot f\cdot(1-e)\cdot(a\cdot d + a - a\cdot d + d - a\cdot d + 1 - d - a + a\cdot d)\\ &+ r\cdot(1-f)\cdot(1-e)\cdot(a + d - d\cdot a + 1 - d - a + d\cdot a)\\ &= r\cdot f\cdot(1-e) + r\cdot(1-f)\cdot(1-e) = r\cdot(1-e)\cdot(f+1-f)\\ &= \boxed{r\cdot(1-e)}\end{aligned} \tag{11}$$

$$\begin{aligned}\textbf{AC}: &\ r\cdot f\cdot(1-a)\cdot e\cdot d + r\cdot f\cdot(1-a)\cdot(1-e)\cdot d\\ &+ r\cdot(1-f)\cdot(1-a)\cdot e\cdot d + r\cdot(1-f)\cdot(1-a)\cdot(1-e)\cdot d\\ &= r\cdot f\cdot(1-a)\cdot d\cdot(e+1-e) + r\cdot(1-f)\cdot(1-a)\cdot d\cdot(e+1-e)\\ &= r\cdot d\cdot(1-a)\cdot(f+1-f)\\ &= \boxed{r\cdot(1-a)\cdot d}\end{aligned} \tag{12}$$

$$\begin{aligned}\textbf{DA}: &\ r\cdot f\cdot(1-a)\cdot e\cdot d + r\cdot f\cdot(1-a)\cdot e\cdot(1-d)\\ &+ r\cdot f\cdot(1-a)\cdot(1-e)\cdot d + r\cdot f\cdot(1-a)\cdot(1-e)\cdot(1-d)\\ &= r\cdot f\cdot(1-a)\cdot e\cdot(d+1-d) + r\cdot f\cdot(1-a)\cdot(1-e)\cdot(d+1-d)\\ &= r\cdot f\cdot(1-a)\cdot e + r\cdot f\cdot(1-a)\cdot(1-e)\\ &= r\cdot f\cdot(1-a)\cdot(e+1-e)\\ &= \boxed{r\cdot f\cdot(1-a)}\end{aligned} \tag{13}$$

$$
\begin{aligned}
\mathbf{MT} : & r \cdot f \cdot a \cdot (1-e) \cdot d + r \cdot f \cdot (1-a) \cdot (1-e) \cdot d \\
& = r \cdot f \cdot (1-e) \cdot d \cdot (a + 1 - a) \\
& = \boxed{r \cdot f \cdot (1-e) \cdot d}
\end{aligned} \tag{14}
$$

A.3.2 Suppositional Theory

Sequential Variant

$$
\begin{aligned}
\mathbf{MP} : & b \cdot c \cdot i + (1-i) \cdot c + b \cdot c \cdot i \cdot (1-c) \cdot (1-i) \cdot s \\
& + b \cdot c \cdot i \cdot (1-c) \cdot (1-i) \cdot (1-s) + b \cdot c \cdot (1-i) \cdot (1-i) \cdot (1-s) \cdot (1-s^*) \\
& + b \cdot c \cdot (1-i) \cdot (1-i) \cdot (1-s) \cdot s^* + b \cdot c \cdot (1-i) \cdot (1-i) \cdot s \cdot (1-s^*) \\
& + b \cdot c \cdot (1-i) \cdot (1-i) \cdot s \cdot s^* + b \cdot c \cdot (1-i) \cdot i \cdot (1-s^*) \\
& + b \cdot c \cdot (1-i) \cdot i \cdot s^* + b \cdot (1-c) \cdot (1-i) \cdot (1-s) \\
& + b \cdot (1-c) \cdot (1-i) \cdot s + b \cdot (1-c) \cdot i \cdot (1-c) \cdot (1-s) \\
& + b \cdot (1-c) \cdot i \cdot (1-c) \cdot s + b \cdot (1-c) \cdot i \cdot c \\
& = b \cdot c \cdot i \cdot [i + c - c \cdot i + (1 - i - c + c \cdot i) \cdot (s + 1 - s) \\
& \qquad\qquad + (1-i) \cdot (s^* + 1 - s^*) + (1-c)] \\
& + b \cdot c \cdot (1-i) \cdot (1-i) \cdot [(1-s) \cdot (1 - s^* + s^*) + s \cdot (1 - s^* + s^*)] \\
& + b \cdot (1-c) \cdot [(1-i) \cdot (1 - s + s) + (i - c \cdot i) \cdot (1 - s + s)] \\
& = b \cdot c \cdot i \cdot (3 - i - c) + b \cdot c \cdot (1 - 2 \cdot i + i^2) + b \cdot (1 - c \cdot i - c + c^2 \cdot i) \\
& = 3 \cdot b \cdot c \cdot i - b \cdot c \cdot i^2 - b \cdot c^2 \cdot i + b \cdot c - 2 \cdot b \cdot c \cdot i + b \cdot c \cdot i^2 + b \\
& \qquad\qquad - b \cdot c \cdot i - b \cdot c + b \cdot c^2 \cdot i \\
& = \boxed{b}
\end{aligned} \tag{15}
$$

$$
\begin{aligned}
\mathbf{AC}: &\ b \cdot c \cdot i \cdot (i + (1-i) \cdot c) + b \cdot c \cdot i \cdot (1-c) \cdot (1-i) \cdot s \\
&+ b \cdot c \cdot i \cdot (1-c) \cdot (1-i) \cdot (1-s) + b \cdot c \cdot (1-i) \cdot (1-i) \cdot (1-s) \cdot (1-s^*) \\
&+ b \cdot c \cdot (1-i) \cdot (1-i) \cdot s \cdot (1-s^*) + bvc \cdot (1-i) \cdot (1-i) \cdot s \cdot (1-s^*) \\
&+ b \cdot c \cdot (1-i) \cdot (1-i) \cdot s \cdot s^* + b \cdot c \cdot (1-i) \cdot i \cdot (1-s^*) \\
&+ b \cdot c \cdot (1-i) \cdot i \cdot s^* \\
&= b \cdot c \cdot i[i + c - c \cdot i + (1 - i - c + c \cdot i) \cdot (s + 1 - s) + (1-i) \cdot (1 - s^* + s^*)] \\
&+ b \cdot c \cdot (1-i) \cdot (1-i) \cdot [(1-s) \cdot (1 - s^* + s^*) + s \cdot (1 - s^* + s^*)] \\
&= b \cdot c \cdot i \cdot (2-i) + b \cdot c \cdot (1 - 2 \cdot i + i^2) \\
&= \boxed{b \cdot c}
\end{aligned}
\tag{16}
$$

$$
\begin{aligned}
\mathbf{DA}: &\ b \cdot c \cdot i \cdot (i + (1-i) \cdot c) + b \cdot c \cdot i \cdot (1-c) \cdot (1-i) \cdot s \\
&+ b \cdot c \cdot i \cdot (1-c) \cdot (1-i) \cdot (1-s) + b \cdot c \cdot (1-i) \cdot (1-i) \cdot (1-s) \cdot s^* \\
&+ b \cdot c \cdot (1-i) \cdot (1-i) \cdot s \cdot s^* + b \cdot c \cdot (1-i) \cdot i \cdot s^* \\
&+ b \cdot (1-c) \cdot i \cdot (1-c) \cdot (1-s) + b \cdot (1-c) \cdot i \cdot (1-c) \cdot s \\
&+ b \cdot (1-c) \cdot i \cdot c \\
&= b \cdot c \cdot i \cdot [i + c - c \cdot i + (1 - i - c + c \cdot i) \cdot (s + 1 - s)] \\
&+ b \cdot c \cdot (1-i) \cdot s^* \cdot (1 - s - i + s \cdot i + s - s \cdot i + i) \\
&+ b \cdot (1-c) \cdot i \cdot (1 - s - c + c \cdot s + s - c \cdot s + c) \\
&= b \cdot c \cdot i + b \cdot c \cdot (1-i) \cdot s^* + b \cdot (1-c) \cdot i \\
&= b \cdot (c \cdot i + c \cdot s^* - c \cdot s^* \cdot i + i - c \cdot i) \\
&= \boxed{b \cdot (c \cdot s^* \cdot (1-i) + i)}
\end{aligned}
\tag{17}
$$

$$
\begin{aligned}
\mathbf{MT}: & \, b \cdot c \cdot i \cdot (i + (1-i) \cdot c) + b \cdot c \cdot i \cdot (1-c) \cdot (1-i) \cdot s \\
& + b \cdot c \cdot (1-i) \cdot (1-i) \cdot s \cdot (1-s^*) + b \cdot c \cdot (1-i) \cdot (1-i) \cdot s \cdot s^* \\
& + b \cdot c \cdot (1-i) \cdot i \cdot (1-s^*) + b \cdot c \cdot (1-i) \cdot i \cdot s^* \\
& + b \cdot (1-c) \cdot (1-i) \cdot s + b \cdot (1-c) \cdot i \cdot (1-c) \cdot s \\
& + b \cdot (1-c) \cdot i \cdot c \\
& = b \cdot c \cdot i \cdot [i + (1-i) \cdot c + (1-i) \cdot (1-s^*) + (1-i) \cdot s^* + (1-c)] \\
& + b \cdot s \cdot [c \cdot i \cdot (1-c) \cdot (1-i) + c \cdot (1-i) \cdot (1-i) \cdot (1-s^*) \\
& \qquad + c \cdot (1-i) \cdot (1-i) \cdot s^* + (1-c) \cdot (1-i) + (1-c) \cdot (1-c) \cdot i] \\
& = b \cdot c \cdot i \cdot [i + c - c \cdot i + (1-i) \cdot (1 - s^* + s^*) + 1 - c] \\
& + b \cdot s[c \cdot i \cdot (1 - i - c + c \cdot i) + c \cdot (1 - 2 \cdot i + i^2) \cdot (1 - s^* + s^*) \\
& \qquad\qquad\qquad + 1 - i - c + c\dot{i} + i \cdot (1 - 2 \cdot c + c^2)] \\
& = b \cdot c \cdot i \cdot (2 - c \cdot i) \\
& + b \cdot s \cdot (c \cdot i - c \cdot i^2 - c^2 \cdot i + c^2 \cdot i^2 + c - 2 \cdot c \cdot i + c \cdot i^2 + \\
& \qquad\qquad\qquad 1 - i - c + c \cdot i + i - 2 \cdot c \cdot i + c^2 \cdot i) \\
& = b \cdot c \cdot i \cdot (2 - c \cdot i) + b \cdot s \cdot (1 - 2 \cdot c \cdot i + c^2 \cdot i^2) \\
& = b \cdot [c \cdot i \cdot (2 - c \cdot i) + s \cdot (1 - 2 \cdot c \cdot i + c^2 \cdot i^2)] \\
& = b \cdot (2 \cdot c \cdot i - c^2 \cdot i^2 + s - 2 \cdot c \cdot i \cdot s + s \cdot c^2 \cdot i^2) \\
& = b \cdot [2 \cdot c \cdot i \cdot (1-s) - c^2 \cdot i^2 \cdot (1-s) + s] \\
& = \boxed{b \cdot [(1-s) \cdot (2 \cdot c \cdot i - c^2 \cdot i^2) + s]}
\end{aligned}
\tag{18}
$$

Exclusive Variant

$$
\begin{aligned}
\mathbf{AC}: & \, b \cdot m \cdot c \cdot i \cdot (i + (1-i) \cdot c) + b \cdot m \cdot c \cdot i \cdot (1-c) \cdot (1-i) \\
& + b \cdot m \cdot c \cdot (1-i) \cdot (1-i) + b \cdot m \cdot c \cdot (1-i) \cdot i \\
& = b \cdot m \cdot c \cdot i \cdot (i + c - c \cdot i + 1 - i - c + c \cdot i) \\
& + b \cdot m \cdot c \cdot (1-i) \cdot (1 - i + i) \\
& = b \cdot m \cdot c \cdot i + b \cdot m \cdot c \cdot (1-i) = b \cdot m \cdot c \cdot i + b \cdot m \cdot c - b \cdot m \cdot c \cdot i \\
& = \boxed{b \cdot m \cdot c}
\end{aligned}
\tag{19}
$$

$$
\begin{aligned}
\mathbf{MP} : & \, b \cdot m \cdot c \cdot i \cdot (i + (1-i) \cdot c) + b \cdot m \cdot c \cdot i \cdot (1-c) \cdot (1-i) \\
& + b \cdot m \cdot c \cdot (1-i) \cdot (1-i) + b \cdot m \cdot c \cdot (1-i) \cdot i \\
& + b \cdot m \cdot (1-c) \cdot (1-i) + b \cdot m \cdot (1-c) \cdot i \cdot (1-c) \\
& + b \cdot m \cdot (1-c) \cdot i \cdot c + b \cdot (1-m) \cdot (1-s) \\
& + b \cdot (1-m) \cdot s \\
& = b \cdot m \cdot c \cdot i \cdot (i + c - c \cdot i + 1 - i - c + c \cdot i) + b \cdot m \cdot c \cdot (1-i) \cdot (1 - i + i) \\
& + b \cdot m \cdot (1-c) \cdot (1 - i + i - i \cdot c + i \cdot c) \cdot b \cdot (1-m) \cdot (1 - s + s) \\
& = b \cdot m \cdot c \cdot i + \cdot b \cdot m \cdot c \cdot (1-i) + b \cdot m \cdot (1-c) + b \cdot (1-m) \\
& = b \cdot m \cdot c \cdot i + b \cdot m \cdot c - b \cdot m \cdot c \cdot i + b \cdot m - b \cdot m \cdot c + b - b \cdot m \\
& = \boxed{b}
\end{aligned}
\tag{20}
$$

$$
\begin{aligned}
\mathbf{DA} : & \, b \cdot m \cdot c \cdot i \cdot (i + (1-i) \cdot c) + b \cdot m \cdot c \cdot i \cdot (1-c) \cdot (1-i) \\
& + b \cdot m \cdot (1-c) \cdot i \cdot (1-c) + b \cdot m \cdot (1-c) \cdot i \cdot c \\
& = b \cdot m \cdot c \cdot i \cdot (i + c - c \cdot i + 1 - i - c + c \cdot i) + b \cdot m \cdot (1-c) \cdot i \cdot (1 - c + c) \\
& = b \cdot m \cdot c \cdot i + b \cdot m \cdot i \cdot (1-c) = b \cdot m \cdot c \cdot i + b \cdot m \cdot i - b \cdot m \cdot c \cdot i \\
& = \boxed{b \cdot m \cdot i}
\end{aligned}
\tag{21}
$$

$$
\begin{aligned}
\mathbf{MT} : & \, b \cdot m \cdot c \cdot i \cdot (i + (1-i) \cdot c) + b \cdot m \cdot c \cdot (1-i) \cdot i \\
& + b \cdot m \cdot (1-c) \cdot i \cdot c + b \cdot (1-m) \cdot s \\
& = b \cdot m \cdot c \cdot i \cdot (i + c - c \cdot i + 1 - i + 1 - c) + b \cdot (1-m) \cdot s \\
& = b \cdot m \cdot c \cdot i \cdot (2 - c \cdot i) + b \cdot (1-m) \cdot s \\
& = \boxed{b \cdot [m \cdot c \cdot i(2 - c \cdot i) + (1-m) \cdot s]}
\end{aligned}
\tag{22}
$$

A.4 ϵ-MMT and Oaksford-Chater

Table 44: Representation of Oaksford-Chater Probabilistic Model parameters using ϵ-MMT parameters

Oaks.	Probability	ϵ-MMT
a	P(X)	$p_3 + p_4$
b	P(Y)	$p_2 + p_4$
ϵ	P(¬X|Y)	$\frac{p_3}{p_3+p_4}$

MP $$1-\epsilon = 1 - \frac{p_3}{p_3+p_4} = \frac{\cancel{p_3} + p_4 - \cancel{p_3}}{p_3+p_4} = \frac{p_4}{p_3+p_4}$$

DA $$\frac{1-b-a\cdot\epsilon}{1-a} = \frac{1-(p_2+p_4)-\cancel{(p_3+p_4)}\cdot\frac{p_3}{\cancel{p_3+p_4}}}{1-(p_3+p_4)}$$

$$= \frac{1-p_2-p_4-p_3}{1-p_3-p_4} = \frac{p_1}{p_1+p_2}$$

AC $$\frac{a\cdot(1-\epsilon)}{b} = \frac{(p_3+p_4)\cdot(1-\frac{p_3}{p_3+p_4})}{p_2+p_4})$$

$$= \frac{\cancel{(p_3+p_4)}\cdot\frac{\cancel{p_3}+p_4-\cancel{p_3}}{\cancel{p_3+p_4}}}{p_2+p_4} = \frac{p_4}{p_2+p_4}$$

MT $$\frac{1-b-a\cdot\epsilon}{1-b} = \frac{1-(p_2+p_4)-\cancel{(p_3+p_4)}\cdot\frac{p_3}{\cancel{p_3+p_4}}}{1-(p_2+p_4)}$$

$$= \frac{1-p_2-p_4-p_3}{1-p_2-p_4} = \frac{p_1}{p_1+p_3}$$

P(Y|¬X) $$\frac{b-a\cdot(1-\epsilon)}{1-a} = \frac{p_2+p_4-(p_3+p_4)\cdot(1-\frac{p_3}{p_3+p_4})}{1-(p_3+p_4)}$$

$$= \frac{p_2+p_4-\cancel{(p_3+p_4)\cdot(\frac{p_3+p_4-p_3}{p_3+p_4})}}{1-p_3-p_4} = \frac{p_2+\cancel{p_4}-\cancel{p_4}}{1-p_3-p_4} = \frac{p_2}{p_1+p_2}$$

P(¬X∨Y) $$1-a\cdot\epsilon = 1-\cancel{(p_3+p_4)}\cdot\frac{p_3}{\cancel{p_3+p_4}} = 1-p_3 = p_1+p_2+p_4$$

$$\mathbf{P(X \wedge Y)} \qquad a \cdot (1 - \epsilon) = (p_3 + p_4) \cdot (1 = \frac{p_3}{p_3 + p_4}) = \cancel{(p_3 + p_4)} \cdot (\frac{\cancel{p_3} + p_4 - \cancel{p_3}}{\cancel{p_3 + p_4}}) = p_4$$

A.5 Main Analysis Tables

In the following, the tables showing all of the analysis results are presented. In Chapter 6 and 7 only the relevant parameter value analysis results were displayed and discussed.

Table 46: Mean percentages of the individuals' parameter values in experiment 1. ('Red' - Reduced Inference, 'Cond' - Conditional, 'Bicond' - Biconditional, 'Form' - Conditional Presentation Form, 'D' - Disablers, 'A' - Alternatives, 'F' - Few, 'M' - Many, 'Par' - Parameter)

		Form					**Form**		
D/A	**Par**	**Red**	**Cond**	**Bicond**	**D/A**	**Par**	**Red**	**Cond**	**Bicond**
F/F	p_1	40.07	35.42	44.85	M/F	p_1	32.58	37.49	37.53
	p_2	5.10	3.32	1.41		p_2	0.50	1.54	1.28
	p_3	3.42	1.25	2.21		p_3	46.70	8.27	18.36
	p_4	51.39	60.01	51.53		p_4	20.22	52.70	42.82
F/M	p_1	50.22	46.10	41.70	M/M	p_1	32.92	31.21	44.36
	p_2	15.41	9.49	4.47		p_2	19.24	16.23	4.32
	p_3	4.11	1.79	1.52		p_3	18.15	6.85	3.94
	p_4	30.28	42.61	52.30		p_4	29.69	45.72	47.38

Table 47: Analysis of the change in parameter values, within participants in experiment 1, for different conditional presentation forms, for each combination of varying amounts of disablers and alternatives. Significant p-values are marked in bold. ('Red' - Reduced Inference, 'Cond' - Conditional, 'Bicond' - Biconditional, 'Form' - Conditional Presentation Form, '1' and '2' - Two different tasks, 'Mean' - Mean of the differences between parameter values expressed as a percentage, 'D' - Disablers, 'A' - Alternatives, 'F' - Few, 'M' - Many, 'Par' - Parameter)

		Form 1 - Form 2					
D/A	**Par**	**Red - Cond**		**Red - Bicond**		**Cond - Bicond**	
		Mean	**p-value**	**Mean**	**p-value**	**Mean**	**p-value**
F/F	p_1	4.66	.433	-4.77	.147	-9.43	**.020**
	p_2	1.78	**.034**	3.69	$<$ **.001**	1.91	**.006**
	p_3	2.17	$<$ **.001**	1.21	.063	-0.96	.445
	p_4	-8.61	.075	-0.14	.891	8.47	.060
F/M	p_1	4.11	.891	8.51	.112	4.40	.422
	p_2	5.91	**.023**	10.94	$<$ **.001**	5.02	**.004**
	p_3	2.32	**.036**	2.59	**.009**	0.27	.281
	p_4	-12.33	.117	-22.03	$<$ **.001**	-9.69	.327
M/F	p_1	-4.91	.217	-4.95	.308	-0.04	.724
	p_2	-1.04	.638	-0.78	.078	0.26	.610
	p_3	38.43	$<$ **.001**	28.34	$<$ **.001**	-10.09	**.044**
	p_4	-32.47	$<$ **.001**	-22.60	**.001**	9.87	.117
M/M	p_1	1.71	.624	-11.44	**.023**	-13.15	.052
	p_2	3.01	.290	14.92	$<$ **.001**	11.91	$<$ **.001**
	p_3	11.30	$<$ **.001**	14.22	$<$ **.001**	2.92	**.005**
	p_4	-16.03	$<$ **.001**	-17.69	**.004**	-1.66	.570

Table 48: Analysis of the change in parameter values, within participants in experiment 1, for varying amounts of disablers and alternatives, for each conditional presentation form. Significant p-values are marked in bold. ('D' - Disablers, 'A' - Alternatives, 'F' - Few, 'M' - Many, '1' and '2' - Two different tasks, 'Mean' - Mean of the differences between parameter values expressed as a percentage, 'Form' - Conditional Presentation Form, 'Red' - Reduced Inference, 'Cond' - Conditional, 'Bicond' - Biconditional, 'Par' - Parameter)

		D/A 1 - D/A 2					
Form	**Par**	**F/F - F/M**		**F/F - M/F**		**F/F - M/M**	
		Mean	**p-value**	**Mean**	**p-value**	**Mean**	**p-value**
Red	p_1	-10.14	.147	7.49	.126	7.15	.183
	p_2	-10.31	**.001**	4.60	< **.001**	-14.14	< **.001**
	p_3	-0.69	**.012**	-43.28	< **.001**	-14.74	< **.001**
	p_4	21.12	**.002**	31.17	< **.001**	21.70	**.001**
Cond	p_1	-10.68	.092	-2.08	.624	4.21	.681
	p_2	-6.18	**.018**	1.78	**.001**	-12.91	< **.001**
	p_3	-0.54	.289	-7.02	.069	-5.60	< **.001**
	p_4	17.40	**.014**	7.31	.224	14.29	**.046**
Bicond	p_1	3.15	.481	7.31	.240	0.49	.210
	p_2	-3.06	.754	0.13	.217	-2.91	.264
	p_3	0.69	.673	-16.15	< **.001**	-1.73	.441
	p_4	-0.77	.544	8.71	.100	4.16	.256
		F/M - M/F		**F/M - M/M**		**M/F - M/M**	
Red	p_1	17.63	**.017**	17.30	**.004**	-0.34	.984
	p_2	14.91	< **.001**	-3.83	.100	-18.74	< **.001**
	p_3	-42.59	< **.001**	-14.05	< **.001**	28.55	< **.001**
	p_4	10.05	**.030**	0.59	.829	-9.46	**.014**
Cond	p_1	8.61	.153	14.89	**.020**	6.29	.281
	p_2	7.95	< **.001**	-6.73	**.004**	-14.69	< **.001**
	p_3	-6.48	**.018**	-5.07	< **.001**	1.41	.308
	p_4	-10.09	**.042**	-3.11	.570	6.98	.389
Bicond	p_1	4.17	.290	-2.66	.367	-6.83	**.036**
	p_2	3.19	.055	0.15	.272	-3.04	**.011**
	p_3	-16.84	**.005**	-2.42	.704	14.42	**.003**
	p_4	9.48	**.028**	4.93	.769	-4.55	.544

Table 49: Mean percentages of the individuals' parameter values in experiment 2. Analysis of parameters for different conditional presentation forms, for each speaker expertise case. Significant p-values are marked in bold. ('Red' - Reduced Inference, 'Cond' - Conditional, 'Form' - Conditional Presentation Form, 'Speak' - Speaker Expertise, 'Non' - Non-Expert, 'Exp' - Expert, '1' and '2' - Two different tasks, 'Mean' - Mean of the differences between parameter values expressed as a percentage, 'Par' - Parameter)

		Form	
Speak	**Par**	**Red**	**Cond**
Non	p_1	34.71	36.17
	p_2	16.10	11.04
	p_3	17.60	11.40
	p_4	31.49	41.18
Exp	p_1	34.42	38.77
	p_2	16.49	10.01
	p_3	17.72	9.04
	p_4	31.20	42.11

		Form 1 - Form 2	
		Red - Cond	
Speak	**Par**	**Mean**	**p-value**
Non	p_1	-1.46	.286
	p_2	5.06	$<$ **.001**
	p_3	6.19	$<$ **.001**
	p_4	-9.69	$<$ **.001**
Exp	p_1	-4.34	.063
	p_2	6.48	$<$ **.001**
	p_3	8.68	$<$ **.001**
	p_4	-10.91	$<$ **.001**

Table 50: Analysis of the change in parameter values, within participants in experiment 2, for different speaker expertise cases, for each conditional presentation form. Significant p-values are marked in bold ('Speak' - Speaker Expertise, 'Non' - Non-Expert, 'Exp' - Expert, '1' and '2' - Two different tasks, 'Mean' - Mean of the differences between parameter values expressed as a percentage, 'Red' - Reduced Inference, 'Cond' - Conditional, 'Form' - Conditional Presentation Form, 'Par' - Parameter)

		Speak 1 - Speak 2	
		Non - Exp	
Form	**Par**	**Mean**	**p-value**
Cond	p_1	-2.60	.178
	p_2	1.03	.123
	p_3	2.36	**.004**
	p_4	-0.93	.628

Table 51: Mean percentages of the individuals' parameter values in experiment 3, for different conditional presentation forms, for each combination of varying amount of disablers and alternatives, and, each suppression effect condition. ('Red' - Reduced Inference, 'Cond' - Conditional, 'Form' - Conditional Presentation Form, 'D'/'Dis' - Disablers, 'A'/'Alt' - Alternatives, 'F' - Few, 'M' - Many, 'Supp' - Suppression Effect Condition, 'Base' - Baseline, 'Par' - Parameter)

			Form					**Form**	
Supp	**D/A**	**Par**	**Red**	**Cond**	**Supp**	**D/A**	**Par**	**Red**	**Cond**
Base	F/F	p_1	35.18	38.60	Dis	F/F	p_1	26.25	32.16
		p_2	5.24	4.68			p_2	6.17	5.21
		p_3	4.09	2.49			p_3	8.58	7.50
		p_4	55.50	54.23			p_4	59.01	55.12
	F/M	p_1	33.16	40.92		F/M	p_1	41.36	41.48
		p_2	11.81	8.12			p_2	13.67	10.12
		p_3	3.16	2.65			p_3	5.67	5.04
		p_4	51.87	48.31			p_4	39.30	43.36
	M/F	p_1	32.27	37.38		M/F	p_1	25.13	35.34
		p_2	2.31	1.90			p_2	2.96	4.18
		p_3	46.56	20.69			p_3	54.42	40.62
		p_4	18.86	40.03			p_4	17.49	19.86
	M/M	p_1	28.48	35.35		M/M	p_1	35.91	34.02
		p_2	24.79	15.16			p_2	19.78	18.78
		p_3	19.18	10.43			p_3	16.09	13.44
		p_4	27.54	39.06			p_4	28.22	33.76
Alt	F/F	p_1	35.44	27.35	Alt	M/F	p_1	22.13	25.55
		p_2	15.58	13.64			p_2	8.47	11.92
		p_3	5.31	4.07			p_3	43.38	23.99
		p_4	43.67	54.95			p_4	26.02	38.54
	F/M	p_1	44.96	36.04		M/M	p_1	32.29	30.69
		p_2	25.01	22.19			p_2	23.69	28.37
		p_3	4.44	2.99			p_3	17.25	9.66
		p_4	25.59	38.77			p_4	26.77	31.28

Table 52: Analysis of the change in parameter values, within participants in experiment 3, for different conditional presentation forms, for each combination of varying amount of disablers and alternatives, and, each suppression effect condition. Significant p-values are marked in bold. ('Red' - Reduced Inference, 'Cond' - Conditional, 'Form' - Conditional Presentation Form, '1' and '2' - Two different tasks, 'Mean' - Mean of the differences between parameter values expressed as a percentage, 'D'/'Dis' - Disablers, 'A'/'Alt' - Alternatives, 'F' - Few, 'M' - Many, 'Supp' - Suppression Effect Condition, 'Base' - Baseline, 'Par' - Parameter)

			Form 1 - Form 2					**Form 1 - Form 2**	
Supp	**D/A**	**Par**	**Red - Cond**		**Supp**	**D/A**	**Par**	**Red - Cond**	
			Mean	**p-value**				**Mean**	**p-value**
Base	F/F	p_1	-3.43	.485	Dis	F/F	p_1	-5.92	.581
		p_2	0.56	.248			p_2	0.95	.374
		p_3	1.59	.144			p_3	1.08	.548
		p_4	1.27	.770			p_4	3.88	.517
	F/M	p_1	-7.77	.280		F/M	p_1	-0.12	.885
		p_2	3.68	.316			p_2	3.55	.442
		p_3	0.52	.269			p_3	0.63	.648
		p_4	3.56	.469			p_4	-4.06	.501
	M/F	p_1	-5.11	.280		M/F	p_1	-10.22	.203
		p_2	0.41	.990			p_2	-1.21	.737
		p_3	25.87	$<$ **.001**			p_3	13.80	**.034**
		p_4	-21.17	**.002**			p_4	-2.37	.981
	M/M	p_1	-6.87	.341		M/M	p_1	1.88	.648
		p_2	9.64	**.026**			p_2	1.00	.904
		p_3	8.76	**.012**			p_3	2.66	.486
		p_4	-11.53	**.034**			p_4	-5.54	.337
Alt	F/F	p_1	8.09	.153	Alt	M/F	p_1	-3.42	.391
		p_2	1.94	.189			p_2	-3.45	**.013**
		p_3	1.25	.110			p_3	19.39	**.009**
		p_4	-11.27	.072			p_4	-12.52	**.015**
	F/M	p_1	8.92	.253		M/M	p_1	1.60	.819
		p_2	2.82	.493			p_2	-4.68	.219
		p_3	1.45	.153			p_3	7.59	**.011**
		p_4	-13.19	.059			p_4	-4.51	.179

Table 53: Analysis of the change in parameter values, within participants in experiment 3, for varying amounts of disablers and alternatives, for each conditional presentation form, and, each suppression effect condition (1/2). Significant p-values are marked in bold. ('D'/'Dis' - Disablers, 'A'/'Alt' - Alternatives, 'F' - Few, 'M' - Many, '1' and '2' - Two different tasks, 'Mean' - Mean of the differences between parameter values expressed as a percentage, 'Form' - Conditional Presentation Form, 'Red' - Reduced Inference, 'Cond' - Conditional, 'Supp' - Suppression Effect Condition, 'Base' - Baseline, 'Par' - Parameter)

			D/A 1 - D/A 2					
			F/F - F/M		**F/F - M/F**		**F/F - M/M**	
Supp	**Form**	**Par**	**Mean**	**p-value**	**Mean**	**p-value**	**Mean**	**p-value**
Base	Red	p_1	2.02	.770	2.90	.354	6.69	.159
		p_2	-6.57	.066	2.93	**.025**	-19.55	< **.001**
		p_3	0.93	.485	-42.47	< **.001**	-15.10	< **.001**
		p_4	3.63	.585	36.64	< **.001**	27.96	< **.001**
	Cond	p_1	-2.32	.620	1.22	.732	3.25	.191
		p_2	-3.45	.395	2.78	.086	-10.48	< **.001**
		p_3	-0.15	.469	-18.20	< **.001**	-7.93	< **.001**
		p_4	5.92	.517	14.20	**.043**	15.17	**.010**
Dis	Red	p_1	-15.11	.055	1.12	.810	-9.66	.124
		p_2	-7.51	**.009**	3.20	**.021**	-13.61	< **.001**
		p_3	2.92	.136	-45.83	< **.001**	-7.51	**.012**
		p_4	19.70	**.007**	41.51	< **.001**	30.79	< **.001**
	Cond	p_1	-9.32	.136	-3.18	.773	-1.86	.923
		p_2	-4.91	**.013**	1.04	.084	-13.57	< **.001**
		p_3	2.46	.113	-33.11	< **.001**	-5.93	**.010**
		p_4	11.76	.456	35.26	< **.001**	21.37	**.007**
Alt	Red	p_1	-9.52	.189	13.31	**.013**	3.15	.668
		p_2	-9.44	**.006**	7.10	**.006**	-8.11	**.019**
		p_3	0.87	.391	-38.06	< **.001**	-11.94	< **.001**
		p_4	18.09	**.002**	17.65	**.008**	16.90	**.018**
	Cond	p_1	-8.69	.063	1.80	.627	-3.34	.304
		p_2	-8.56	**.004**	1.72	.549	-14.73	< **.001**
		p_3	1.08	.331	-19.93	< **.001**	-5.60	**.005**
		p_4	16.17	**.010**	16.40	**.004**	23.67	< **.001**

Table 54: Analysis of the change in parameter values, within participants in experiment 3, for varying amounts of disablers and alternatives, for each conditional presentation form, and, each suppression effect condition (2/2). Significant p-values are marked in bold. ('D'/'Dis' - Disablers, 'A'/'Alt' - Alternatives, 'F' - Few, 'M' - Many, '1' and '2' - Two different tasks, 'Mean' - Mean of the differences between parameter values expressed as a percentage, 'Form' - Conditional Presentation Form, 'Red' - Reduced Inference, 'Cond' - Conditional, 'Supp' - Suppression Effect Condition, 'Base' - Baseline, 'Par' - Parameter)

			D/A 1 - D/A 2					
Supp	**Form**	**Par**	**F/M - M/F**		**F/M - M/M**		**M/F - M/M**	
			Mean	**p-value**	**Mean**	**p-value**	**Mean**	**p-value**
Base	Red	p_1	0.89	.929	4.67	.809	3.79	.603
		p_2	9.50	**.005**	-12.99	**.013**	-22.49	< **.001**
		p_3	-43.40	< **.001**	-16.02	< **.001**	27.38	< **.001**
		p_4	33.01	< **.001**	24.33	**.005**	-8.68	**.036**
	Cond	p_1	3.54	.849	5.57	.551	2.03	.869
		p_2	6.23	**.004**	-7.03	**.004**	-13.26	< **.001**
		p_3	-18.05	< **.001**	-7.78	**.001**	10.27	.381
		p_4	8.28	.151	9.24	.151	0.97	.949
Dis	Red	p_1	16.23	**.049**	5.45	.810	-10.78	**.034**
		p_2	10.71	< **.001**	-6.11	**.044**	-16.82	< **.001**
		p_3	-48.75	< **.001**	-10.42	**.001**	38.33	< **.001**
		p_4	21.81	**.002**	11.08	.072	-10.73	**.006**
	Cond	p_1	6.14	.178	7.46	.130	1.32	.648
		p_2	5.94	**.002**	-8.67	**.003**	-14.61	< **.001**
		p_3	-35.58	< **.001**	-8.40	< **.001**	27.18	**.001**
		p_4	23.50	**.002**	9.61	.097	-13.90	**.011**
Alt	Red	p_1	22.83	**.006**	12.67	.072	-10.16	.076
		p_2	16.54	< **.001**	1.33	.475	-15.22	< **.001**
		p_3	-38.94	< **.001**	-12.81	< **.001**	26.12	< **.001**
		p_4	-0.44	.710	-1.18	.954	-0.75	.954
	Cond	p_1	10.50	.092	5.35	.253	-5.15	.346
		p_2	10.27	**.032**	-6.17	**.014**	-16.45	< **.001**
		p_3	-21.00	**.001**	-6.67	< **.001**	14.33	.059
		p_4	0.23	.819	7.50	.886	7.27	.219

Table 55: Analysis of the change in parameter values, between participants in experiment 3, for different suppression effect conditions, for each conditional presentation form, and, each combination of varying amounts of disablers and alternatives (1/2). Significant p-values are marked in bold. ('D'/'Dis' - Disablers, 'A'/'Alt' - Alternatives, 'F' - Few, 'M' - Many, '1' and '2' - Two different tasks, 'Mean' - Mean of the differences between parameter values expressed as a percentage, 'Form' - Conditional Presentation Form, 'Red' - Reduced Inference, 'Cond' - Conditional, 'Supp' - Suppression Effect Condition, 'Base' - Baseline, 'Par' - Parameter)

			Supp 1 - Supp 2					
Form	**D/A**	**Par**	**Base - Dis**		**Base - Alt**		**Dis - Alt**	
			Mean	**p-value**	**Mean**	**p-value**	**Mean**	**p-value**
Red	F/F	p_1	8.93	**.017**	-0.25	.496	-9.19	.067
		p_2	-0.93	.270	-10.33	$<$ **.001**	-9.41	**.001**
		p_3	-4.49	**.043**	-1.22	.310	3.28	.161
		p_4	-3.50	.180	11.82	**.025**	15.31	**.009**
	F/M	p_1	-8.18	.126	-11.80	**.043**	-3.60	.315
		p_2	-1.87	.111	-13.22	**.001**	-11.34	**.019**
		p_3	-2.51	.064	-1.28	.367	1.23	.105
		p_4	12.57	.093	26.29	**.003**	13.74	**.040**
	M/F	p_1	7.13	.111	10.14	.066	3.00	.245
		p_2	-0.66	.482	-6.17	**.018**	-5.51	**.032**
		p_3	-7.85	.129	3.18	.303	11.05	.086
		p_4	1.37	.391	-7.16	.222	-8.55	.127
	M/M	p_1	-7.42	.095	-3.80	.277	3.59	.206
		p_2	5.00	.175	1.10	.389	-3.91	.147
		p_3	3.10	.066	1.95	.324	-1.16	.175
		p_4	-0.68	.433	0.77	.227	1.43	.329

Table 56: Analysis of the change in parameter values, between participants in experiment 3, for different suppression effect conditions, for each conditional presentation form, and, each combination of varying amounts of disablers and alternatives (2/2). Significant p-values are marked in bold. ('D'/'Dis' - Disablers, 'A'/'Alt' - Alternatives, 'F' - Few, 'M' - Many, '1' and '2' - Two different tasks, 'Mean' - Mean of the differences between parameter values expressed as a percentage, 'Form' - Conditional Presentation Form, 'Red' - Reduced Inference, 'Cond' - Conditional, 'Supp' - Suppression Effect Condition, 'Base' - Baseline, 'Par' - Parameter)

			Supp 1 - Supp 2					
Form	**D/A**	**Par**	**Base - Dis**		**Base - Alt**		**Dis - Alt**	
			Mean	**p-value**	**Mean**	**p-value**	**Mean**	**p-value**
Cond	F/F	p_1	6.44	.108	11.23	**.009**	4.79	.289
		p_2	-0.53	.279	-8.96	**.001**	-8.42	**.005**
		p_3	-5.01	**.005**	-1.57	.283	3.43	**.020**
		p_4	-0.89	.405	-0.71	.367	0.17	.378
	F/M	p_1	-0.57	.351	4.92	.412	5.43	.315
		p_2	-2.00	**.048**	-14.07	$<$ **.001**	-12.08	**.008**
		p_3	-2.39	**.019**	-0.34	.367	2.04	**.025**
		p_4	4.94	.166	9.51	.109	4.59	.089
	M/F	p_1	2.04	.461	11.82	.098	9.78	.444
		p_2	-2.28	.475	-10.02	$<$ **.001**	-7.75	$<$ **.001**
		p_3	-19.92	**.003**	-3.27	.297	16.61	**.015**
		p_4	20.17	$<$ **.001**	1.50	.360	-18.71	**.001**
	M/M	p_1	1.32	.412	4.68	.210	3.33	.322
		p_2	-3.62	.084	-13.21	**.002**	-9.58	**.020**
		p_3	-3.01	.153	0.77	.419	3.77	.143
		p_4	5.32	.115	7.79	.128	2.50	.481

Table 57: Mean percentages of the individuals' parameter values in experiment 4, for different conditional presentation forms, for each combination of varying amount of disablers and alternatives, and, each suppression effect condition. ('Red' - Reduced Inference, 'Cond' - Conditional, 'Form' - Conditional Presentation Form, 'D'/'Dis' - Disablers, 'A'/'Alt' - Alternatives, 'F' - Few, 'M' - Many, 'Supp' - Suppression Effect Condition, 'Base' - Baseline, 'Par' - Parameter)

			Form					Form	
Supp	**D/A**	**Par**	**Red**	**Cond**	**Supp**	**D/A**	**Par**	**Red**	**Cond**
Base	F/F	p_1	33.82	43.34	Dis	F/F	p_1	33.49	39.72
		p_2	4.93	1.94			p_2	3.70	2.63
		p_3	3.40	0.91			p_3	8.48	7.75
		p_4	57.85	53.81			p_4	54.34	49.90
	F/M	p_1	51.86	44.06		F/M	p_1	44.25	38.38
		p_2	9.57	6.46			p_2	8.48	7.65
		p_3	4.83	2.73			p_3	7.79	5.75
		p_4	33.75	46.74			p_4	39.49	48.21
	M/F	p_1	46.96	35.53		M/F	p_1	35.35	29.80
		p_2	2.20	1.68			p_2	3.11	2.75
		p_3	32.40	13.64			p_3	41.05	36.03
		p_4	18.45	49.16			p_4	20.49	31.42
	M/M	p_1	24.91	33.58		M/M	p_1	28.27	28.90
		p_2	24.10	11.72			p_2	21.69	16.66
		p_3	13.26	6.82			p_3	17.94	17.27
		p_4	37.72	47.88			p_4	32.10	37.18
Alt	F/F	p_1	21.42	20.25	Alt	M/F	p_1	33.49	29.40
		p_2	23.15	17.84			p_2	9.55	11.57
		p_3	5.05	4.68			p_3	37.68	25.37
		p_4	50.38	57.23			p_4	19.28	33.67
	F/M	p_1	50.71	40.57		M/M	p_1	23.16	32.84
		p_2	21.61	15.14			p_2	31.18	26.51
		p_3	4.12	2.16			p_3	18.90	8.85
		p_4	23.56	42.13			p_4	26.76	31.81

Table 58: Analysis of the change in parameter values, within participants in experiment 4, for different conditional presentation forms, for each combination of varying amount of disablers and alternatives, and, each suppression effect condition. Significant p-values are marked in bold. ('Red' - Reduced Inference, 'Cond' - Conditional, 'Form' - Conditional Presentation Form, '1' and '2' - Two different tasks, 'Mean' - Mean of the differences between parameter values expressed as a percentage, 'D'/'Dis' - Disablers, 'A'/'Alt' - Alternatives, 'F' - Few, 'M' - Many, 'Supp' - Suppression Effect Condition, 'Base' - Baseline, 'Par' - Parameter)

			Form 1 - Form 2					**Form 1 - Form 2**	
			Red - Cond					**Red - Cond**	
Supp	**D/A**	**Par**	**Mean**	**p-value**	**Supp**	**D/A**	**Par**	**Mean**	**p-value**
Base	F/F	p_1	-9.51	.150	Dis	F/F	p_1	-6.23	.327
		p_2	2.99	**.001**			p_2	1.07	.318
		p_3	2.49	**.002**			p_3	0.73	.922
		p_4	4.04	.482			p_4	4.43	.583
	F/M	p_1	7.80	.177		F/M	p_1	5.86	.378
		p_2	3.10	.177			p_2	0.82	.217
		p_3	2.10	.058			p_3	2.04	.158
		p_4	-12.99	**.028**			p_4	-8.72	.153
	M/F	p_1	11.43	.191		M/F	p_1	5.55	.610
		p_2	0.52	.103			p_2	0.35	.799
		p_3	18.76	**.001**			p_3	5.03	.308
		p_4	-30.71	**.001**			p_4	-10.93	.096
	M/M	p_1	-8.67	.090		M/M	p_1	-0.63	.829
		p_2	12.38	**.011**			p_2	5.03	**.012**
		p_3	6.45	**.009**			p_3	0.68	.739
		p_4	-10.16	**.016**			p_4	-5.08	.131
Alt	F/F	p_1	1.17	.829	Alt	M/F	p_1	4.09	.754
		p_2	5.31	.122			p_2	-2.02	.347
		p_3	0.38	.875			p_3	12.32	**.015**
		p_4	-6.85	.318			p_4	-14.39	**.044**
	F/M	p_1	10.14	.122		M/M	p_1	-9.68	**.023**
		p_2	6.48	**.023**			p_2	4.67	.164
		p_3	1.95	.378			p_3	10.05	**.002**
		p_4	-18.57	**.010**			p_4	-5.04	.829

Table 59: Analysis of the change in parameter values, within participants in experiment 4, for varying amounts of disablers and alternatives, for each conditional presentation form, and, each suppression effect condition (1/2). Significant p-values are marked in bold. ('D'/'Dis' - Disablers, 'A'/'Alt' - Alternatives, 'F' - Few, 'M' - Many, '1' and '2' - Two different tasks, 'Mean' - Mean of the differences between parameter values expressed as a percentage, 'Form' - Conditional Presentation Form, 'Red' - Reduced Inference, 'Cond' - Conditional, 'Supp' - Suppression Effect Condition, 'Base' - Baseline, 'Par' - Parameter)

			D/A 1 - D/A 2					
			F/F - F/M		F/F - M/F		F/F - M/M	
Supp	**Form**	**Par**	**Mean**	**p-value**	**Mean**	**p-value**	**Mean**	**p-value**
Base	Red	p_1	-18.03	**.033**	-13.14	.247	8.91	.239
		p_2	-4.64	.206	2.73	**.001**	-19.18	< **.001**
		p_3	-1.43	.991	-29.00	< **.001**	-9.86	< **.001**
		p_4	24.10	**.004**	39.40	< **.001**	20.13	**.018**
	Cond	p_1	-0.72	.721	7.80	.456	9.76	.482
		p_2	-4.52	.068	0.27	.157	-9.78	< **.001**
		p_3	-1.82	.275	-12.73	**.001**	-5.91	**.001**
		p_4	7.07	.304	4.65	.510	5.93	.315
Dis	Red	p_1	-10.76	.131	-1.86	.829	5.22	.318
		p_2	-4.78	**.006**	0.59	.096	-17.99	< **.001**
		p_3	0.69	.557	-32.57	< **.001**	-9.46	**.001**
		p_4	14.85	**.034**	33.84	< **.001**	22.23	**.001**
	Cond	p_1	1.34	.814	9.92	.075	10.82	**.010**
		p_2	-5.03	**.004**	-0.12	.096	-14.03	< **.001**
		p_3	2.00	**.028**	-28.28	< **.001**	-9.52	< **.001**
		p_4	1.69	.481	18.49	**.010**	12.72	**.005**
Alt	Red	p_1	-29.29	< **.001**	-12.07	.217	-1.74	.456
		p_2	1.54	.845	13.60	**.002**	-8.03	.055
		p_3	0.94	.088	-32.63	< **.001**	-13.85	< **.001**
		p_4	26.82	< **.001**	31.10	< **.001**	23.61	< **.001**
	Cond	p_1	-20.32	**.007**	-9.15	.122	-12.59	**.002**
		p_2	2.70	.256	6.27	**.004**	-8.67	**.020**
		p_3	2.51	**.038**	-20.69	< **.001**	-4.17	**.016**
		p_4	15.10	**.040**	23.57	**.001**	25.43	< **.001**

Table 60: Analysis of the change in parameter values, within participants in experiment 4, for varying amounts of disablers and alternatives, for each conditional presentation form, and, each suppression effect condition (2/2). Significant p-values are marked in bold. ('D'/'Dis' - Disablers, 'A'/'Alt' - Alternatives, 'F' - Few, 'M' - Many, '1' and '2' - Two different tasks, 'Mean' - Mean of the differences between parameter values expressed as a percentage, 'Form' - Conditional Presentation Form, 'Red' - Reduced Inference, 'Cond' - Conditional, 'Supp' - Suppression Effect Condition, 'Base' - Baseline, 'Par' - Parameter)

			D/A 1 - D/A 2					
			F/M - M/F		**F/M - M/M**		**M/F - M/M**	
Supp	**Form**	**Par**	**Mean**	**p-value**	**Mean**	**p-value**	**Mean**	**p-value**
Base	Red	p_1	4.90	.567	26.95	**.003**	22.05	**.013**
		p_2	7.37	**.005**	-14.54	< **.001**	-21.91	< **.001**
		p_3	-27.57	< **.001**	-8.43	**.002**	19.13	< **.001**
		p_4	15.30	**.028**	-3.97	.417	-19.27	**.003**
	Cond	p_1	8.53	.191	10.48	.074	1.95	.871
		p_2	4.79	**.007**	-5.26	**.002**	-10.05	< **.001**
		p_3	-10.91	**.001**	-4.08	**.001**	6.82	.294
		p_4	-2.42	.721	-1.14	.347	1.28	.787
Dis	Red	p_1	8.89	.085	15.98	**.021**	7.08	.217
		p_2	5.37	**.002**	-13.21	< **.001**	-18.58	< **.001**
		p_3	-33.26	< **.001**	-10.15	< **.001**	23.11	< **.001**
		p_4	18.99	**.004**	7.38	.518	-11.61	**.010**
	Cond	p_1	8.58	.142	9.48	.055	0.90	.814
		p_2	4.90	**.003**	-9.00	< **.001**	-13.90	< **.001**
		p_3	-30.27	< **.001**	-11.51	< **.001**	18.76	< **.001**
		p_4	16.79	**.014**	11.03	**.044**	-5.76	.096
Alt	Red	p_1	17.22	**.019**	27.55	< **.001**	10.33	.158
		p_2	12.06	**.001**	-9.56	**.014**	-21.63	< **.001**
		p_3	-33.57	< **.001**	-14.78	< **.001**	18.78	**.007**
		p_4	4.28	.445	-3.20	.505	-7.48	.085
	Cond	p_1	11.18	.085	7.73	.281	-3.44	.357
		p_2	3.57	**.025**	-11.37	**.003**	-14.94	< **.001**
		p_3	-23.20	< **.001**	-6.68	< **.001**	16.52	**.046**
		p_4	8.47	.505	10.33	.248	1.86	.544

Table 61: Analysis of the change in parameter values, between participants in experiment 4, for different suppression effect conditions, for each conditional presentation form, and, each combination of varying amounts of disablers and alternatives (1/2). Significant p-values are marked in bold. ('D'/'Dis' - Disablers, 'A'/'Alt' - Alternatives, 'F' - Few, 'M' - Many, '1' and '2' - Two different tasks, 'Mean' - Mean of the differences between parameter values expressed as a percentage, 'Form' - Conditional Presentation Form, 'Red' - Reduced Inference, 'Cond' - Conditional, 'Supp' - Suppression Effect Condition, 'Base' - Baseline, 'Par' - Parameter)

			Supp 1 - Supp 2					
Form	**D/A**	**Par**	**Base - Dis**		**Base - Alt**		**Dis - Alt**	
			Mean	**p-value**	**Mean**	**p-value**	**Mean**	**p-value**
Red	F/F	p_1	0.33	.430	12.41	**.036**	12.07	**.010**
		p_2	1.23	.130	-18.22	$<$ **.001**	-19.45	$<$ **.001**
		p_3	-5.08	**.005**	-1.65	.104	3.43	**.042**
		p_4	3.51	.204	7.46	.087	3.96	.272
	F/M	p_1	7.60	.180	1.15	.384	-6.47	.162
		p_2	1.09	.459	-12.05	$<$ **.001**	-13.14	$<$ **.001**
		p_3	-2.97	**.003**	0.71	.361	3.67	**.001**
		p_4	-5.74	.230	10.19	.150	15.93	**.021**
	M/F	p_1	11.60	.104	13.47	**.032**	1.86	.311
		p_2	-0.91	.482	-7.35	**.001**	-6.44	**.001**
		p_3	-8.65	.150	-5.26	.339	3.37	.282
		p_4	-2.05	.248	-0.82	.323	1.21	.384
	M/M	p_1	-3.35	.172	1.75	.465	5.11	.113
		p_2	2.41	.488	-7.08	**.027**	-9.49	**.017**
		p_3	-4.68	**.042**	-5.64	.113	-0.96	.316
		p_4	5.61	.253	10.96	**.032**	5.34	.116

Table 62: Analysis of the change in parameter values, between participants in experiment 4, for different suppression effect conditions, for each conditional presentation form, and, each combination of varying amounts of disablers and alternatives (2/2). Significant p-values are marked in bold. ('D'/'Dis' - Disablers, 'A'/'Alt' - Alternatives, 'F' - Few, 'M' - Many, '1' and '2' - Two different tasks, 'Mean' - Mean of the differences between parameter values expressed as a percentage, 'Form' - Conditional Presentation Form, 'Red' - Reduced Inference, 'Cond' - Conditional, 'Supp' - Suppression Effect Condition, 'Base' - Baseline, 'Par' - Parameter)

			Supp 1 - Supp 2					
Form	**D/A**	**Par**	**Base - Dis**		**Base - Alt**		**Dis - Alt**	
			Mean	**p-value**	**Mean**	**p-value**	**Mean**	**p-value**
Cond	F/F	p_1	3.62	0.367	23.10	0.001	19.47	0.000
		p_2	-0.69	.066	-15.90	< **.001**	-15.21	< **.001**
		p_3	-6.83	< **.001**	-3.77	**.004**	3.07	**.025**
		p_4	3.90	.217	-3.40	.356	-7.33	.108
	F/M	p_1	5.68	.092	3.52	.318	-2.19	.427
		p_2	-1.20	.176	-8.67	**.002**	-7.48	**.009**
		p_3	-3.02	**.001**	0.57	.071	3.59	< **.001**
		p_4	-1.46	.329	4.59	.195	6.08	.211
	M/F	p_1	5.73	.389	6.11	.323	0.41	.494
		p_2	-1.08	.176	-9.89	< **.001**	-8.81	< **.001**
		p_3	-22.38	< **.001**	-11.74	.078	10.66	**.014**
		p_4	17.74	**.037**	15.51	.050	-2.25	.311
	M/M	p_1	4.68	.161	0.73	.418	-3.94	.362
		p_2	-4.93	**.015**	-14.80	< **.001**	-9.85	**.014**
		p_3	-10.44	< **.001**	-2.03	.127	8.42	**.001**
		p_4	10.71	**.010**	16.10	**.002**	5.38	**.040**

Table 63: Mean percentages of the individuals' parameter values in experiment 5, for prological and counterlogical conditionals, for each instruction type. ('Pro' - Prological, 'Count' - Counterlogical, 'Instr' - Instructions, 'Ded' - Deductive, 'Ind' - Inductive, 'Par'- Parameter)

		Type				**Type**	
Instr	**Par**	**Pro**	**Count**	**Instr**	**Par**	**Pro**	**Count**
Ded	p_1	37.92	41.40	Ind	p_1	53.17	46.70
	p_2	20.72	12.46		p_2	11.20	4.40
	p_3	3.70	5.96		p_3	1.79	16.04
	p_4	37.38	40.14		p_4	33.84	32.86

Table 64: Analysis of the change in parameter values, within participants in experiment 5, for prological and counterlogical conditionals, for each instruction type. Analysis for deductive and inductive instructions, for each conditional type. Significant p-values are marked in bold. ('Pro' - Prological, 'Count' - Counterlogical, 'Instr' - Instructions, 'Ded' - Deductive, 'Ind' - Inductive, '1' and '2' - Two different tasks, 'Mean' - Mean of the differences between parameter values expressed as a percentage, 'Par' - Parameter)

		Type 1 - Type 2				**Type 1 - Type 2**	
		Pro - Count				**Pro - Count**	
Instr	**Par**	**Mean**	**p-value**	**Instr**	**Par**	**Mean**	**p-value**
Ded	p_1	-3.49	.648	Ind	p_1	6.47	.288
	p_2	8.26	.452		p_2	6.80	**.001**
	p_3	-2.27	**.014**		p_3	-14.24	$<$ **.001**
	p_4	-2.76	.798		p_4	0.97	.936

		Instr 1 - Instr 2				**Instr 1 - Instr 2**	
		Ded - Ind				**Ded - Ind**	
Type	**Par**	**Mean**	**p-value**	**Type**	**Par**	**Mean**	**p-value**
Pro	p_1	-15.26	**.025**	Count	p_1	-5.30	.211
	p_2	9.52	**.042**		p_2	8.06	.426
	p_3	1.91	**.042**		p_3	-10.07	$<$ **.001**
	p_4	3.54	.327		p_4	7.27	.274

Table 65: Mean percentages of the individuals' parameter values in experiment 6, for prological, counterlogical and neutral conditionals, for each instruction type. ('Pro' - Prological, 'Count' - Counterlogical, 'Neut' - Neutral, 'Instr' - Instructions, 'Ded' - Deductive, 'Ind' - Inductive, 'Par'- Parameter)

		Type					**Type**		
Instr	**Par**	**Pro**	**Count**	**Neut**	**Instr**	**Par**	**Pro**	**Count**	**Neut**
Ded	p_1	25.27	29.59	39.14	Ind	p_1	34.42	34.40	37.17
	p_2	43.24	33.39	15.20		p_2	29.20	19.86	4.98
	p_3	1.78	4.57	3.83		p_3	2.16	13.58	17.02
	p_4	29.80	32.62	41.68		p_4	33.92	32.06	40.65

Table 66: Analysis of the change in parameter values, within participants in experiment 6, for prological, counterlogical and neutral conditionals, for each instruction type. Significant p-values are marked in bold. ('Pro' - Prological, 'Count' - Counterlogical, 'Neut' - Neutral, 'Instr' - Instructions, 'Ded' - Deductive, 'Ind' - Inductive, '1' and '2' - Two different tasks, 'Mean' - Mean of the differences between parameter values expressed as a percentage, 'Par' - Parameter)

		Type 1 - Type 2					
Instr	**Par**	**Pro - Count**		**Pro - Neut**		**Count - Neut**	
		Means	**p-value**	**Means**	**p-value**	**Means**	**p-value**
Ded	p_1	-4.31	.252	-13.87	**.011**	-9.56	.052
	p_2	9.85	**.038**	28.04	$<$ **.001**	18.19	$<$ **.001**
	p_3	-2.79	**.018**	-2.05	$<$ **.001**	0.74	.206
	p_4	-2.82	.517	-11.89	**.031**	-9.07	.134
Ind	p_1	0.01	.559	-2.76	.329	-2.77	.318
	p_2	9.34	**.034**	24.22	$<$ **.001**	14.87	$<$ **.001**
	p_3	-11.42	$<$ **.001**	-14.87	$<$ **.001**	-3.44	.672
	p_4	1.87	.541	-6.73	.088	-8.60	**.031**

Table 67: Analysis of the change in parameter values, between participants in experiment 6, for deductive and inductive instructions, for each conditional type. Significant p-values are marked in bold. ('Pro' - Prological, 'Count' - Counterlogical, 'Neut' - Neutral, 'Instr' - Instructions, 'Ded' - Deductive, 'Ind' - Inductive, '1' and '2' - Two different tasks, 'Mean' - Mean of the differences between parameter values expressed as a percentage, 'Par' - Parameter)

		Instr 1 - Instr 2				**Instr 1 - Instr 2**	
		Ded - Ind				**Ded - Ind**	
Type	**Par**	**Mean**	**p-value**	**Type**	**Par**	**Mean**	**p-value**
Pro	p_1	-9.14	**.003**	Count	p_1	-4.82	**.002**
	p_2	14.03	.192		p_2	13.53	.453
	p_3	-0.37	**.009**		p_3	-9.01	$<$ **.001**
	p_4	-4.13	**.014**		p_4	0.56	.067

		Instr 1 - Instr 2	
		Ded - Ind	
Type	**Par**	**Mean**	**p-value**
Neut	p_1	1.96	.490
	p_2	10.22	.357
	p_3	-13.18	$<$ **.001**
	p_4	1.03	.459

Table 68: Mean percentages and analysis of the individuals' parameter values in experiment 7 and 8, for inductive and deductive reasoning. Significant p-values are marked in bold. ('Ind' - Inductive, 'Ded' - Deductive, 'Mean' - Mean of the differences between parameter values expressed as a percentage, 'Par'- Parameter)

Par	**Ind**	**Ded**
p_1	21.87	31.26
p_2	7.60	25.06
p_3	32.00	1.68
p_4	38.47	42.00

Par	**Ind - Ded**	
	Mean	**p-value**
p_1	-9.41	.363
p_2	-17.46	.466
p_3	30.32	$<$ **.001**
p_4	-3.53	.289

A.6 Monty Python Witch Logic

Why is the conditional "If the woman weighs the same as a duck, then she is a witch" true, according to the 1975 British movie "Monty Python and the Holy Grail".

1. She looks like a witch.
2. She is dressed like a witch.
3. She has a wart.
4. She turned someone into a newt.
5. One burns witches.
6. One also burns wood.
7. Witches burn because they are made of wood.
8. Also bridges are made of wood.
9. However, bridges can also be made of stone → Making a bridge out of her cannot prove that she is a witch.
10. Wood floats in water.
11. Aside from bread, apples, very small rocks, cider, gravy, cherries, mud, churches and lead, also ducks float in water.
12. It follows that if she weighs the same as a duck, she is made of wood because she will float in water and if she is made of wood then she is a witch.
13. Therefore, by transitivity, if she weighs the same as a duck then she is a witch, holds.

This list of points has been summarized from the movie's scene script found on `http://www.montypython.net/scripts/HG-witchscene.php`.