# Strategies of Human Belief Revision in Deductive Reasoning

Inaugural Dissertation
In partial fulfillment of the requirements for the degree of
*Doctor rerum naturalium* (Dr. rer. nat.)

Submitted to
the faculty of Psychology and Sports Science
of the Justus Liebig University of Giessen
Germany

By

**Ann Gabriella Wolf**
From
Geleen, The Netherlands

Giessen, 2010

First Supervisor and Reviewer:

Prof. Dr. Markus Knauff (Psychology, Giessen)

Second Reviewer:

Prof. Dr. Michael R. Waldmann (Psychology, Göttingen)

## Acknowledgements

First, I would like to express my gratitude to my supervisor Prof. Dr. Markus Knauff for giving me the opportunity to explore my own research ideas and for his insightful and helpful guidance along the way. I thank Prof. Dr. Michael Waldmann for acting as second reviewer of my dissertation and Prof. Dr. Gebhard Sammer and Prof. Dr. Albrecht Beutelspacher for serving as members of my defense committee. Also, the work would not have been possible without Susann Rieger. Her help with the data collection and the constructive and pleasurable discussions were invaluable. Next, I am heartily thankful to all my colleagues for their support, fruitful and funny discussions about research in general and for keeping up the warm 'family-like' atmosphere in the group. Kai, vielen dank für die Korrektur der Zusammenfassung. I offer my further regards to all of those who have helped me in any respect during the completion of the project. A big thank you to my father and brother for always being there and making the doctoral period as comfortable as possible. To all my friends, thank you for the nice get-togethers. They have always been a wonderful distraction that made me keep in mind there is a world outside the research lab. Last but not least, my deepest thank you go to my husband Ervice for his unlimited love, support and patience and our son Noah, whose most beautiful smiles and coos helped me make the final corrections.

## Zusammenfassung

Die Forschung im Bereich des menschlichen Schlussfolgerns hat wiederholt aufgezeigt, dass Menschen es relativ einfach finden Unstimmigkeiten zu entdecken. Die immer noch offene Frage dabei ist, wie Menschen ihre Überzeugungen revidieren, um diese Unstimmigkeiten zu überwinden. Vergangene Forschung zur Revision von Überzeugungen hat hauptsächlich versucht, eine einzige Revisions-Strategie zu identifizieren. Das Hauptziel der vorliegenden Arbeit war es zu untersuchen, ob Menschen eine einzige Revisions-Strategie anwenden oder eher mehrere unterschiedliche Strategien. Sieben Experimente zur Überzeugungs-Revision aus dem Bereich der Theorien zum schlussfolgernden Denken werden berichtet. In jedem Experiment, mussten die Teilnehmer Überzeugungs-Revisionen zu modus ponens (MP) und modus tollens (MT) Inferenz Probleme durchführen. Jedes Inferenz Problem bestand aus jeweils drei widersprüchlichen Aussagen. Die Wahrheit der ersten beiden Aussagen war nicht sicher, während die letzte Aussage, die eine Inkonsistenz verursacht, eine Tatsache darstellte. Die Aufgabe der Teilnehmer war, es eine Präferenz zur Annahme der ersten oder zweiten Aussage an zu zeigen, um die Konsistenz wiederzuerlangen. Ihre Überzeugungs-Präferenz und Entscheidungszeit wurden als abhängige Variablen erfasst. Der Schwerpunkt lag auf der Untersuchung, ob Menschen ihre Überzeugungen als Funktion von mentalen Modellen oder aber anhand Inhalt und Kontext-faktoren revidieren. Dies beinhaltet einen Kontrast zwischen einem globalen Logik-basierten Schlussfolgerungs-Prozess und einem spezifischen psychologisch-basierten logisches Schlussfolgerungs-Prozess. Der Einsatz dieser beiden unterschiedlichen Inferenz Typen ermöglichte die Untersuchung einer eventuellen Beteiligung mentaler Modelle. Eine mögliche Rolle von Inhalt und Kontext Faktoren in der Revision menschlicher Überzeugungen wurde durch die Manipulation der Inhalts-Faktoren Wahrscheinlichkeit (Experimente 1, 2 und 3), Vertrautheit (Experiment 3) und der Kontext-Faktoren Aufgabeninstruktion (Experiment 4) sowie der Vertrauenswürdigkeit der Quellen (Experimente 5, 6 und 7) geprüft. Die Ergebnisse zeigten, dass zwei unterschiedliche Revisions-Strategien existieren, eine basierend auf mentalen Modellen und eine basierend auf Wahrscheinlichkeit. Darüber hinaus zeigten

sowohl die Entscheidungen bei der Überzeugungsrevision wie auch die Entscheidungszeiten, dass die beiden Strategien nicht unabhängig voneinander sind, sondern den Prozess der Überarbeitung von Überzeugungen erleichtern oder erschweren, je nachdem ob sie parallel zueinander laufen oder die Denkprozesse einer Person in verschiedene Richtungen führen. So existieren sich gegenseitig ausschließende Strategien der Überzeugungsrevision, die die Revision von Überzeugungen gleichzeitig, aber auch jeweils auf ihre eigene Weise beeinflussen.

## Abstract

Research in the field of human reasoning has shown repeatedly that people find it reasonably easy to detect inconsistencies. The question that still remains is how people revise their beliefs to undo these inconsistencies. Past research on belief revision has mainly focused on trying to identify a single belief revision strategy. The main goal of this thesis was to investigate whether people apply one single belief revision strategy or rather multiple ones. This thesis reports on seven experiments on belief revision investigated within the arena of human reasoning theories. In each experiment, participants had to make belief revision choices on modus ponens (MP) and modus tollens (MT) problems. Each inference problem consisted of sets of three inconsistent statements. The truth of the first two statements was not certain and the last statement, that caused an inconsistency, was a fact. The participants' task was to indicate belief preference in the first or second statement to gain consistency. Their belief preference and their decision time were measured. The focus laid on investigating whether people revise their beliefs as a function of mental models or content and context factors. This captures a contrast between a more global logic-based reasoning process and a more specific psychological-based reasoning process. The use of the two different inference types made it possible to test a possible involvement of mental models. A possible role of content and context in human belief revision was examined by manipulating the content factors probability (Experiments I, II, and II), familiarity (Experiment III), and the context factors task instruction (Experiment IV) and source trustworthiness (Experiments V, VI, VII). The results showed that two main belief revision strategies exist, one based on mental models and one based on probability. Furthermore, both the belief revision choices and the decision time results revealed that the two strategies are not independent of each other, but either facilitate or hamper the process of belief revision depending on whether they run parallel or guide a person's mind in different directions. Thus, there exist mutually exclusive belief revision strategies affecting the belief revision process simultaneously and each in their own way.

# TABLE OF CONTENTS

# Chapter 1

# Theoretical background

## 1.1. Belief Revision

The vast and rich amount of choices we are confronted with in our daily lives can seem overwhelming. We are placed in a social network consisting of such diverse entities such as family, work, religion, and friends. But also our mind is an interconnected network consisting of cognitions, desires, opinions, norms and values, and beliefs. Yet, somehow we find our way through this cognitive maze by means of reasoning. The current thesis considers the choices we make with regard to beliefs. A great part of reasoning consists of manipulating a set of beliefs. The operation of manipulating beliefs is a combined process of taking into account evidence from the outside world and the personal motivation to maintain some beliefs over others. Beliefs are formed over a lifetime and reside deeply in our minds. Most often we welcome a new belief in our mind with little difficulty and few consequences. However, at times we receive a piece of factual information that does not fit very well with the existing knowledge about the respective topic that comprises our current belief state. This causes an inconsistency in the mind. The belief set or knowledge base needs to be updated to regain consistency. We are faced then with the challenge which belief to discard or how to change the 'importance' or 'strength' of each belief. This process is known as *belief revision.* In general, belief revision is concerned with how our view of the world changes in the view of new evidence. The study of belief revision originates from the areas of artificial intelligence (AI) and philosophy (Alchourron, Gärdenfors, & Makinson, 1985, Gärdenfors, 1988, Harman, 1986). Theorists in artificial intelligence as well as logicians constructed several theories of belief revision, of which the most well-known are minimal change (Harman, 1986) and epistemic entrenchment (Gärdenfors, 1988, 1992). According to the principle of minimal change, the belief that will be rejected is the one that causes the most amount of change in the existing belief set. The notion of epistemic entrenchment, on the other hand, holds that beliefs differ in explanatory power or informational content. The belief with the least value (be it related to personal or world knowledge) will be the one discarded.

This dissertation, however, is concerned with the investigation of human belief revision.

Human belief revision has typically been investigated by using conditional reasoning tasks. They encompass a widely established method to study reasoning (Evans, Newstead, & Byrne, 1993; Evans & Over, 1996) and since belief revision is closely tied to the former process, these tasks in turn offer an empirical sound way to investigate belief revision. The most widely used deductive reasoning problems in belief revision are modus ponens (MP) and modus tollens (MT) inferences. MP is the inference rule of the form 'if p then q; p; q'. MT is the inference rule of the form 'if p then q; not-q; not-p'. In belief revision tasks, the conclusions for MP and MT inference problems are instead ¬q and p, respectively. Both these conclusions cause an *inconsistency* between the conclusion on the one hand and the conditional and categorical premises on the other hand. For example:

*If Chris goes to work, then he will take the car*

*Chris does not take the car*

*Chris goes to work*

The reasoner, then, is typically instructed to assume that the conclusion is true but that the truth status of the premises is uncertain and decide thereupon whether to abandon his or her belief in the conditional or the categorical premise. Notwithstanding this switch from a descriptive to an empirical approach, so far the studies of belief revision still flavoured the approaches and accounts from logic and AI. From a psychological research perspective, the topic of belief revision is still lacking an established ensemble of empirical studies and a theoretical foundation. In this regard, the study of belief revision resembles an open field that is highly fertile and is awaiting the seeds of knowledge to be planted.

Because belief revision is the resulting process from an inconsistency in reasoning, the experiments in this dissertation are placed within the context of human reasoning theories. In this dissertation, belief revision is subjected to a series of experiments in which several factors, which are thought to be indicative of belief, are controlled and placed under empirical scrutiny with the goal to derive at a model for human belief revision. Before going into detail about what has discovered so far about human belief

revision, the introduction to belief revision is better served when it starts with an outline of the origin of this type of cognitive process. Even though emphasis in this thesis is put on the psychological approach to belief revision, the domains of logic and artificial intelligence are covered first. The knowledge gathered in those domains served as the building blocks for the more empirical approach to belief revision by psychologists. After that, a chapter will be devoted to an outline of human reasoning theories, since a premium is placed on psychological theories to understand human belief revision. Finally, there will be a switch from the more formal theories of belief revision to an empirical perspective that reports on human belief revision experiment done so far and introducing the so far only existing theory on human belief revision that will ultimately set out the aim of the current thesis.

### 1.1.1. Belief Revision in Logic and Artificial Intelligence

Belief revision and knowledge are closely intertwined. Knowledge, which is the fundamental benchmark of intelligence is not static but is instead constantly adapting to the environment. For this to happen, knowledge needs to be updated in the light of new information. This is where *belief change* comes into play. Belief change has been of great interest to the Artificial Intelligence arena because it cannot know how knowledge is represented in artificial agents without knowing how an artificial knowledge set handles new information (Kern-Isberner, 2001). Artificial Intelligence (A.I.) constitutes the investigation of how to make computers' capabilities equalize or exceed human performance (Rich, 1986). However, before being recognized by those studying artificial agents, belief revision first became of interest to philosophers in the 1970s.

Classical theories on belief revision fall under one of two main camps of belief revision: The 'foundation approach' of belief revision and 'the coherence approach' of belief revision (Bochman, 2001; Doyle, 1992.; Harman, 1986b; Rott, 2000). What distinguishes these two approaches is their view on the role of justification in belief acceptance (Bochman, 2001). The basic idea behind the foundation view is that one needs to keep track of the original reasons for one's beliefs in order to justify beliefs (Bochman, 2001: Harman, 1986). The guiding factor behind belief change is the dependency among beliefs and belief change occurs when someone adopts or withdraws the reasons for a belief. An example of a foundation approach is the Truth Maintenance System (TMS) by Doyle (1979). The foundation approach has been critized because it asks for a great amount of computational effort and it is not in line with observed psychological behaviour (Gärdenfors, 1990, as cited in Doyle, 1992). The coherence approach maintains that beliefs are held as long as they are consistent and logically adhere with other beliefs (Bochman, 2001; Doyle, 1992). The principal idea behind all the belief revision theories under the coherence approach was that the change in the belief system ought to be as minimal as possible. This came to be known as the idea of *informational economy* (Rott, 2000). According to Harman (1986), minimal change in resolving an inconsistency is driven by the simplest measure possible which is (Harman, 1986, pp. 59):

*Simple Measure of Change in View*: Take the sum of the number of (explicit) new beliefs added plus the number of (explicit) old beliefs given up.

Harman explicates this with an example consisting of five inconsistent beliefs:

P, Q, if P then R, if Q then R, not-R

Using the Simple Measure would lead to disbelieving *not-R*, which entails the fastest and most obvious choice to reconcile consistency. However, not everyone notices the inconsistency and proceed with reasoning. One might infer *not-P* from *If P then R* and *not-R*. Similarly, one might deduce *not-Q* from *If Q then R* and *not-R*. This would results in the following set of beliefs:

P, Q, if P then R, if Q then R, not-R, not-P, not-Q

With such a set of beliefs, the minimal change would be to give up *P and Q*, because *not-R* cannot be abandoned without also letting go of the beliefs *not-P* and *not-Q* (Harman, 1986).

Another more well-established research programme on belief revision was accomplished by Alchourrón, Gärdenfors, and Makinson (1958; 1988) and came to be known as *the AGM approach* to belief change. They viewed minimal change as a function of assigning different degrees of belief values to the statements. They called this 'epistemic entrenchment'; the sentence or belief with the highest value is most entrenched and is retained (Gärdenfors, 1992). The language of the AGM is derived from first-order logic. The elements in a belief state are coded as formulas that are represented as sentences or propositions. All sentences together constitute the language L. Sentences or pairs of sentences can be assigned the following Boolean operators: $\neg$ $\wedge$ (negation), $\vee$ (conjunction), (disjunction), and $\rightarrow$ (implication). The Greek symbols $\Phi$, $\psi$, $\chi$ represent the variables over sentences in L. Finally, the symbols $\top$ and $\bot$ represent "truth" and "falsity. The AGM approach starts with a set of basic elements (Areces & Becher, 2001; Doyle, 1992.;Gärdenfors , 1988):

1) the belief system $K$ consisting of sets of propositions

2) a propositional language $L$ over the standard sentential connectives ( $\neg$, $\wedge \vee$, $\rightarrow$, $\leftrightarrow$)

3) a piece of new information $\Phi$

4) further individual propositions denoted as $\alpha \beta, \gamma$

From these basic elements, the AGM constructed three possible theory change:

1) Expansion, $K + \Phi$ = information is added to the belief system, irrespective of whether the expanded belief system is consistent.

2) Revision, $K * \Phi$ = a new piece of information is added to the belief system that is inconsistent with it. To maintain consistency in the belief system some of its old beliefs are deleted.

3) Contraction, $K - \Phi$ = information is eliminated from the belief system without adding new facts.

A belief set must satisfy the following *integrity constraint*:

(I) If K logically entails ψ, then ψ Є K.

This shows that the belief set contains all the sentences that are accepted in the belief state. Since the emphasis in the current dissertation lies on belief revision, belief expansion and belief contraction will not be considered further. Gärdenfors (1988) defines belief revision as follows: "A belief revision occurs when a new piece of information that is inconsistent with the present belief system (or database) is added to that system in such a way that the result is a new consistent belief system" (pp.3). He offered the following example to illustrate (pp.1):

Imagine a database containing the following pieces of information:

α: All European Swans are white

B: The bird caught in the trap is white

Γ: The bird caught in the trap comes from Sweden

δ: Sweden is part of Europe

If the database is equipped with the capability to compute logical inferences then it follows from α - δ:

ε: The bird caught in the trap is white

Gärdenfors then presents the hypothetical situation in which you find out that the bird is actually black. That means that you want to add ¬ε, the negation of ε to your database. This will consequently make your database *inconsistent*. In order to regain consistency, you need to revise it. You don't want to give up all of your beliefs, this would be an unnecessary loss of information. So you need to choose between α, B, Γ, and δ. Gärdenfors denotes that logic alone is inadequate in informing you which belief to give up; this needs to be decided by other means. However, he does say that algorithms are needed for computing revision functions. For example, a computer scientist would need to develop algorithms to compute the right revision function for any belief set in order to understand belief revision. However, to judge the success of a revision function, it is necessary to know what an appropriate revision function is. To judge the appropriateness of revision functions, Gärdenfors developed a series of *rationality postulates*. The overall idea with respect to these postulates is that the revision function K*Φ of K with respect to Φ should represent the minimal change of K that is needed to allow Φ into the belief set consistently.

Belief sets are used as models of belief states. Gärdenfors subsequently developed the following postulates for revision defined over such belief sets:

(K * 1) For any sentence Φ and any belief set K, K*Φ is a belief set

(K * 2) Φ Є K*Φ

(K * 3) $K^*\Phi \leq K+ \Phi$

(K * 4) If $\neg \Phi \notin K$, then $K+ \Phi \leq K^*\Phi$

(K * 5) $K^*\Phi = K^{\perp}$ if and only if $\vdash \neg \Phi$

(K * 6) If $\vdash \Phi \leftrightarrow \psi$, then $K^*\Phi = K^* \psi$

(K * 7) $K^*\Phi \wedge \psi \leq (K^*\Phi) + \psi$

(K * 8) If $\neg \psi \notin K^*\Phi$, then $(K^*\Phi) + \psi \leq K^*\Phi \wedge \psi$

The first postulate tells you that the output of a revision function has to be a belief set. The second postulate assures that the sentence $\Phi$ is accepted in the belief set $K^*\Phi$. When a sentence contradicts what is already in K, this is denoted as $\neg\Phi \in K$. An expansion is at least as equally large as a revision, as shown by the third postulate. However, if nothing in K contradicts $\Phi$, then $K^*\Phi$ is equal to or larger than the expansion $K+ \Phi$ as shown by the fourth postulate. The goal of a revision is to reach a new consistent belief set. That means that $K^*\Phi$ should be consistent unless $\Phi$ is logically impossible. A belief set is false if the new sentence contradicts with the already present beliefs in the set, as is shown by the fifth postulate. Finally, the sixth postulate formulises that when the sentences $\Phi$ and $\psi$ have the same content, then the revision is also the same (Gärdenfors, 1992). Belief revision, according to Gärdenfors, is not syntax- but knowledge-based, referred to as "the Irrelevance of syntax" (Elio & Pelletier, 1997). That means that sentences of different syntactical structure but of logical equivalence should result in identical revisions. Postulates 1 through six comprise the basic set of postulates. The last two postulates are composite postulates because they combine revision with expansion. They show that when $\psi$ is to be included in $K^*\Phi$ this should be done by expansion of $K^*\Phi$ whenever possible. In other words, the minimal change to K to include both $\Phi$ and $\psi$, that is $K^*\Phi \wedge \psi$, should be the same as expanding $K^*\Phi$ by $\psi$, as long as $\psi$ does not cause a contradiction in $K^*\Phi$. Then, how does the AGM explain how an agent maintains minimal change in its belief set? The AGM declares that even though an agent accepts beliefs as long as they are logically consistent, they inherit different degrees of value. Certain pieces of knowledge about the world are more

important than others to the individual (Gärdenfors, 1992). Gärdenfors captures this idea with the term "epistemic entrenchment". Epistemic entrenchment is the ordering of propositions with regard to their valuableness (Doyle, 1992). This ordering is typically expressed with the symbol ≤ (Dragoni & Giorgini, 2001).

Gärdenfors (1992) characterizes epistemic entrenchment by the following postulates:

| | | |
|---|---|---|
| (≤ 1) | If $\Phi \leq \psi$ and $\psi \leq \chi$, then $\Phi \leq \chi$ | (*transitivity*) |
| (≤ 2) | If $\Phi \vdash \psi$, then $\Phi \leq \psi$ | (*dominance*) |
| (≤ 3) | For any $\Phi$ and $\psi$, $\Phi \leq \Phi \wedge \psi$ or $\psi \leq \Phi \wedge \psi$ | (*conjunctiveness*) |
| (≤ 4) | When $K \neq K^{\perp}$, $\Phi \notin K$, iff $\Phi \leq \psi$, for all $\psi$ | (*minimality*) |
| (≤ 5) | If $\psi \leq \Phi$ for all $\psi$, then $\vdash \Phi$ | (*maximality*) |

These postulates capture the idea that a good revision can be translated into providing a good ordering of epistemic entrenchment (Gärdenfors, 1992).

Even though belief revision theory into focus originally by philosophers, it was the work performed by logicians in AI that motivated the interest of psychologists to incorporate belief revision theory into human research. It was mainly Gärdenfors's view on minimal change, the notion of epistemic entrenchment, that caught the attention of scholars in psychology. However, a major difference between artificial agents and humans is that the former do not have an a priori embedded program for belief revision but instead it needs to be implemented. On the contrary, humans do have an innate system to deal with belief revision; the challenge here is to find out how it works. A better starting point other than AI approaches to belief revision, then, is to look into well-established theories on human thinking and reasoning processes. A plethora of work in this area has been devoted to higher-order thinking that can be of valuable meaning to the investigation of human belief revision. After all, theories of belief revision developed within AI (e.g. AGM developed by Alchourron et al., 1985) are grounded on formal logic. Moreover, AI theories mainly focused on formulizing competence theories of

belief revision. Little attention has been devoted to how belief revision is actually performed in practice by artificial agents (Elio & Pelletier, 1997). After all, the initial intention to study belief revision in humans was to use the gathered knowledge to better elucidate belief revision processes in artificial agents. For this to happen, we first need a firm theoretical framework for human belief revision. In the sections 1.2 to 1-4, the development of knowledge of theory of human reasoning is set out before discussing their contribution to human belief revision.

## 1.2. Human Deductive Reasoning

Reasoning takes a prominent place in our cognitive habitat. It is of central concern to the understanding of human intelligence and knowledge structure. In essence, what happens when we reason? We take as a starting point, by way of inference, some kind of information - stored knowledge, perceptual observations, statements, beliefs, imaginations, principles, general rules - and derive at a conclusion that is novel and will guide our action (Evans et al., 1993). For example, understanding utterances as part of natural language is an ongoing process of making inferences. Whereas the study of human belief revision derives mainly from artificial intelligence, the study on human deductive reasoning originates entirely from logic. This chapter will thus start by describing the basics of logic used in reasoning. Reasoning has a long history, dating back as far as the ancient Greek. A main figure at that time occupying himself with logic was Aristotle (384 BC – 322 BC), who is considered the founder of formal logic and deductive inference. Aristotle and his contemporaries occupied themselves with the question 'what makes an argument valid'? They embarked solely on syllogistic reasoning which became the only type of reasoning of interest for several millennia (Evans et al., 1993; Garnham & Oakhill, 1994). A syllogism is constructed as follows:

*All men are mortal*
*Socrates is a man*
*Socrates is mortal*

Logic has since than been the foremost normative theory of deductive reasoning. It is a theory of how valid conclusions are inferred and is used as a measure of understanding logical validity (Evans et al., 1993).

One can define several types of reasoning. However, the main demarcation is made between inductive reasoning and deductive reasoning. Induction is the process whereby one reasons from the particular to the general, for example:

*All observed swans are white*
*Therefore, all swans are white*

We engage in deduction when we reason from the general to the particular. The main classes of deduction are *propositional*, *relational*, and *quantificational* reasoning (Johnson-Laird & Byrne, 1991). The current thesis is concerned only with the deductive propositional reasoning, and will therefore be of only further focus. Deductive reasoning is, according to logic, considered to be monotonic. Monotonic reasoning holds that a conclusion that followed from a set of premises cannot be falsified by any additional premise (Johnson-Laird & Byrne, 1991).

While the Greek and Romans were interested in the relationship between argument and human reasoning, the practitioners of modern formal logic took on a more abstract doctrine and developed formal systems that stand in little relationship to human thought. This modern formal logic differs from traditional syllogistic logic in that it bears more on mathematics. Deductive arguments build on the *predicate calculus,* the most well-established of many logical systems. To fit the current context, this part will concentrate only on *propositional calculus* which is a part of the predicate calculus (Garnham & Oakhill, 1994). Propositional calculus can be formulized by a *proof-theoretic method* and a *model-theoretic method*. According to the proof-theoretic method, the form of the sentences characterize whether the argument is valid. This method applies rules to sentences which depend on the sentences' form. According to the model-theoretic approach, an argument is described by the meaning of its sentences and the validity of the argument is determent by manipulating mental models (Manktelow, 2004). Theories belonging to these two camps will be presented in part 1.3. For now, the goal is to lay out the basic logic of propositional calculus. Propositional calculus defines well-structured

sentences as consisting of simple sentences and logical operators. These logical operators are denoted by the following symbols: ⌐, ∧ ∨, →, ↔. In common language, these operators are words that connect the linguistic propositions. Thus, the connectives in ordinary English language that correspond to the logical operators are "not", "and", "or", "if", "if….then", and "if and only if" (Garnham & Oakhill, 1994; Manktelow, 2004). Propositional calculus formulizes arguments by means of the relationship between the propositions; it is not concerned with the internal structure of sentences (Garnham & Oakhill, 1994). Deductive reasoning of this kind is called propositional reasoning. The conditional inference tasks used for the belief revision studies in this thesis are solely dealing with the word "if".

The parts that make up a major premise in propositional reasoning are called atomic propositions (Manktelow, 2004). In propositional logic, truth values are assigned to each atomic proposition of an inference involving connectives; these values can either be true or false. A truth-table lists these different values or truth conditions and in doing so depicts the validity of an argument (Evans et al., 1993; Manktelow, 2004). Below is shown the truth-table of the conditional assertion. The corresponding connective in this propositional calculus is called the *material implication* (Johnson-Laird & Byrne, 1991; Manktelow, 2004):

| p | q | if p then q |
|---|---|---|
| true | true | true |
| true | false | false |
| false | true | true |
| false | false | true |

*Table 1.* **Schematic overview of the truth table for the material implication for the conditional 'if p then q'.**

Each row depicts a possible combination of the truth values. Because the material implication consists of two propositions and each proposition can only have two different truth functions, there are four combinations possible (Johnson-Laird & Byrne, 1991). As can be seen from the table, the statement is false when p is true and q is false. Here follows an example of a conditional to clarify this: "If someone has a coffee mug, then the person drinks coffee". This means that in the case where someone has a coffee mug but does not drink coffee, the statement is false. In other cases where someone does not have a coffee mug but drinks coffee is by logic true, because it does not falsify the statement (Manktelow, 2004). Any truth table can be singly captured in a negation or an inclusive disjunction. The truth table for material implication relates to the inclusive disjunction 'not-p or q' (Johnson-Laird & Byrne, 1991). Moreover, the truth table for material implication is closely linked to the conditions of *sufficiency* and *necessity*. With sufficiency is meant that the antecedent is a sufficient condition for the consequent to occur. Necessity implies that the consequent is a necessary condition for the antecedent (Johnson-Laird & Byrne, 1991; Manktelow, 2004).

The formal properties of a conditional inference can also be described or depicted in a way very similar to the Aristotelian syllogism:

If p then q
p
———————
?

The above format consists of two premises from which a conclusion should follow. The first premise is often referred to as the *conditional* or the *major premise*. The p part of the conditional is referred to as the *antecedent* and q part as the *consequent*. The second premise represents either the p or the q or a negation of either one and is often called the *categorical* or the *minor premise*. Reasoning with conditionals results in the following four common inferences:

***Table 2***. **Schematic overview of the four conditional inferences.**

| | |
|---|---|
| *Modus Ponens (MP)* | |
| If p then q | If someone has coffee mug, then he/she drinks coffee |
| p | Peter has a coffee mug |
| _____ | |
| q | Peter drinks coffee |
| *Modus Tollens (MT)* | |
| If p then q | If someone has coffee mug, then he/she drinks coffee |
| not q | Peter does not drink coffee |
| _____ | |
| not p | Peter does not have a coffee mug |
| *Affirmation of the Consequent (AC)* | |
| If p then q | If someone has coffee mug, then he/she drinks coffee |
| Q | Peter drinks coffee |
| _____ | |
| P | Peter has a coffee mug |
| *Denial of the Antecedent (DA)* | |
| If p then q | If someone has coffee mug, then he/she drinks coffee |
| not p | Peter does not have a coffee mug |
| _____ | |
| not q | Peter does not drink coffee |

Now comes the question whether these inferences are valid. A valid inference is one that must be true if the premises are true. According to the material implication, only the MP

and MT inferences are considered valid; the AC and DA inferences are fallacies. This can be explained with the help of the Euler circle depicted in Figure 1. Under the assumption that the claim 'people who own a coffee mug always drink coffee' is true, there are other ways of drinking coffee other than using a coffee mug. This shows the validity of the MP inference: all the cases that someone owns a coffee mug (p) are embedded in all the cases that someone drinks coffee (q). A similar argument can be made for the MT inference. The minor premise is 'someone does not drink coffee'. As everyone who drinks coffee is entrenched within the case 'someone has a coffee mug', that conversely means that anyone who does not drink coffee, does not have a coffee mug as these case are outside the q circle. With DA, knowing someone has no coffee mug offers no information whether that person drinks coffee or not. The ¬ p population can be inside the q field, outside the q field, or overlapping it. The absence of this information makes the DA invalid. Finally, with AC, there are more people who drink coffee than people who drink coffee and own a coffee mug. The person may own a coffee mug or not. Similar to DA, the absence of this information makes the AC invalid.

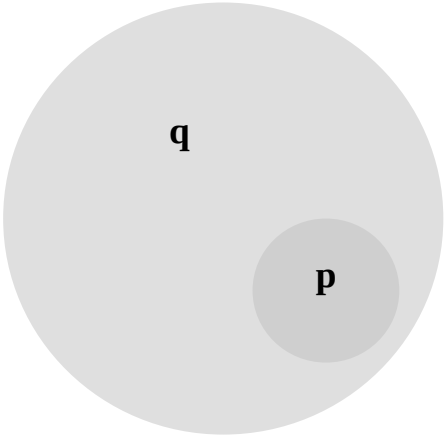



*Figure 1.* Euler circle representation for implication reading of conditional sentences. Taken from Manktelow (2003, pp. 42)

Thus, an argument is deductively correct if the conclusion follows logically from the arguments by applying a givens set of rules. People's ability to understand and apply rules of logic is called *Deductive competence* (Evans, Over, & Manktelow, 1993). Widely used tests that assess *deductive competence* are the Wason selection task, syllogistic reasoning tasks, and deductive reasoning tasks. Deductive performance was tested originally with classical syllogisms. Over the course of time, the deductive inference rules (MP, MT, DA, and AC) became more popular. In a typical set-up, subjects are first presented with the premises of an argument. They are then asked to indicate which conclusion follows logically from the premises or to choose the best candidate from a list of conclusions. Logic proscribes that content of the material has influence on the performance; reasoning performance should depend solely on the syntactic structure of the sentences. However, two main findings persistently showed up: First, people made many logical errors and second their responses were influenced by the supposedly illogical content of the material. The conflict between the prescriptions of logic and actual performance boosted the development of several classical human reasoning theories, which will be presented in the following section.

## 1.3. Theories of reasoning

### 1.3.1. Theories based on logic: Rules vs. Models

There exist several theoretical explanations that aim to explain what mental processes underlie people's performance on making inferences. Theories on reasoning have gone through a developmental process from pure logic to more psychological theories. The logic-based theories, which will be presented first, are the formal inference rule theories and the mental models theory.

#### 1.3.1.1. Formal inference rules theories

The first several theories on reasoning can be grouped together as theories that postulate that reasoning is achieved by abstract general reasoning rules. The human

mind has inherent a mental logic consisting of a set of formal, abstract inference rules or schema's that can be applied in all reasoning contexts, irrespective of content. The first of this kind are the *formal inference rule theories* (e.g. Braine, 1978; Braine & O'Brien, 1991; Rips, 1983). Those belonging to this group of theories treat the proof-theoretic formalization of the propositional calculus as a mental natural deduction system, which was developed by logicians. People construct language-like representations of the premises and make use of cognitive processes that are similar to the natural deductive method (Evans et al., 1993; Garnham & Oakhill, 1994). According to the formal inference rule theory by Rips (1983), deductive reasoning is a process of applying mental inference rules to the premises and conclusion of an argument. These inference rules determine the validity of an argument. If there is a match between the argument and the inference rule, then the argument is valid (Rips, 1983). When people make an inference, they follow three mental steps. First, they seek to detect the logical form of an argument. Second, they apply inference rules to proof whether the conclusion is valid (Evans et al. 1993; Rips, 1983). This mental proof consists of a sequence of steps that are applied to the inference to determine the validity of the conclusion. Each step within this sequence is a logical consequence of the previous steps and they are derived by mean of those inference rules. Modus ponens is such an inference rule. The inference rules are content free and the final step consists of translating the content-free conclusion back into the original content of the premises (Rips, 1983). The inference-rule theory has been criticized for its failure to account for content effects on reasoning (e.g. Cheng & Holyoak, 1985). It builds on classical logic and is purely monotonic in nature whereas everyday commonsense reasoning is typified by non-monotonicity (Byrne, 1989; Chater & Oaksford, 1991; Holyoak & Spellman, 1993). Moreover, simply stating the human mind contains a logic is not sufficient (e.g. Evans, 1991). First, the inference rule theory did not specify which mechanism encodes content into abstract rules and how it decodes these rules back in the original format. Second, a mechanism is missing that explains how the rules are selected and applied to make the inference (.i.e., the comprehension step) (Evans 1991; Byrne, 1989).

### 1.3.1.2. The Mental Models Theory

The *mental models theory* answered more specifically the questions how reasoning problems are represented in the mind and what the mechanism is that acts on these representations (Evans, 1991; Johnson-Laird & Byrne, 1991). The mental model theory states that people construct mental models of the possible state of affairs described in the premises. Johnson-Laird has stated that his model is fundamentally different from the mental rule theory (Johnson-Laird, 1995, as stated in Manktelow, 2004), in that reasoning from premises is a semantic rather than a syntactic process. Syntax captures a set of rules about the form of expressions or statements. Grammar is one such set of rules, as is logic. Semantics, on the other hand, concerns the connection between the premises and the outside world. The mental models theory infers that the reasoning process is not based on inference rules but rather on the construction and evaluation of mental models (Evans, et al., 1993). These mental models represent possible states of the world. The mental theory states that people engage in a three-step thought process to derive at a deduction or judge the validity of an inference. In the first stage, called comprehension, reasoners construct mental models of the possible state of affairs described in the premises. Hereby they make use of the knowledge embedded in the premises as well as prior world knowledge. The key assumption of the mental model theory is '*The principle of truth*'. This means that people construct mental models of propositions in which each model represents a true possibility and whereby each mental model makes explicit only that clause of the proposition that is true in that possibility. For example, the conditional premise of the form *if p, then q* has three mental models that are true possibilities: *p and q*, *not-p and q*, and *not-p and not-q* (Johnson-Laird, Girotto, & Legrenzi, 2004; Johnson-Laird, Legrenzi, & Girotto, 2004; Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999). In most cases, people construct only the mental model *p and q*, called the explicit model, which is explained to be the cause of limited working memory capacity. People do however make 'mental footnotes' of the further implicit models *not-p and q*, and *not-p and not-q,* which can be fully fleshed out upon instruction (Johnson-Laird & Byrne, 1991, 2002). In the second stage, reasoners attempt at formulating a putative conclusion based on models they constructed. This

conclusion should offer new information that was not already explicitly present in the forgoing premises. This mental model of this putative conclusion should hold a possibility in which all premises are true (Johnson-Laird, Legrenzi, Girotto, & Legrenzi, 2000), called the principle of *models of consistency* (Legrenzi, Girotto, & Johnson-Laird, 2003). In case reasoners cannot find such a conclusion, they infer that nothing follows from the premises. In the third and final stage, reasoners search for alternative models of the premises in which their presumptive conclusion does not hold. When they cannot find such a model, then their conclusion is valid. In case they do find such a model, the reasoners go back to stage 2 and look if there is any conclusion true in one of the already constructed models. If so, then they will need to search for counterexamples to it that are consistent with the premises but not with the conclusion. In case they find such a counterexample, it invalidates the inference (Evans, 1991; Johnson-Laird & Byrne, 1991). Below is a schematic overview of this process:

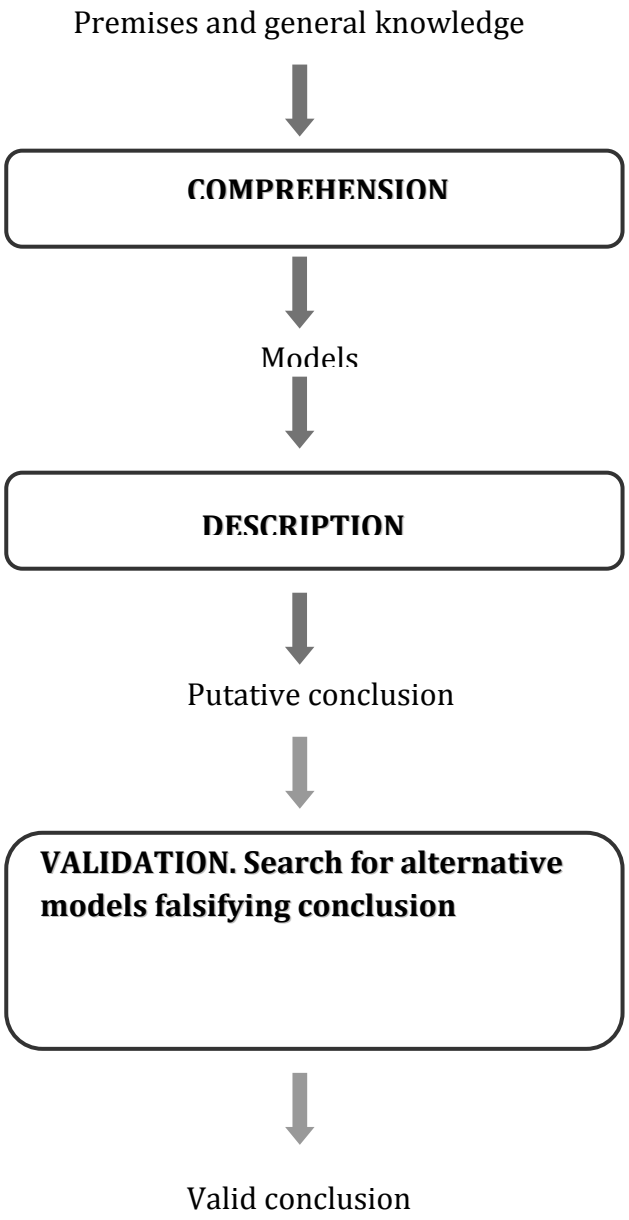



*Figure 2*. The three stages of deduction according to the mental model (taken from Johnson-Laird & Byrne, 1991, pp 36)

To illustrate the process, lets take a conditional reasoning problems that consists of the following two premises:

*If it is a triangle, then it is blue*
*It is a triangle*

According to the mental model theory, the conditional premise is represented as the following mental model:

*[traingle] blue*

*…*

The brackets around 'triangle' indicate that this part is exhaustively represented with respect to 'blue'. That is, 'triangle' cannot occur in any other model unless 'blue' is presented in that model as well. The dotted line presents the remaining models derived from the conditional statement but that are not at first attempt explicitly represented due to working memory constraints. The mental models theory claims to explain why accordingly the modus ponens is an easier inference for people to make than the modus tollens. Because the additional premise with MP informs one that 'traingle' is present then it immediately follows that it must be blue. With Modus Tollens, on the other hand, the additional premise offers the information that it is not blue. Since the possibility '¬ blue' (not-blue) is not explicitly represented, no inference follows directly. To draw the inference, the additional mental models have to be fleshed out:

*[traingle] blue*

*¬traingle [¬blue]*

*¬traingle [blue]*

The only mental model compatible with the given premises is the second model. Because Modus Tollens requires more effort and not everyone succeeds in doing so, it is not made as frequently as the Modus Ponens.

Although Johnson-Laird declares that the MMT deviates from pure logic and is a psychological theory, some argue to the contrary. First, it has been alleged that the MMT shows commonalities with the inference rule-theories in that it adheres to the truth-table of material implication (Evans et al., 1993). In order to make a valid deductive inference one must understand the truth condition of the respective connective (Evans, 1991). Furthermore, it is argued that the search for counterexamples entails a certain logic (e.g. Chater & Oaksford, 2001). More precisely, the initial explicit model and implicit models that form the core of the mental model theory are constructively highly similar to the truth table in propositional logic, in a sense making the assumptions of the mental model theory equal to the material implication (as explained in e.g. Oberauer & Wilhelm, 2003; Johnson-Laird & Byrne, 1991; Johnson-Laird & Byrne, 2002; Johnson-Laird, Byrne, & Schaeken, 1992). The material implication postulates that the human mind adopts the inference rules of a standard logical system (e.g. Garnham & Oakhill, 1994). However, Johnson-Laird & Byrne (2002) argue that they acknowledge the apparent similarity but point out that mental models differ from the lines of a truth table in that mental models represent possibilities and not truth values. Second and related to the first point, in its original format, the MMT stated that reasoning should be unaffected by content. Only later did it make a reference to the role of content (e.g. Johnson-Laird, et al., 1999). The role of content, however, still seems very limited. That is, the content of the models may be increased by additional information from background information possessed by the reasoner but it will not influence the structure of the mental models (Evans et al., 1993). However, exactly how content influences mental models has not been adequately explained (e.g. Oberaurer & Wilhelm, 2003). They also extended their theory to include the effect of probability of the material, in that it affects the number of models constructed (Johnson-Laird, et al., 1999). We will return back to this in the discussion section of this paper.

### 1.3.2. Theories on content and context effects in reasoning

Over the course of development of theories on reasoning, everyday reasoning was placed more on the foreground. That is, pragmatics was recognized to have a substantial influence on inference making. The prime difference between the logic-based theories

and the ones that will be outlined in this section concerns the interpretation of people's performance on laboratory deductive reasoning tasks. The former one explains poor performance as a misunderstanding of basic logic. The current set of theories claim that logical inconsistencies need not necessarily be reasoning failures, but instead might represent pragmatically valid inferences that mirror people's knowledge of events and relations between these in the real world (Cheng & Holyoak, 1985; Cheng, Holyoak, Nisbett, & Oliver, 1986; Evans, et al., 1993). The research reported in this section was motivated by the conviction that everyday life inferences capture a non-monotonic uncertain character and need not be logically valid (Oaksford & Chater, 2009).

The theories and studies laid out in this section concentrate more on comprehension processes which are sensitive to discourse context. Thompson (1994) argues that theories that concentrated mainly on inferential processes, such as the MMT, paid little attention to interpretative processes. Thompson identified two main classes of content factors that are central to these interpretative processes: pragmatic contexts and necessity/sufficiency relations, both of which will be outlined in this section.

### 1.3.2.1. Pragmatic reasoning Schema

The first of such approach to reasoning and that looked at interpretative factors was the theory of pragmatic reasoning schemas by Cheng and Holyoak (1985; Cheng et al., 1986). Cheng and Holyoak (1985) argued that people's reasoning is not based on syntax, context-free rules or specific instances derived from memory. Rather, people reason by means of pragmatic reasoning schemas, which are "…abstract knowledge structures induced from ordinary life experiences…" (pp. 395). Examples of such schemas are permissions, obligations, and causations (Cheng & Holyoak, 1985). For example, permissions are situations whereby an action may only be taken if a certain precondition to this action has been satisfied. If a statement contains the semantics that typify it as a permission situation, then all the rules in general about permission can be applied, including "If action p is to be taken, then precondition q must be satisfied", "Action p is to be taken only if precondition q is satisfied", "if precondition q is not satisfied, then action

p must not be taken". The schema for situations that involve obligations (the occurrence of p infers the necessity that action q must be taken) are similar (Cheng et al., 1986). These knowledge structures are grouped in classes of goals (such as taking action or making predictions about future events) whereby particular relationships are mapped onto these goals (such as cause and effect for causal schemas) (Cheng & Holyoak, 1985; Cheng et al., 1986). Cheng and Holyoak (1985) state that the permission schema is similar to the material implication. However, they point out that the permission schema is not the same as material implication in standard propositional logic because the former is context-sensitive. To explain further, it takes into account terms from formal logic such as "if-then" and "only-if" but the schema's also help to interpret "nonlogical" terms such as 'cause' and 'predict'. However, even though the merits of this theory go more into the direction of explaining human reasoning in context, it only focuses on a restricted set of contexts (e.g. Thompson, 1994). Moreover, it was developed for and applies mainly to the Wason selection task (Wason, 1966) which is a type of deduction task that goes beyond the scope of this dissertation. Later theories acknowledged that context and content should be placed in a much broader framework

### 1.3.2.2. Necessity/sufficiency relations

According to propositional logic, conditional inferences are truth-functional, i.e. they adhere to the logics of a truth-table. Furthermore, conditionals are subjected to the notions of *logical sufficiency* and *logical necessity* (Johnson-Laird & Byrne, 1991). Truth-functional necessity and sufficiency are based on the semantic function imposed on the connective of the truth-functional conditional. According to this function, the antecedent is reckoned sufficient for the consequent and the consequent is reckoned necessary for the antecedent, regardless of the content of the reasoning material (Cummins, 1995). In contrast, psychological theories on human reasoning constitute that necessity and sufficiency denote pragmatic rather than semantic relations between constituent propositions. Because they rely heavily on the reasoner's interpretation of the conditional, they are called *perceived necessity* and *perceived sufficiency* (Cummins,

1995; Cummins, Lubert, Alksnis, & Rist, 1991; Thompson, 2000). The way reasoners represent causal deductive inferences is influenced by their beliefs concerning the causal necessity and sufficiency of a rule's components (Cummins, 1995). By logic, for a relationship to be sufficient, q always occurs when p does (MP) and if q has not happened then p has not either (MT) (Thompson, 1994). However, disabling conditions raise doubt on the sufficiency of the cause to bring about the effect and the reasoner will think of possible disabling conditions. If none come to mind, then the cause is deemed sufficient to have produced the effect. By the same token, alternative causes insert doubt in the reasoner's mind concerning a reported cause for bringing about an effect and therefore the reasoner will consider these other possible causes. When no other cause comes to mind then the cause in question is considered necessary and must have occurred (Cummins, 1995).

The effects of perceived necessity and perceived sufficiency have been empirically manipulated in two different ways. In the first line of research they have been presented as additional premises that represented counterexamples. These additional premises were said to represent conditions that suppress the conditional inference (Byrne, 1989; George, 1997, 1999; Politzer & Bourmond, 2002; Stevenson & Over, 1995) Staudenmayer (1975) and others found support for this with the inferences AC and DA. Byrne (1989) further found that it also holds for the MP and MT inferences. She added a second premise to the first one which makes the antecedent of this first premise insufficient, for example (MP):

*If she has an essay to write then she will study late in the library.*

*She has an essay to write*

*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*

Subjects received either simple conditional arguments as the first example illustrates or conditional arguments containing *additional antecedents* as in the second example. An additional antecedent represents an additional requirement that must also hold.

The subjects received the following conclusions they had to choose from:

*a)* *She will study late in the library*

*b)* *She will not study late in the library*

*c)* *She may or may not study late in the library*

Byrne found that the additional antecedent suppresses accepting the valid inferences MP and MT. If people think the antecedent is not sufficient for the consequent, then they will less likely endorse the inference. The endorsement of the modus ponens dropped from 96% with simple inferences to 37% with inferences containing an additional argument. With modus tollens, it decreased from 92% to 33%. Byrne found that AC and DA are more affected by perceived necessity and MP and MT more by perceived sufficiency. This study and related ones have additionally shown that human reasoning is by no means monotonic. To reiterate, deductive reasoning is according to logic considered to be monotonic. Monotonic reasoning holds that a conclusion that followed from a set of premises cannot be falsified by any additional premise. By contrast, non-monotonic reasoning is retracting a conclusion in the light of new information (Johnson-Laird & Byrne, 1991).

In the second line of research the two mechanisms were implicitly implanted in the conditional of the ordinary argument form (e.g. Cummins, et al., 1991; Cummins, 1995; Liu et al., 1996). Cummins et al. (1991) used the terms “possible causes” ad “possible disabling conditions” to denote perceived necessity and perceived sufficiency, respectively. Subjects were given arguments consisting of a rule and a fact and had to rate to what extent they could draw the conclusions. The conditionals were chosen from a pilot study in which another group of subject had to generate alternative causes and disabling conditions to causal conditional statements. Below are given an example of a low-sufficient MP and a high-sufficient MP:

*If I eat candy often then I have cavities*

*I eat candy often*

*Therefore, I have cavities*

*If my finger is cut, then it bleeds*

*My finger is cut*

*Therefore, it bleeds*

They found that the acceptance rate of the inference problems was low when these problems were based on conditionals with many disabling conditions than when they were based on conditionals with few disabling conditions.

These studies show that the interpretations assigned to conditions are not only sensitive to conversational implications (as in the case of the pragmatic schema theory) but to a broader spectrum of knowledge-based facts; it conforms more to a general reasoning strategy (Cummins et al., 1991).

### 1.3.2.3. Probability theory

Everyday life is full of uncertainty for which we have to find ways to cope with. We aim to reach our goals, make minor and even vital decisions, and of course update our beliefs. Our reasoning in all these cognitive enterprises is far from well aligned with logic. Logic dictates that a valid deduction is one whose conclusion must be true if the premises are true. The theories by Brain & O'Brien (1991) and Johnson-Laird and Byrne (1991; 2001) are logic-based and were modelled to explain people's reasoning behaviour when they are instructed to assume that the major and minor premises are true. However, indicative conditionals can rarely simply be assumed true (Over & Evans, 2003). Human reasoning is characterized not by true statements but instead by uncertain statements. In all

corners of working society, from business to science, there would be no progress if all actions based on plausible conclusions would be discarded.

The finding that people's confidence in a conclusion hinges on the plausibility of the premises, as reported in the former section, led Evans and Over (1996) to argue that people's reasoning inherits a probabilistic approach. Several authors have followed the idea that the mental representation of a conditional is probabilistic. They postulate that the ordinary conditional if p then q is somehow represented psychologically as the conditional probability of q given p (George, 1997, 1999; Liu, Lo, and Wu, 1996; Politzer, 2005; Stevenson and Over, 1995). For example, Liu, Lo, and Wu (1996) tested a probabilistic account by manipulating the perceived probability of q given p. The categories of perceived probability were high, medium, and low. Subjects had to decide whether the conclusion followed logically from the premises on valid and invalid inference problems. They found that the higher the perceived probability of the problems, the higher the correct responses. In a follow-up test, they also found a positive relationship between the probability judgment of the conditional and the endorsement of the set of premises. Evans, Handley, & Over (2003) have shown in a similar fashion that people are less likely to endorse a conditional when the antecedent has a low probability.

The probabilistic approach to human reasoning does not imply logical probability, which is a numerical approach to probabilistic reasoning; it refers to objective situations such as true games of chance. The probabilistic approach to reasoning, rather, is qualitative. Its point of focus is on *degrees of belief* and allows individual judgement (Manktelow, 2004). The idea stems from the assumption that erroneous reasoning is actually an expression of successful human reasoning under uncertainty. In this regard, probability theory is not considered a mathematical calculus, but represents a calculus for rationally updating beliefs. The accompanying normative system for probability judgement is called Bayes' theorem (Manktelow, 2004). A number of authors have taken the Ramsey test as an inspiration and starting point for building their probability theory.

Oaksford and Chater (2001; Oaksford, Chater, & Larkin, 2000) were the first to introduce a probabilistic approach to human reasoning. They later called their qualitative approach to probabilistic reasoning *Bayesian Rationality* (BR; Oaksford &

Chater, 2009). BR consists of four key ideas: First, the probability of a conditional is the conditional probability. This is called the Equation, which assumes that *P* (*if p then q*) = *P* (*q*/*p*) which has long been argued by philosophical logicians (e.g. Edgington, 1995, as cited in Over & Evans, 2003). Second, inferring probability is a subjective process, that is, it is based on *degrees of belief.* Third, a psychological process called the "Ramsey test" (Ramsey, 1931, as cited in Over & Evans, 2003) induces conditional probability. To give the following MP inference as an example: *If it is raining, then Jon will play indoor tennis*. You first hypothetically assume *it is raining* and adjust your other beliefs so they fit with this initial supposition. From these revised beliefs you then "read off" your hypothetical belief that *Jon plays indoor tennis* (Oaksford & Chater, 2009). In doing so, the probability of *P*(1) *It is raining and Jon plays indoor tennis* is compared with the probability of *P*(2) *It is raining and Jon does not play indoor tennis*. To the extent that *P*(1) is higher than *P*(2) then *P* (q | p) is high and *P*(if p then q) is judged to have a high probability. If, by contrast, *P*(1) is lower than *P*(2) then *P* (q | p) is low and *P*(if p then q) is judged to have a low probability (Evans et al., 2003). The fourth and final idea concerns *conditionalization*. This holds that the categorical is not merely assumed, but rather believed or known to be true. With the above example, conditionalization entails knowing that or assigning a high degree of belief to the event *that it rains.* Subsequently, the new degree of belief that Jon plays indoor tennis should be revised to the degree of belief one has in *if it rains, then Jon plays indoor tennis,* i.e. one's degree of belief in the conditional (Oaksford & Chater, 2009). Put formally, the Equation tells you that *P0* (*if it rains then Jon plays indoor tennis*) equals *Po* ( *Jon plays indoor tennis* | *it is raining*). P0 indicates *prior* degree of belief and P1 indicates *posterior* degree of belief. When acquiring that it is raining, then *P1* (it is raining) = 1. Upon this knowledge of conditionalizing, the new degree of belief in Jon plays indoor tennis, *Po* (*Jon plays indoor tennis*) should equal *Po* ( *Jon plays indoor tennis* | *it is raining*), or:

*P1*(*q*) = *P0*(*q p*), when *P1*(*p*) = 1

To exemplify quantitatively, suppose one believes that *Po* ( *Jon plays indoor tennis* | *it is raining*) = 0.8, then finding out that *it rains (P1(it rains) = 1)* the new degree of belief that *Jon plays tennis* should be 0.8 *(P1(Jon plays tennis = 0.8).*

In the psychological literature, the Equation (also called the conditional probability hypothesis) has been experimentally supported (Evans et al, 2003; Oberauer & Wilhelm, 2003; Over, Hadjichristidis, Evans, Handley, & Sloman, 2007). The conditional probability hypothesis has been typically contrasted with the mental model theory. The outcome predictions, put formally, for judging the probability or truth function for the conditional were as follows:

The conditional probability: $P(q \mid p) = P(pq)/P(pq) + P(p\neg q)$
The material implication: $P(pq) = P(pq) + P(\neg pq) + P(\neg p\neg q)$
The conjunctive probability: $P(pq) = P(pq)$

The material implication (or the material conditional hypothesis by Over, et al., 2007) concerns the explicit mental model together with the fully fleshed out cases of the implicit mental models. The conjunctive probability is equal to the explicit model of the conditional statement as provided by the model theory (Oberaurer & Wilhelm, 2003).

Evans et al. (2003) were the first to test directly people's assessment of the probability of the conditional statement. Their aim was to assess whether people's probability judgment of a conditional statement, *if p then q*, equalled that of the conditional probability ($P(q \mid p)$) , the material implication or the conjunctive probability. The experimental setup concerned a story about a deck of cards that consisted of cards on which was printed either a circle or a diamond and were either yellow or red, distributed as follows:

1 yellow circle
4 yellow diamonds
16 red circles
16 red diamonds

The subjects had to rate the possibility of a claim to be true on a scale from 1 to 5, such as:

If the card is yellow then it has a circle printed on it

The probabilities according to the three hypotheses are 5/37, 33/37, and 1/37, respectively. The probability conditional to be true is improbable, probable, or highly improbable depending on which hypothesis is used to derive the truth of the claim. The authors found support for both the conditional probability and the conjunctive probability. However, the strongest support was found for the conditional probability in that ratings for the conditional increased as the probability of pq increased and decreased as the probability of p¬q increased. In addition, ratings also decreased with higher frequency of ¬p cases, which according to the material implication would actually result in higher ratings. Furthermore, there was a low correlation between judging the probability of pq and ¬p¬q, which speaks against the material implication, since this hypothesis regards the two possibilities logically equal. Oberauer and Wilhelm (2003) turned to thematic material and their findings corroborated those of Evans et al. (2003). Over et al. (2007) argued that these foregoing experiments were limited in that they only concerned conditionals about frequency distributions. These experiments used basic conditionals, which are conditionals that have a neutral content not affected by context and background knowledge. These conditionals resemble gambling situations which fall under logic probability, as pointed out earlier. Over et al. state that the conditional probability hypothesis is instead about *subjective* probability judgements. They, therefore, tested the hypothesis with respect to non-basic conditionals related to common knowledge, which are omnipresent in everyday reasoning. They assessed implicit probability judgements by asking subjects to give probability judgements about four natural language conjunctions that are common in ordinary contexts. These correspond to the four possible states of affair with concern to the probability of a conditional as well as the truth table cases for a conditional rating, p & q, p & not-q, not-p & q, and not-p & not-q. The four probability estimates had to add to 100%. The conditional probability was by far the strongest predictor for probability judgements when compared with ratings for

the same statements framed in an explicit task. There was very little support for the conjunctive hypothesis, even less so than in the experiments by Evans et al. and Oberauer and Wilhelm in which abstract material was used. There was no support at all for the material implication.

These studies have shown that people represent statements as epistemic mental models that can present degrees of beliefs or subjective probabilities. Their findings pose a challenge to the truth-functional account of the psychological meaning of conditionals (Oberauer & Wilhelm, 2003).

## 1.5. Belief Revision in Humans

Many of the concepts and distinctions of the above presented deduction theories resemble those of the classical belief revision theories, with the distinction that human deduction theories explain the performance part. The goal of this part is to offer an overview of the stand of psychological research on belief revision conducted to date. Belief revision found its way into the area of psychology by means of several pioneering studies performed by Revlis and colleagues (Revlis, Lipkin, Hayes, 1971; Revlis & Hayes, 1972; Revlis, 1974). Whereas they concentrated primarily on belief revision with syllogistic reasoning problems, relatively more recent research took on conditional deductive inference problems as a framework for belief revision. Apart from this difference, the overall approach is to present a reasoner with two or more statements, presented as beliefs, followed by a belief-contravening assumption that is to be accepted as true. The reasoner's task is then to resolve the logical contradiction between the beliefs and the assumption by choosing to reject one of the beliefs.

### 1.5.1. Empirical studies

Revlis and colleagues used syllogistic inference problems and found that universal quantifiers of the form "All A's are B's" are favoured over particular quantifiers of the form "This Y is a Z". They claimed to confirm an assumption from philosophy that belief revision is based on a preferential ordering (Revlis, et al., 1971; Revlis & Hayes, 1972;

Revlis, 1974). People perform a preferential ordering of statements which accords with a modal ordering. Generalities are assigned a high priority because they represent a suitable rule of inference. Such rules are similar to laws and definitions in science (Revlis & Hayes, 1972). As a consequence, general laws (universal quantifiers) are immune to rejection and get assigned a truth more frequently than particular facts (particular quantifiers) (Revlis, et al., 1971; Revlis & Hayes, 1972; Revlis, 1974). It was further found that universal quantifiers with concrete subject terms such as "All snakes in this forest are rattlers" are accepted more often than universal quantifiers that contained abstract subject terms such as "All creatures in this forest are reptiles" (Revlis & Hayes, 1972).

After two decennia of silence, the study of belief revision was revived by Elio and Pelletier (1997). They framed their experiments in relation to the notion of epistemic entrenchment and the minimal change principle, both central to belief revision in AI, as well as to psychological theories on human deductive reasoning. For their experiments, they manipulated problem form and presentation form, using MP and MT inference rules of either a symbolic or a science-fiction form. The goal of their study was two-fold. First, they aimed to investigate what belief revision choices people make. In this regard, the symbolic problems were included as a baseline to find out whether a problem's conditional premise received entrenchment due to its syntactic form. Second, they used the deductive inference problems in order to link the findings to deductive reasoning theories. This way they could explore whether the inference rule affects the type of subsequent belief revision rule. An example of an MT problem they used in experiment 1 is a follows (p.431):

*On Monday you know the following are true:*

*If an ancient ruin has a protective force field, then it is inhabited by the aliens called Pylons.*

*The tallest ancient ruin is not inhabited by Pylons.*

*Therefore, the tallest ancient ruin does not have a protective force field.*

*On Tuesday, you then learn:*

*The tallest ancient does have a protective force field.*

An example of the symbolic version was as follows:

*If Lex's have a P, then they also have an R*

*Max is a Lex that has a P*

*Therefore, Max has an R*

After presenting the additional information causing an inconsistency, the subjects were offered several "theories" that would reconcile the pieces of information received at consecutive times. Thereby they had the choice to disbelieve the conditional, disbelieve the categorical, or to disbelieve the conditional and render the ground sentence uncertain. The results showed that the subjects preferred to disbelieve the conditional, both for MP and MT problems. This finding was even stronger on natural language problems (.i.e. those with the science fiction material). However, it was not completely symmetrical across the two kinds of inferences, for the effect was stronger for MP problems. For this reason, they inferred that conditionals are disbelieved on syntactic grounds. They further applied a theory formation to their findings which states that conditionals express regularities about the world. If new data contradicts the regularity then the regularity must be flawed. This stands in contrast to the findings by Revlis and colleagues (Revlis, et al., 1971; Revlis & Hayes, 1972; Revlis, 1974).

Since Elio & Pelletier's findings have refuted the claim from AI that conditionals deserve more entrenchment because they describe law-like generalisations, researchers caught up on their research and tried to identify what else entrenchment can be identified as.

Elio (1997) investigated if different kinds of information expressed in the conditional have an effect on its entrenchment. The study was inspired by the work done by Cummins et al. (1991), reported earlier. The different knowledge types were: causal, promise, unfamiliar definition, and familiar definition. With respect to the causal problems, the number of alternative causes and disablers were manipulated. Elio looked, as well, into the influence of type of inference rule, MP vs. MT, on belief revision. The set-up was similar: an initial conclusion followed a conditional and a categorical premise only after which the contradictory information was presented. The subject had to both

assign a degree of belief to the statements and make a forced choice to indicate which belief they would retain and which they would abandon. The most important finding concerns the effect of the number of disablers. When an inference problem involved a many-disabler conditional, then the categorical received a higher degree of belief. When the inference problem involved a few-disabling conditional, then the conditional received a higher degree of belief. The pattern of results was the same for the forced-choice task. This supports findings from deductive reasoning studies by Byrne (1989) and Cummins et al. (1991). In accordance with Elio & Pelletier's (1997) finding, the conditional received higher ratings of belief on the modus ponens problems than on the modus tollens problems. Elio interprets the results in a domain independent fashion, arguing that entrenchment is a matter of alternative possible worlds to the conditional statement. Elio tested in a second experiment whether this influence of knowledge type on belief revision requires *domain knowledge*. She used 'unfamiliar' conditionals composed of the knowledge types 'definition', 'promise', 'cause-and-effect', and 'prediction'. An example of an unfamiliar cause-and-effect conditional is '*If there is a death in a Meorian tribe, then the tribe relocates its camp*'. The findings were similar to the science fiction problems used by Elio & Pelletier, namely that belief in the conditional was lower than in the categorical, and that this was more pronounced for MP than MT problems. Moreover, the conditional gained more belief when it expressed a causal relationship then when it concerned a prediction or a definition. Even when the domain is unfamiliar, people understand a conditional is about a type of relationship between the antecedent and the consequent. Elio argues that belief revision trigger belief-based reasoning. Elio views epistemic entrenchment as a function of the plausibility of alternative worlds. However, it was not mentioned whether the unfamiliar problems were fiction or truth-based. And in case the problems were truth-based, whether the subjects were made aware of it.

A similar view on entrenchment was offered by Politzer and Carles (2001). Belief revision is the result of reassigning strenghts of beliefs to the different statements in a set. In the first of two experiments, they challenged Elio & Pelletier's argumentation that entrenchment is a function of syntactic structure. They did so by including disjunctive and conjunctive syllogisms which are also inference problems with a major premise but not of the conditional kind. They also changed the answer options to give the

subjects the choice to express uncertainty about one of the premises. The answer options were: negation of the conditional, uncertainty about the condition, negation of the categorical, uncertainty about the categorical. Like in Elio and Pelletier's study, the content involved science fiction. They found that uncertainty responses (77.8%) were favoured over categorical denials (16.6%). Also, in accordance with the findings by Elio and Pelletier, the conditional premise was selected two to three times more often than the categorical premise. More importantly, this was also the case with the disjunctive and conjunctive arguments. Because of this, they argued that it is not the conditional nature of the major premise leading to its revision but instead its status quo of being the major premise; its compound nature makes it's more likely to be the erroneous statement. In their second experiment they tested the assumption that the preference for uncertain responses reflects a reassignment of degrees of belief, as was found by Elio (1997). They used high-plausibility and low-plausibility revision problems. For example, the counterpart of the high-plausible conditional '*If his car is beyond repair, Alex will change his car*' was '*If his neighbour buys a new car, Bastien will change his car*'. They found that entrenchment of the conditional depended on its level of plausibility. When the conditional's plausibility was high, then it was favoured over the categorical. However, when the conditional's plausibility was low, then the categorical premise was preferred. Politzer and Carlos subsequently defined entrenchment as a function of degrees of belief and rendered the argument that belief revision is a function of syntax hard to maintain.

Revlin and colleagues examined belief revision on what they called combining problems (similar to DA) and rending problems (similar to MP) which expressed either class-inclusion relations or property-assignment relations. An example of each syllogistic problem expressing a class-inclusion relation is presented below:

*(1a) all snakes in the forest are reptiles*

*(1b) this animal is not a snake*

*(1c) this animal is not a reptile*

*(1d) * assume this animal is a snake*

*(1a) all snakes in the forest are reptiles*

*(1b) this animal is a snake*

*(1c) this animal is a reptile*

*(1d) * assume this animal is a snake*

A property assignment relation would be '*All snakes in the forest have cold blood*'. They also looked at whether belief preferences are dependent on a real-world belief value (like the upper problems) or whether belief acceptance of a generality is merely a function of its role within the structure of the problem. For comparison, they used abstract problems whose propositions were neutralized of probability. The subjects had to indicate, upon receiving the last statement which of the previous ones they wished to reject in order to establish consistency. Across the experiments, the authors found that, in general, the general statement was identified as true more often than the particular statement. However, at closer inspection, the preference for the general statement varied with the logic of the problem. The general statement was assigned truth value more often than the particular statement only on combining problems. Moreover, this preference was stronger on problems with real-world belief value like the problem presented above than on arbitrary problems with abstract propositions such as '*All animals in the forest are mammals*'. On rending problems, in contrast, there was no reliable preference for the general statement for real-world problems and a straight out rejection of the generality when the propositions were abstract. The authors, however, although setting out different theories, did not make conclusive as to which theory explained the results best (Revlin, Cate, & Rouss, 2001).

Dieusaert, Schaeken, De Neys, and d'Ydewalle (2000), like Politzer and Carles (2001), also made modifications to the design of Elio & Pelletier (1997). First, they did not define "beliefs" as "sentences that people are told to accept as true, in the context of resolving some (subsequent) contradiction arising from new information that is provided", and by treating "beliefs" as "propositions that are believed to be true or to be false (or sometimes

to have a belief status of "uncertain")" (Elio & Pelleteire, 1997; p. 421). Rather, their subjects were offered raw data from which they had to obtain the inference rule themselves. Second, they included answer options that gave the subjects the opportunity to indicate doubt rather than merely the forced choice of doubt. Subjects received a cover story about a protagonist investigating the success of certain diets. Along with the cover story came a notebook listing the different diets and whether it led to a loss of weight. The table of the notebook was accompanied with the following instruction: "*In the notebook, the investigator has written down the name of the diet and the result, such as: "Person used 'Fat burn', person lost weight". You may study this table for two minutes. Do you find a regularity in the table of the form "if…then…"? If so, write it down below the table*" (p. 280). These tables were manipulated such that they resulted in a conditional rule with either 100%, 90% or 75% certitude (i.e. the probability that the diet led to weight loss). Each subject received two problems of each certitude level. After the subjects had time to write down their own rule, they received the additional information: "*The investigator has found a particular conditional rule in the table. This rule is ""If "Fat Burn" is used, one loses weight". The investigator decides to make one more observation and discovers a contradiction of the rule. He discovers that "Fat Burn" was used. In line with the rule, he expects the questioned person to have lost weight, but that person gained weight.*" (p. 282). The contradiction in modus tollens form was as follows: "The questioned person didn't lose weight. In line with the rule, he expected the questioned person not to have used "Fat Burn", but the person had used "Fat Burn". The subjects were subsequently given seven answer options:

*Believe rule, disbelieve observation*
*Believe rule, uncertain about observation*
*Disbelieve rule, believe observation*
*Uncertain about rule, believe observation*
*Uncertain about rule, uncertain about observation*
*Disbelieve rule, uncertain about observation*
*Disbelieve rule, disbelieve observation*

In overall, the results showed that expressing doubt was preferred over total disbelief, which is in accordance with Elio's (1997) and Politzer and Carles's (2001) findings. In addition, doubt was applied more often to the conditional premise than to the categorical premise and greater doubt was expressed about conditionals in MP problems than in MT problems. Concerning certitude level, in the 100% certitude level the preferred choice was to believe the rule and disbelieve or express uncertainty about the observation (71%) or express uncertainty about the rule and believe the observation (76%). For the 90% and 75% certitude levels, the preferred choice was to disbelieve the rule and believe the observation (55%) or be uncertain about the rule and believe the observation (90%).

Byrne and Walsh (2005) took a closer look into the nature of revisions. In their first experiment, they focused on the revision process of 'unfamiliar' beliefs. The subjects received a fiction-based cover story, in modus ponens and modus tollens form, that was highly similar to the one applied by Elio & Pelletier (1997). However, after receiving the contravening fact they had to write out their own revision; what they thought to be true after reconciling all the information. The results corroborated earlier findings that people more readily focus on the conditional for revision. However, the authors found an asymmetry in the nature of the subject's revision. To explain, those who concentrated on the conditional (45%) were more inclined to modify (37%) it rather than outright reject (8%) or doubt it. Those who concentrated on the categorical premise (28%) instead were more readily to reject it (24%) than to modify it (4%). In a second experiment, they examined familiar beliefs (Politzer & Carles, 2001; Dieussaert et al., 2000; Dieussaert, Schaeken, & D'ydewalle, 2002; Elio, 1997) stemming from the conviction that they are more entrenched. The inferences contained either causal conditionals, e.g. "*If water was poured on the campfire, the fire went out*" or definitional conditionals, e.g. "*If the animal was a Labrador it was a dog*". Again, the subjects were given modus ponens and modus tollens inferences. After inferring what they thought would follow from the first two premises they were given additional information that came to be known at a later point in time that contradicted their conclusion. In this case, "*The fire did not go out*". They had to write down what they believed to be true taking into account all the information they had at that point. Contrary to the first experiment, they seldom revised the causal or definitional conditionals. Instead, they focused on the categorical. For the conditional

inferences, the difference was 48% and 15% for revising the categorical and conditional, respectively. That is, when the inferred "*It was not a Labrador*" after receiving the above-mentioned conditional and the categorical premise "*It was not a dog*", they chose to reject the categorical fact after receiving the contradiction "*It was a Labrador*". Likewise, for the causal conditional, the difference was 48% and 19% for revising the categorical and conditional, respectively. That is, when the inferred "*The fire went out*" after receiving the above-mentioned conditional and the categorical premise "*Water was poured on the campfire*", they chose to reject the categorical fact after receiving the contradiction "*The fire did not go out*". However, the authors failed to report which plausibility category or categories they chose (e.g. highly plausible, medium plausible, or low plausible). As in the first experiment, when subject focused on the categorical they preferred to outright reject it. By contrast, when they instead focused on the conditional, they rather chose to modify it. The authors conclude that with familiar problems, people concentrate on the categorical and prefer to reject it. This is because conditionals express laws or rules which make them more deeply entrenched and immune to rejection. Furthermore, regardless of familiarity, people more readily reject categoricals and more often modify conditionals.

The above outline gave an overview of empirical studies of human belief revision to date. What comes clear from these studies is that no clear answer has been found yet as to what exactly constitutes belief revision. There is no answer on logical grounds (Revlis, et al., 1971) nor has there been found any other exact mechanism behind belief revision. A contradiction only signals that something is wrong, but does not pinpoint what is wrong (Byrne & Walsh, 2005).

### 1.5.2. Drawbacks from previous belief revision studies

Only a handful of studies have so far been conducted on human belief revision and these have not yet resulted in a satisfying understanding of the workings of belief revision in humans. Any attempts to reconcile these earlier findings have failed. A major cause of this has been the absence of an overall empirical approach. Methodological differences between these studies make it difficult to clarify the contradictions in

findings. The studies differ mainly in task instructions, content of the inference problems' material, and presentation of the inference problems. The methods used in the different studies were too far apart to make any pre-conclusive statements that could serve as building blocks of a firm theory of human belief revision.

A first problem that resulted from these differences is the wide variety of interpretations given to the meaning of entrenchment. Since researchers in AI failed to capture the exact meaning of entrenchment, researchers in the field of human belief revision studies took on this task. However, they were too narrow-focused in that they were looking for a single meaning of entrenchment. Defining entrenchment seemed to rely heavily on content of inference problems and the findings. I will first concentrate here on the studies that found that a particular statement was favoured over a general statement (Byrne & Walsh, 2005; experiment 1; Dieussaert, et al., 2000; Elio, 1997, experiment 2; Elio & Pelletier, 1997; Politzer & Carles, 2001, experiment 1). To reiterate, Elio & Pelletier ascribed it to the syntactic nature of the conditional. It expresses a regularity and from an empiricist viewpoint a regularity merely summarizes data and this makes it fragile for mistakes. Politzer and Carles argue that the status qua of the major premise makes it susceptible for mistakes. Byrne and Walsh state that conditionals get rejected more often than categoricals only when they are unfamiliar. However, these studies, apart from Dieussaert et al., used science-fiction material. Those who found that subjects favoured the conditional over the categorical share the viewpoint that conditionals are more entrenched because they express laws (Byrne & Walsh, 2005, experiment 2; Revlin, et al., 2001; Revlis, et al., 1971; Revlis & Hayes, 1972; Revlis, 1974). Byrne (1989) argued in addition that this holds only for familiar inference problems. However, in these studies the conditional contained a high-probability.

The main problem with respect to this contradiction, is that critical factors such as probability and familiarity were not correctly operationalized and manipulated The science-fiction material was referred to as 'unfamiliar' and conditionals with a real-world belief value, often referred to as 'familiar' inference problems, had a high probability. Diuessaert at al (2000) found that even with high-probable conditionals the conditional is doubted more than the categorical. However, the manner in which the problem was

framed was more of inductive kind, whereby a single observation will make the rule false. Findings from two studies support the above-sketched argument (Elio, 1997, experiment 2; Politzer & Carles, 2001, experiment 2). In both studies the plausibility of the conditional was manipulated and this affected the degree of belief in the conditional.

A second problem is that studies concentrated either only on AI philosophy-based theories or took on a more psychological view whereby only one particular psychological theory was considered. Concerning the first group, in some studies the motivation for the experiments and the interpretation of the results share a very high liking to the ideas of AI belief revision approaches. For example, Politzer and Carlos (2001) linked their findings of to the theories of Harman (1986). Revlin, et al., (2001) and Elio and Pelletier (1997) made an effort to link their findings to deductive reasoning theories but these were logic-based theories that were mainly abstract and descriptive in nature and made little reference to daily life reasoning. Moreover, Elio and Pelletier (1997) constructed their own interpretation of how the mental model theory might explain belief revision as at that time the mental model theory had not yet been extended to explain belief revision. Follow-up studies came closer to the ideas postulated by more psychological theories about human reasoning. Several studies following Elio and Pelletier (1997) found that belief revision depends on prior knowledge and belief (Elio, 1997; Politzer & Carles, 2001). For example, Elio (1997) found that belief in the conditional depends on the number of disabling conditions and that people prefer to assign uncertainty to premises rather than straight out denial. She came to view entrenchment as a function of alternative possible worlds to the conditional statement. Her findings align with the findings of reasoning studies on necessity and sufficiency relations (e.g. Byrne, 1989; Cummins et al., 2001). Politzer and Carles (experiment 2) manipulated probability defined entrenchment as a function of degrees of belief. Although they did not make an explicit reference to any deduction theory, their results parallel the most recent one, that of the probability theory on reasoning.

However, each of these studies failed to design their experiments such that the logic-based theories and more psychologically based theories could be tested independently. This resulted each time in a one-sided view on belief revision such that former researchers believed there was only way in which people revise their beliefs. A study on

belief revision should use a comparison or control group or design the material in any other manner so to allow the comparison of different predictions.

### 1.5.3. The Mismatch principle

So far, there exists only one theory of human belief revision which is called the mismatch principle. The mismatch principle as part of the mental model theory is the first developed theory on human belief revision (Johnson-Laird, 2006; Johnson-Laird & Byrne, 2002; Johnson-Laird, et al., 2004). Recall that people construct mental models of the possibilities that the situation embedded in a set of premises might represent. If people fail to find such a model, they perform the so-called *mismatch principle* to decide which premise of an inference problem will be revised (by which they mean 'discarded'). (Hasson & Johnson-Laird, 2003; Johnson-Laird, 2006; Johnson-Laird, et al., 2004). According to this principle, the statement that will be revised or believed less is that statement, whether it be conditional or categorical, that has a mental model that not only mismatches but also conflicts with the mental model of the contradictory fact. With MP inference problems, the contradicting fact's mental model is 'not-q', this mismatches and conflicts with the mental model of the conditional 'p q' and would therefore be abandoned. With MT problems, the contradicting fact's mental model is 'p', which conflicts with the model of the categorical 'not-q' and therefore people would disbelieve the categorical instead. Note that in this case, the contradicting fact 'p' also forms a match with the antecedent of the conditional statement, a point we will return to later.

Johnson-Laird and his colleagues (Johnson-Laird, 2006; Johnson-Laird et al., 2004) demonstrate the strength of their theory by harmonizing the results of former studies with the mismatch principle. For example, Elio and Pelletier (1997) found that the belief revision was a function of which counter fact followed the belief set. When it was a negation of the consequent, then subjects tended to reject the conditional and believe the categorical statement more. However, when the counter fact was of the form p then they believed the conditional statement more.

Although the mismatch principle has a sound basis, it still lacks sufficient experimental support.

# Chapter 2

# Aim and layout of the dissertation and hypotheses

## 2.1. Aim of the dissertation

Theories on reasoning have seen a development, over a period of decades, from logic consisting of abstract rules to psychological theories more akin to daily-life reasoning. These psychological theories are characterized by a further demarcation between theories focused on domain-general reasoning and theories that emphasize domain-specific reasoning. Human belief revision comprises a relatively new research ground. Due to its being in a very early developmental phase, it still relies heavily on logic-based theories. A human belief revision theory is still awaiting its growth spurt. A good way to induce change is to help guide it into the direction of psychological theories. Even though both the AI belief revision accounts and the psychological theories on human deduction explain how beliefs or sentences are represented, the deduction theories, unlike the AI theories, go further to explain how people generate and bear upon those representations. Moreover, whereas the AI theories and philosophical logic-based theories of belief revision were developed for artificial agents, theories on human (higher-order) thinking have been developed to explain cognitive operations. Finally, theories of belief revision developed within AI (e.g. AGM developed by Alchourron et al., 1985) are grounded on formal logic and the majority of them have barely proven applicable to human belief revision performance. After all, the initial intention to study belief revision in humans was to use the gathered knowledge to better elucidate belief revision processes in artificial agents. For this to happen, we first need a firm theoretical framework for human belief revision.

In this thesis, belief revision is empirically investigated in the context of human reasoning theories. The literature on deductive reasoning shows support for both the MMT and the probabilistic reasoning theories for explaining how people represent premises in their mind and perform deductive reasoning tasks. However, despite the abundant knowledge gathered concerning human deductive reasoning, still very little is know how these theories apply to belief revision. In recent years, belief revision has received attention from scholars in psychology. Although little disagreement exists about the concept of belief revision as being a cognitive process in humans, the mechanisms that underlie this process are still fairly unknown. The overall goal is to perform a

systematic empirical investigation to explore factors that might serve as belief revision strategies within the framework of theories of human deductive reasoning and human belief revision. The overall aim of the current dissertation is to identify factors that convey belief and to find out whether there is one unified strategy of human belief revision or rather a multiplicity of strategies.

Research on belief revision has mainly focused on trying to identify a single belief revision strategy (e.g. Elio & Pelletier, 1997). The demarcation between the use of mental models and prior knowledge in deductive reasoning has been characterized as a difference between general-domain and specific-domain modes of reasoning, respectively (e.g. Thompson, 1994). The current rationalization is that if people's reasoning is based on both domain-general and domain specific modes of thinking, this dual mode of thinking might very well be adopted by people when they engage in belief revision.

The current study pursued two main goals. The first goal was to find out whether believability is affected by mental models or by content and context factors. Do people revise their beliefs by comparing mental models or do they instead judge believability of the material by taking into account knowledge and prior belief? By content factors is meant the linguistic content of a statement. By context factors is meant any factors referring to the social and physical situation in which a statement is uttered. Note that this is different from how Johnson-Laird and Byrne (2002) define context of an utterance by which they mean, among a few other implications, both the linguistic content of an utterance and its social and physical situation. The experiments were constructed such that made it possible to discriminate between the predictions from the two lines of theories. The mismatch principle could be tested by the use of modus ponens and modus tollens problems. Involvement of content and context effects was made possible to test by manipulating several content and context factors.

The first content factor that was explored in this study was *probability* (Experiment I, II, III). The notion that beliefs may be subjected to probability has gained recognition in logic (e.g. Dubois, 2008; Dubois & Prade, 1992, 1993), but very little so far in human belief revision. Given the prominent role of uncertainty and probability in life this might be a feasible factor to consider in belief revision research. The second content factor under investigation was *familiarity* (Experiment III). Most of the studies on belief

revision studied beliefs about either unfamiliar (Elio & Pelletier, 1997; Politzer & Carles, 2001; Experiment 1) or familiar statements (Politzer & Carles, 2001, Experiment 2; Dieussaert, et al., 2000; Elio, 1997). There is, however, only a scarce amount of studies that addressed the effect of familiarity on deductive reasoning (Cummins, 1995; Markovits, 1986; Richardson & Ormerod, 1997), and the topic of familiarity has only been touched upon once in the research of belief revision (Byrne & Walsh, 2005). Byrne and Walsh (2005) adopted the term 'relative entrenchment' and argued that it is a function of familiarity. They found that when the conditional was unfamiliar, the conditional was disbelieved more than the categorical. When the conditional statement had familiar content, people tended to show a preference to disbelieve the categorical. The reason for this, they stated, is that familiar conditionals may express rules and laws that are deeply entrenched and therefore people rather choose to revise the categorical fact. However, the familiar conditionals they used had a high intrinsic probability close to being factual of nature, therefore stripped from any subjective probability. Not all familiar statements express rules and laws and a familiar conditional need not necessarily be high in probability. With respect to the unfamiliar domain, conditionals referred to an imaginary world (similar to Elio & Pelletier, 1997; Politzer & Carles, 2001). People might find it difficult to assign any meaning to the causal relationship depicted in the conditional because they cannot imagine the situation sketched before them and therefore prefer to believe the categorical instead.

The first context factor under investigation was *phrasing of task instruction* (Experiment IV). Belief revision can be a process of rejecting (or decreasing degree of belief) versus accepting (or increasing degree of belief). Consequently, the methodology of studies of belief revision has also varied. These have reflected both the all-or-nothing approach (rejecting versus accepting) and the degree-of-belief approach. Participants have been asked to indicate which statement they outright disbelief or doubt (Elio & Pelletier, 1997; Politzer & Carles, 2001; Byrne & Walsh, 2005), to assign a degree of belief to each statement (Elio, 1997), to select from different options (e.g. negation vs. doubt) (Politzer & Carles, 2001; Elio & Pelletier, 1997), to choose the answer option they support the most (Dieussaert et al., 2000), to choose the statement they wish to reject (Revlin, et al., 2001), to choose which statement to retain and which to reject (Revlin,

Calvillo, & Ballard, 2005), and to choose which statement is true (Byrne & Walsh, 2005). Evans and colleagues have already shown that task instructions influence the search for models in deductive reasoning tasks (Evans, Newstead, Allen, & Pollard, 1994; Evans, Handley, Harper & Johnson-Laird, 1999). It is expected that instructions can not only influence the search for models but also the application of models. Therefore, the influence of task instruction on belief revision warrants investigation. The second context factor that was investigated was *source trustworthiness* (Experiments V, VI, and VII). The factor 'source of the utterance' was first put to the attention by Hovland and Weiss (1951). Source credibility is considered to consist of two factors: source expertise and source trustworthiness (Kelman & Hovland, 1953). The current study focused on the latter one. Source trustworthiness has been defined as the extent to which a source's message is perceived as honest, sincere, and motivated to supply accurate information ( Kelman & Hovland, 1953; Mills & Jellision, 1967). Social psychology has bundled a great amount of studies showing a robust effect of source trustworthiness on persuasion and attitude change, cognitive operations very close to belief revision (Eagly, Wood, & Chaiken, 1978; Hovland & Weiss, 1951; Kelman & Hovland, 1953; Mills & Jellison, 1967; O'Hara, Netmeyer, & Burton, 1991; Priester & Petty, 1995; Rhine & Severance, 1970; Tormala & Clarkson, 2008; Tormala, Briñol, & Petty, 2005; Wiener, & Mowen, 1986). So far, only one study in the area of human reasoning has shown that the source credibility factor 'expertise' influenced how certain people judged the conclusion of a deductive argument (Stevenson & Over, 2001). In their reasoning problems, the conditional and categorical were preceded by a high- (e.g. medical doctor) or low-expertise (medical student) source. They found that in evaluating the validity of the inference, participants relied more on the statement from the high-expertise source. However, no studies have been reported on the effect of source trustworthiness on belief revision.

The second goal was to gain insight in the cognitive processes that lie beneath the belief revision choices and to elucidate the ease of revision strategy. To achieve this, in all four experiments peoples' decision time of their belief revision responses were recorded. Decision time has been investigated in deductive reasoning as a means to bring to the surface the cognitive processes that underlie reasoning, (Evans, et al., 1993;

Johnson-Laird, 1972; Van der Henst, 2002), but decision time in belief revision has not been reported previously.

## 2.2. Manipulation of the variables

### 2.2.1. Probability

Probability was manipulated by inducing different levels of probability onto the conditional statement. In Experiment I, inference problems contained either a high-probable or a low-probable conditional. In Experiment II, the conditional statement of each inference problem expressed either a high, medium or low probability. Since the goal was to find out if conditional probability affects belief revision, no effort was taken at this stage to pinpoint the exact mechanism of mental calculation behind a possible probabilistic approach to belief revision. The subjects were merely asked which of the first two statements they believed more.

### 2.2.2. Familiarity

Familiarity was manipulated by using medium-probability inference problems with either familiar or unfamiliar content. The familiar inference referred to common everyday situations. The unfamiliar content also referred to ordinary real-life situations but from knowledge domains (micro-biology, archaeology, physics, and music instruments) that were unfamiliar to those tested.

### 2.2.3. Task instruction

The context factor 'task instruction' was investigated by using familiar, medium-probability inference problems whereby the participants are asked either which statement to believe more or which to believe less.

### 2.2.4. Source Trustworthiness

This part consists of three experiments. In each experiment, the first two statements were uttered by people whose level of trustworthiness was manipulated. In Experiment

V and VI, the sources were presented as occupations of high, medium, or low trustworthiness. Earlier studies have already shown that people assign different levels of trustworthiness to common occupations (Chun, Campbell, & Yoo, 1975; Rotter & Stein, 1971; Ryckman & Sherman, 1974). In Experiment V, the content of the inference problems was abstract. This was done because the experiment aimed at first finding out if source trustworthiness has any influence on belief revision, hereby avoiding possible unwanted influence of content of the premises. In Experiment VI, content-rich material was used to relate the problems more to real-life situations. In Experiment VII, the sources were presented as the names of two people that were regarded as either high or low trustworthy depending on a positive or negative person description, respectively. In the area of social psychology, it is a well-known phenomenon that whether or not people will accept or be persuaded by a person's statement is influenced by that person's trustworthy personality description (e.g. Petty & Cacioppo, 1986). Again, the content of the material expressed real-life daily actions.

## 2.3. Methodology

Here follow some further methodological issues:

- To stay in line with previous belief revision research and ease cross-study comparisons, MP and MT inference problems were used across all experiments. In most belief revision studies, the focus was on science fiction material or higher-order regularities. Subject were presented with inference problems whose content referred to situations one will likely not find him/herself in. In real life, however, people reason about more mundane issues. In the current investigation a premium is placed on real-life or daily life situations. The inference problems in the current dissertation pertain to real-life situations whereby the if-clause of the conditional premise contained a person's first name. The inference problems consisted of natural language statements. The use of the inference problems in

combination with baseline conditions allowed us to discriminate between competing theories of belief revision.

- Across all studies, the independent variables were *belief preference* and *decision times*. The dependent variables were inference problems (pertains to all studies), probability (Studies I and II), familiarity (Study III), task instruction (Study IV), and source trustworthiness (Studies, V, VI, and VII).
- The term 'revision' within belief revision research has over time developed a negative connotation in the sense that it is referred more to the 'abandonment of a belief' rather than its equally plausible counteraction the 'attainment of a belief'. A more neutral formalization of the concept of belief revision will be set out here. By this, belief revision can be a process of a categorical kind of abandoning versus retaining beliefs as well as a more conservative kind of merely adjusting the degrees of still existing beliefs while welcoming the newly acquired belief.
- A degree-of-belief framework was applied (e.g., Kyburg, 1983; 1984) in the current investigation. Such an approach has already received support in earlier studies of belief revision, in which was found that people prefer to indicate uncertainty or doubt in a premise over straight out rejection (Byrne & Walsh, 2005; Elio 1997; Politzer & Carles, 2001). The degree-of-belief framework resembles the qualitative approach to probability estimates, since people's degree of belief is measured here and not some kind of quantitative measure of belief resembling a mathematical approach (Oaksford & Chater, 2009; Oaksford, Chater, & Hahn, 2008). Several studies on belief revision have indeed shown that people attribute degrees of belief (Elio, 1997; Politzer & Carles, 2001) or certitude levels (Dieussaert et al., 2000) to statement instead of labeling them as categorically true or false. The studies on probabilistic reasoning demonstrate that people attribute degrees of belief to premises. However, in the deduction studies laid out in this section, the participants were asked to either generate a conclusion from premises with ranging certainty levels or rate the certainty of the conclusion. In belief revision tasks, the method and the related cognitive process are reversed.

The new information is introduced as a new fact about the world and induces uncertainty in the premises already held.

## 2.4. Main Hypotheses

Below will be outlined the overall predictions in order to answer the two main aims of the current research:

**Main hypothesis 1:** Both mental models and content and context factors affect belief revision. It is postulated that people will in general be guided by content factors when these clearly confer believability.

**Hypothesis 1.a.:** It is predicted that **probability** is such a factor. That is, probability will guide belief revisions when the event expressed by the conditional expresses either a very low or very high probability. That is, the conditional would gain more belief when it is highly probably and less belief when it is improbable. Inference problems with medium-probable conditionals, however, would activate the use of mental models.

**Hypothesis 1.b.:** **Familiarity** on the other hand, when kept at a medium probability level and expressing real-life situations, is expected not to be an indicator of believability. In this case, it is predicted that people would turn to the use of mental models and apply **the mismatch principle**. This means that people will invest believe in that statement that has a mental model most identical to the one captured by the contradictive fact.

**Hypothesis 1.c.:** With respect to **task instruction**, because the inference problems will be of medium-probability, it is expected that people will apply **the mismatch principle**. However, the effect will be more robust when asked what to believe *more* in stead of *less* because the former mirrors our everyday experience in natural language use.

**Hypothesis 1.d.:** Concerning **source trustworthiness**, it is expected to affect belief revision as evidenced by its robust influence on persuasion and attitude. That is,

belief in the conditional would be highest with a high-trustworthy source, lowest with a low-trustworthy source and at an intermediate level with a medium-trustworthy source. Furthermore, Source trustworthiness would show an effect regardless of how the source was presented.

**Main hypothesis 2:** Our second prediction is that decision times give insight into which belief revision strategy is easier to perform.

# Chapter 3

# Mental models, probability, and familiarity in human belief revision

## 3.1. Introduction

In experiment 1, we aimed to investigate whether probability of conditionals fulfils a believability role in belief revision or whether this carries little value and instead people adopt the mismatch principle and compare mental models to update their belief set (Johnson-Laird, 2006; Johnson-Laird, et al., 2004). If people judge conditionals in a probabilistic manner as voiced by many researchers (e.g. Liu, et al., 1996; Oaksford & Chater, 2001; Oaksford, et al., 2000), they may incorporate probabilistic information embedded in the premises into their belief revision process and base their belief revision choice thereon. To test this, we used MP and MT inference problems with either high or low probability conditionals containing real-life situations.

**Hypothesis 1:** It was hypothesized that when the "if p then q" relationship in the conditional has a clear high or low chance of occurring then people will use this information to guide their belief revision. This holds that people will believe the conditional statement more when it expresses a high probability but instead believe the categorical more when the conditional statement expresses a low probability, irrespective of inference rule.

**Hypothesis 2:** The DTs will reveal which belief revision problem is easiest to perform.

## 3.2. Evaluation study I

A norming study was conducted first to create the material for the experiment.

### 3.2.1 Methods

#### 3.2.1.1. Participants

The pool of subjects consisted of 20 undergraduate psychology students (17, females, 3 males) aged 19 to 34 ($M$ = 22.00, $SD$ = 3.81) who were recruited from seminars and tutorials of a first year course and were tested in the class rooms.

#### 3.1.1.2. Materials

The material for the norming study consisted of 36 conditional statements that expressed 18 low-probable and 18 high-probable events. The conditional statements referred to everyday life "if-then" course of actions that were either very likely or very unlikely to occur based on common knowledge. Each statement referred to a single person whose name appeared in the 'if clause'. Booklets were prepared with general instructions appearing on the first page and the 36 conditional statements presented individually on separate pages, randomized per participant. The booklets were distributed among the participants. The instruction on the first page read as follows (translated from German):

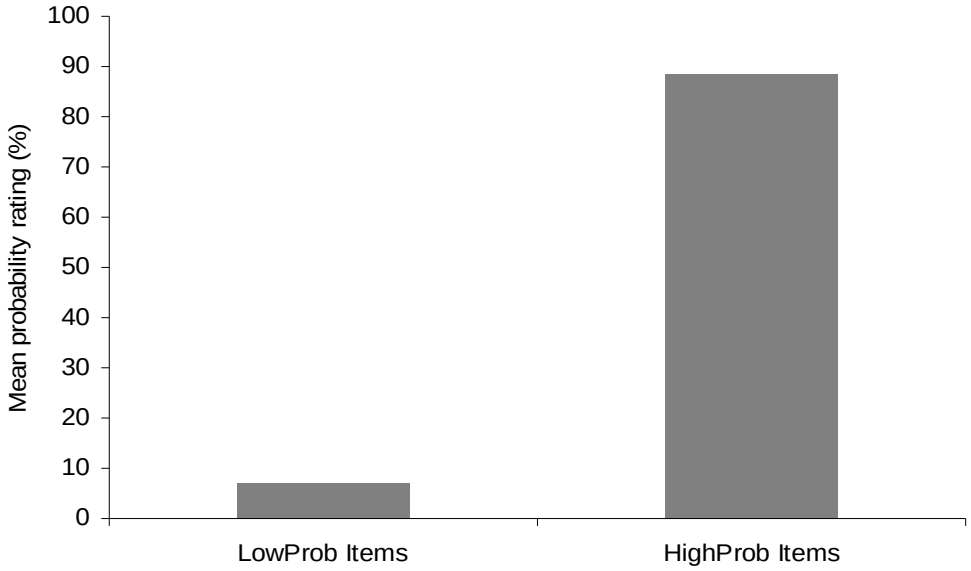
"*On each of the following pages you will be presented with a statement, which is uttered by a person. A rating scale is presented under each statement. On this scale, please rate the possibility that this person is speaking the truth. 0% means "very unlikely" and 100% means "very likely". Please rate the statements in the order they are being presented to you. Please do not go back to previous statements.*"

The scale employed intervals of 10%. Following Girotto and Johnson-Laird's (2004) suggestion, we asked the participants how possible it is that the statement is true. The statements were randomly presented to the subjects and they were allowed to work at their own pace.

### 3.2.2. Results

Descriptive statistics were run over all the statements to find eight statements with the lowest mean probability rating and eight statements with the highest mean probability rating. The overall mean of the eight low-probability statements was 7.19 % (*SD* = 5.70) and the overall mean of the eight high-probability statements was 88.44 % (*SD* = 6.85). A Wilcoxon Signed Ranks Test showed that the difference between these two means was significant, z = -11.030, p < .001.

***Figure 3.*** **Mean probability rating for the two groups of items.**



## 3.3. Experiment I: Belief revision as a function of mental models, and high-and low probability

### 3.3.1 Methods

#### 3.3.1.1. Participants

The participants were 30 students (17 females, 13 males) aged 19 to 31 ($M$ =22.93, $SD$ =2.98) from the University of Giessen. No psychology students were included to ensure the participants had no pre-knowledge of basic logic and reasoning (This applies to Experiments II and III as well). All participants received a small monetary compensation (4 Euro) for their cooperation.

#### 3.3.1.2. Design.

The experiment had a 2 (Inference problem: MP vs. MT) x 2 (Probability of the conditional: High vs. Low) within-subject design. The dependent variables were revision choice and decision time (DT).

#### 3.3.1.3. Materials

From each of the 16 statements from the norming study (8 high-probable and 8 low-probable), a set of MP and a set of MT inference problems were generated resulting in a total 32 problem sets. To create an inconsistency among each set of statements, the third statement was always the opposite of the normative logic conclusion that was expected to follow from the first two statements. This means that for the MP sets of premises, the third sentence was always a negation and for the MT sets of premises, the third sentence was always an affirmation. Table 3 presents the 16 conditional statements used in the experiment. Below are presented two examples.

Example of a low-probability MP problem:

*If Dan goes to work, then he takes the hot air balloon* (p  q)

*Dan goes to work* (p)

*Dan does not take the hot air balloon* (¬q)

Example of a high-probability MT problem:

*If Chris goes to work, then he takes the car* (p  q)

*Chris does not take the car* (¬q)

*Chris goes to work* (p)

The presentation of the 32 problem sets was randomized within subjects. The presentation software package *Super lab 4.0* was used for presenting the items on a computer screen.

***Table 3*. List of conditionals used in Experiment I**

| *High Probability* |
|---|
| If Heiko owns a house, then he lives in it |
| If Nicole has a cat, then she buys cat litter |
| If Laura goes outside, then she keeps her clothes on |
| If Sophia eats in a fine restaurant, then she eats with silverware |
| If Johanna is in her office, then she sits in a chair |
| If Volker is on a business trip, then he stays at a hotel |
| If Kerstin wants to get a tan, then she sits in the sun |
| If Christian is fishing, then he is quiet |

| *Low probability* |
| --- |
| If Anna goes to a job interview, then she wears her oldest pair of jeans |
| If Jan is allergic to pets, then he takes a dog |
| If Marie eats in an expensive restaurant, then she orders rotten fish |
| If Knut goes to work, then he takes the hot-air balloon |
| If Nils receives a salary raise, then he resigns from his job |
| If Tobias turns the heating up higher, then it is warm outside |
| If Lara has a cat, then she gives it a dog bone |
| If Tatjana is playing tennis, then she is wearing high heels |

### 3.3.1.4. Procedure

Participants were tested individually and were seated in a quiet laboratory room in front of a computer. Instructions were presented orally and on the computer screen. Four test items followed to familiarize the participants with the experiment. During the whole experiment, the participants used a response box that consisted of a spacebar-like button to proceed through the experiment and two buttons located to the left and right of this button that were used for making the belief revision choices. The statements within each problem set were presented in successive order in the middle of a black screen. The participants could read the statements in a self-paced manner. The conditional and the categorical statement appeared in white font colour. The third statement appeared in red to signify its counterfactual nature and to aware the participants they had to make a belief revision choice. The participants were explained that the three statements in each problem set were inconsistent and that they had to imagine that each statement was uttered by a different person. They were further informed that the truth status of the

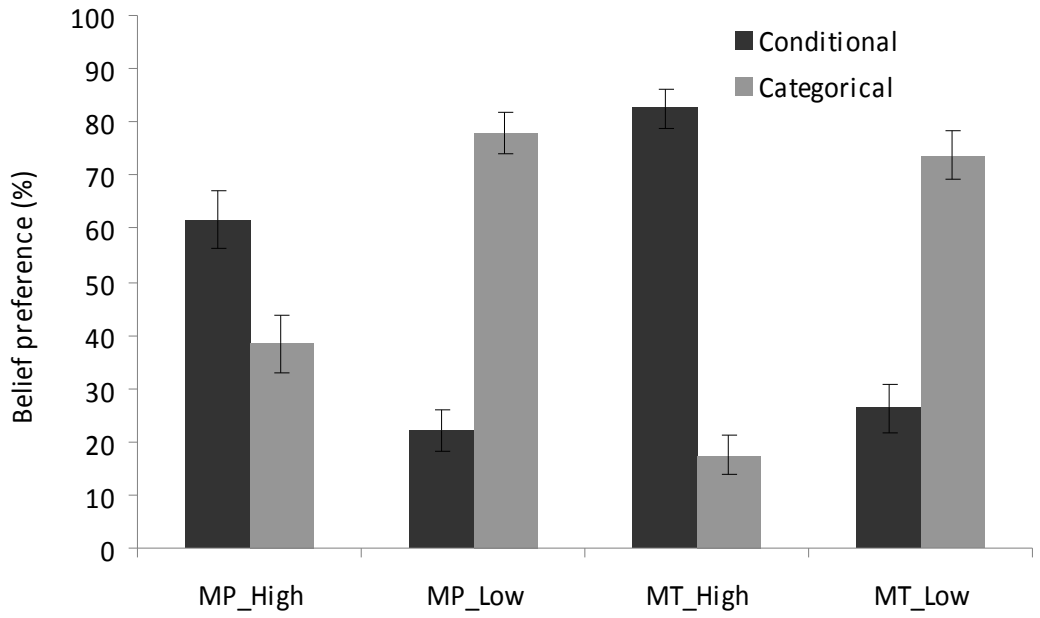
first two statements was uncertain but that the third statement was always true. The third statement then, which was contradictive, formed the basis for their decision. After reading this third statement, their task was to decide which of the first two statements they believed more. The designation of the left and right button to conditional and categorical statements was counterbalanced across participants. The participants had the possibility to take a short break between the presentations of the problem sets.

## 3.3.2. Results

### 3.3.2.1. Belief revision choices

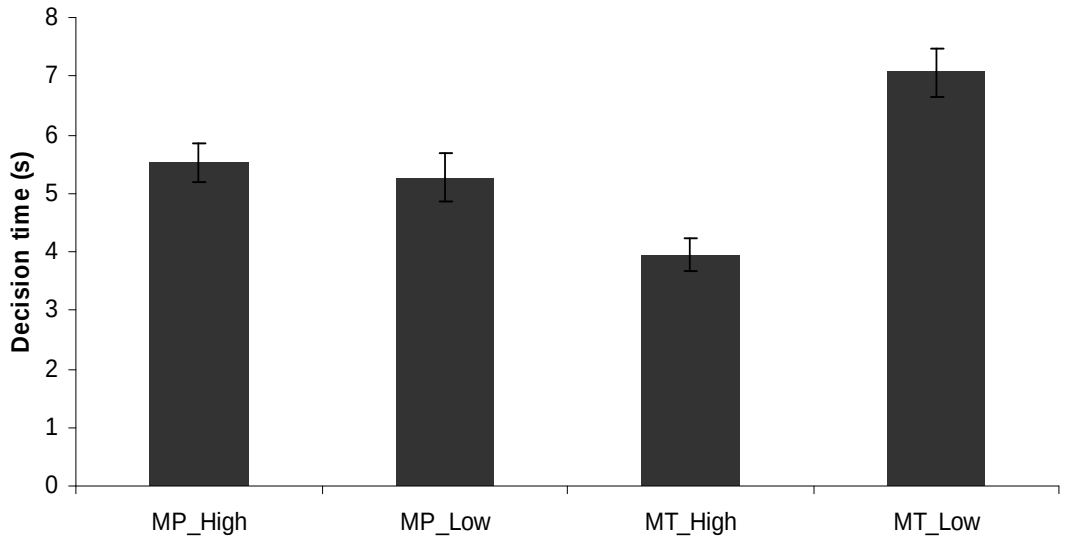
The overall belief in the conditional was 48.12 % (*SD* = 15.19). Belief revision choices in the four conditions are depicted in Figure 4 as the mean preference for choosing the conditional or the categorical statement. In the following, all effects are reported as significant at p<. 05. A 2 x 2 repeated measures analysis of variance (ANOVA) showed a significant main effect of Inference problem, $F(1, 29) = 11.60$, $MSE = .040$, $p = .002$, reflecting a higher endorsement of the conditional for the MT ($M = 54.4\%$) than for the MP problems ($M = 41.9\%$). The results also indicated a significant main effect of Probability, $F(1, 29) = 81.86$, $MSE = .084$, $p < .001$. The conditional was believed more with high-probability ($M = 72.1\%$) than with low-probability inference problems ($M = 24.2\ \%$). Finally, a significant Inference problem x Probability interaction emerged, $F(1, 29) = 9.51$, $MSE = .022$, $p = .004$. Paired-samples t-tests revealed that the conditional was believed significantly more with the high- than with the low-probability inference problems, both for MP ($M = 61.67\%$ vs. $M = 22.08\%$, $t(29) = 6.809$, $p < .0001$) and MT ($M = 82.50\%$ vs. $M = 26.25\%$, $t(29) = 9.259$, $p < .0001$) problems, but this difference was greater for the MT inference problems.

***Figure 4.*** **Mean percentages of belief preferences as a function probability level in Experiment 1.**




**Notes: MP_High = Modus Ponens high-probability; MP_Low = Modus Ponens low-probability; MT_High = Modus Tollens high-probability; MT_Low = Modus Tollens low-probability.**

***Figure 5.*** **Mean decision times of belief preferences as a function of probability level in Experiment 1.**




## 3.3.2.2. Decision times

Figure 5 depicts the mean DTs for the 4 conditions. A 2 x 2 repeated measures ANOVA did not reveal a main effect of Inference Problem, $F(1, 29) < 1$, $MSE = 1.70$. A main effect did show up for Probability, $F(1, 29) = 46.760$, $MSE = 1.33$, $p<.0001$. The decision time was higher for low-probability ($M = 6.17$ s, $SD = 2.06$) than for high-probability inference problems ($M = 4.73$ s, $SD = 1.48$). However, a significant Inference problem x Probability interaction accounted for this effect, $F(86.489)$, $MSE = 0.98$, $p<.0001$. Paired samples T-tests revealed that the difference in decision times between high-and low-probability problems was only significant for MT inference problems. It took subjects significantly longer to decide which statement to believe more with low-probability MT problems ($M = 7.07$ s, $SD = 2.27$) than with high-probability MT problems ($M = 3.95$, $SD = 1.47$), $t(29) = 10.660$, $p<.0001$. With MP inference problems, the decision times were actually higher for high- ($M = 5.51$, $SD = 1.79$) than for low-probability problems ($M =$

5.26, *SD* = 2.25). However, this difference did not reach significance, $t$ (29) = .940, $p$ = .355.

### 3.3.3. Discussion

The findings support our first hypothesis. Belief revision was guided by probability of the conditionals. The conditional was believed more when it conveyed a high probability and the categorical was believed more when the conditional expressed a low probability. This shows that people take the probability of the conditional into account when they make a belief revision choice. This supports the claims made by advocates of probability theories in conditional reasoning (Chater & Oaksford, 2001; Evans & Over, 2004; Evans, 2006; Evans, et al., 2003; Over & Evans, 2003; Oberauer & Wilhelm, 2003; Over et al., 2007). The current findings further extend Politzer and Carles' (2001) finding by showing that probability-based belief revision also holds for MT problems. However, although probability-based reasoning seemed to be the prevailing belief revision strategy, the manipulation of mental models was applied as well. The belief in the conditional was greater for MT problems than it was for MP problems. This finding aligns with the mismatch principle (Hasson & Johnson-Laird, 2003; Johnson-Laird, 2006; Johnson-Laird & Byrne, 2002; Johnson-Laird, et al., 2004).

The DTs findings support our second prediction in that they offered a revelation as to the ease of belief revision performance. The DTs show that the ease of revising one's beliefs varies across content and inference problem. They revealed that the highest DTs were found when the predictions of the two camps of theories would lead to the same belief preference.

Finally, the interaction effects of both the revision choices and the DTs point to a mutual confluence of the use of probability estimates and mental models; the strategy of finding inconsistencies between mental models interacted with the strategy of making probability judgments. This was most clearly reflected by the high percentage of belief for the conditional found for the high-probability MT problems. To explain, with both the high- and low-probability MT set of statements, the mental model of the incontrovertible fact (p) matches the explicit mental model of the conditional statement (p and q). But the subjects believed the conditional more only with the high-probability inference problems.

In that case, then, both strategies converge to the same belief and this strengthened the belief in the conditional. It similarly explains the fastest decision times with the high-probability MT problems; it was most likely the easiest to perform for the subjects. Likewise, the highest DT was found for the low-probability MT problems. With these problems, subjects chose to believe the categorical statement more than the conditional, which is in accordance with the probability theories. However, the mental model of the given fact (p) does not fit the mental model of the categorical statement (not-q). This might have caused a discrepancy in the reasoner's mind.

## 3.4. Experiment II: Belief revision as a function of mental models, and high-medium- and low-probability

In Experiment I we found that both probability and mental models function as believability indicators and guide belief revision, whereby probability-based reasoning was the predominant belief revision strategy. But what happens when the conditional statement is deserted from probability information? We expect that people in that case will rely on mental models. In Experiment 2, we tested this prediction. We included inference problems with a medium probability.

Hypothesis 1: The belief revision choices with the low- and high-probability inference problems will show the same trend as in Experiment I.

Hypothesis 2: With MP medium-probability problems, the categorical statement is expected to be believed more. On the contrary, with MT medium-probability problems the conditional statement should be believed more.

Hypothesis 3: The DTs will be higher for the medium-probability inference problems than for the high-probability- or low-probability inference problems.

## 3.4.1. Methods

### 3.4.1.1. Participants

Forty students (22 female, 18 male) aged 18 to 29 (*M* =22.63, *SD* =2.29) from the University of Giessen were tested in exchange for monetary compensation (4 Euro).

### 3.4.1.2. Design

The experiment applied a 2 (Inference problem: MP vs. MT) x 3 (Probability of the conditional: High vs. Low vs. Medium) within-subject design. The dependent variables were belief revision choice and decision time (DT).

### 3.4.1.3. Materials

For the high- and low-probability inference problem sets, we took four of the eight high-probable and four of the eight low-probable conditionals that were used in Experiment I. Four of the conditionals that were rated around 50% probability in the norming study were used to create the sets of medium-probability inference problems. Table 4 lists the conditionals used in the current experiment. Wilcoxon Signed Ranks Tests showed that the three groups of conditionals differed significantly from each other with respect to mean probability estimate: High- (89.63) versus low-probable (8.25), $z = 3.928$, $p < .0001$, high- versus medium-probable (49.47), $z = 3.922$, $p < .0001$, and low- versus medium probable, $z = 3.924$, $p < .0001$. Figure 6 depicts the mean probability rating for the three groups of items. Form each conditional, a MP problem and a MT problem was constructed, which resulted in a total of 24 sets of inference problems and in the following six conditions: four high-and four low-probability MP problems, four high- and four low-probability MT problems, four medium-probability MP and four medium-probability MT problems. Again, the software program *Super Lab 4.0* was used for presenting the items.

*Figure 6.* Mean probability rating for the three groups of items.

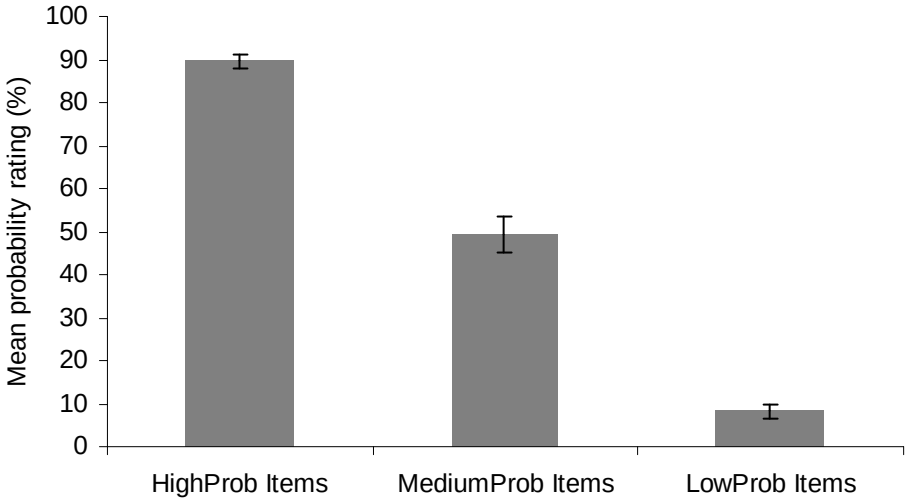


*Table 4.* List of conditionals used in Experiment II

| |
|---|
| *High probability* |
| If Nicole has a cat, then she buys cat litter |
| If Sophia eats in a fine restaurant, then she eats with silverware |
| If Volker is on a business trip, then he stays at a hotel |
| If Kerstin wants to get a tan, then she sits in the sun |
| *Medium probability* |
| If Bruno is hungry, then he eats a cracker |
| If Hannah remains on shore, then it is stormy weather |
| If Anke is resting, then she listens to classical music |
| If Hugo is allergic to fur-bearing animals, then he gets a fish |
| *Low probability* |
| If Anna goes to a job interview, then she wears her oldest pair of jeans |
| If Jan is allergic to pets, then he takes a dog |
| If Nils receives a salary raise, then he resigns from his job |
| If Tobias turns the heating up higher, then it is warm outside |

### 3.4.1.4. Procedure

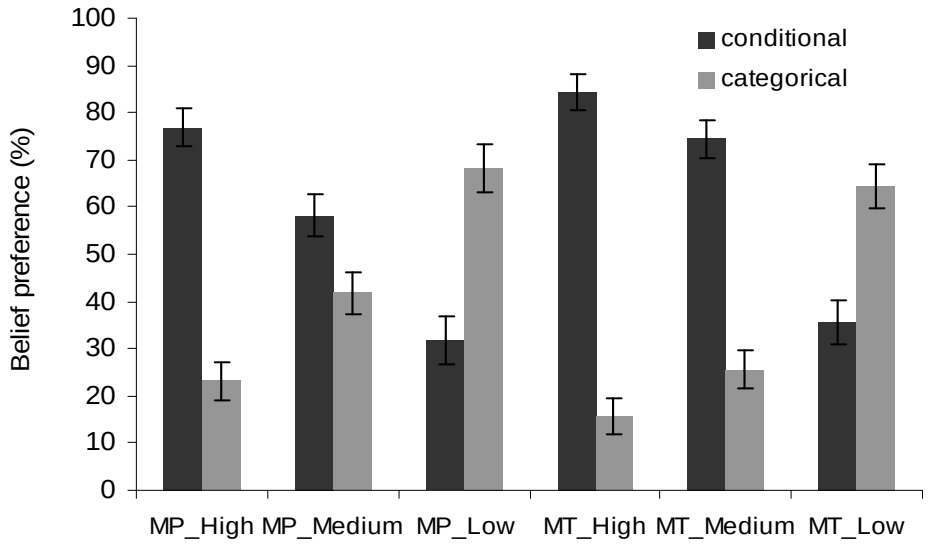
The procedure of Experiment II was the same as in Experiment I.

## 3.4.2. Results

### 3.4.2.1. Belief revision choices

The overall belief in the conditional was 60.21 % (*SD* = 15.15). Mean percentages of belief revision choices in the four conditions are depicted in Figure 7. The factors Inference problem and Probability were submitted to a Repeated Measures ANOVA. All effects are reported as significant at $p<.05$. As in Experiment I, a main effect of Inference Problem was detected, $F(1, 39) = 8.921$, $MSE = .057$, $p = .005$, such that the conditional was believed significantly more often with MT ($M = 64.80$ %) than with MP problems ($M = 55.60$ %). Also, a main effect of Probability emerged, $F(1.59, 62.14) = 47.441$, $MSE = .122$, $p<.0001$. Contrasts revealed that the conditional was believed significantly more often with high-probability inference problems ($M = 80.60\%$) than with medium-probability inference problems ($M = 66.30\%$), $F(1, 39) = 16.492$, $p<.0001$, and in turn significantly more often with medium-probability inference problems than with low-probability inference problems ($M = 33.80\%$), $F(1, 39) = 38.828$, $p<.001$. An interaction between Inference problem and Probability leaned toward significance, $F(2, 78) = 2.503$, $MSE = .033$, $p = .088$.

***Figure 7.*** **Mean percentages of belief preferences as a function of probability level.**

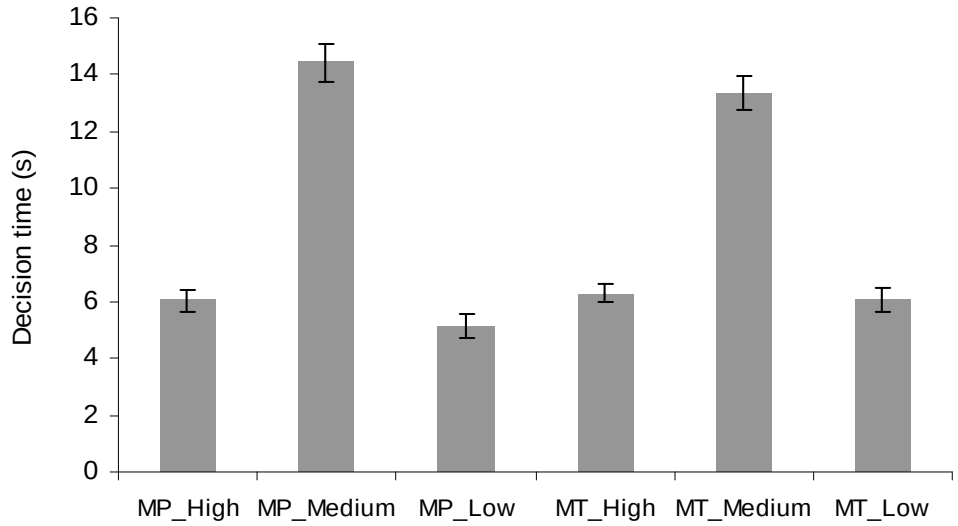



## 3.4.2.2. Decision times

Figure 8 depicts the mean decision times for all the conditions. A 2 x 3 analysis of variance (ANOVA) was carried out to test the effects of Inference problem (MP vs. MT) and Probability (High vs. Low vs. Medium). No main effect of Inference problem was found, $F(1, 39) < 1$, $MSE = 3.26$. There was a significant main effect of Probability, $F(1.65, 64.35) = 257.852$, $MSE = 8.06$, $p<.001$. Post-hoc tests showed that all comparisons were significant: High- versus medium probability, $t(29) = 4.061$, $p < .0001$; High- versus low-probability, $t(29) = 8.135$, $p < .0001$; Medium- versus low-probability, $t(29) = 6.231$, $p < .0001$. There was also a significant Inference problem x Probability interaction effect, $F(1.65, 64.37) = 5.670$, $MSE = 4.50$, $p =.008$. With regard to MP problems, the DTs differed significantly between all three probability levels: High- ($M = 6.04$) versus medium-probability ($M = 14.42$), $t(39) = 14.717$, $p < .0001$; High- versus low-probability ($M = 5.13$), $t(39) =2.073$, $p = .045$; Medium- versus low-probability, $t(39) = 15.681$, $p < .0001$. With respect to MT problems, the DTs differed significantly between high- ($M =6.31$) and

medium-probability ($M$ =13.35), $t$ (39) = 14.861, $p$ < .0001, between low- ($M$ =6.07) and medium-probability, $t$ (39) = 12.222, $p$ < .0001, but not between high- and low-probability inference problems, $t$ (39) = .723, $p$ = .474.

*Figure 8.* **Mean decision times in seconds as a function of probability level.**




## 3.4.3. Discussion

The first prediction was confirmed. The responses on the high- and low-probability inference problems again demonstrate support for the probability theories (Chater & Oaksford, 2001; Evans & Over, 2004; Evans, 2006; Evans et al., 2003; Over & Evans, 2003; Oberauer & Wilhelm, 2003; Over et al., 2007). Furthermore, the belief revision choice pattern portrayed, as in Experiment I, an additional but a less strong influence of mental models in that the conditional was in overall believed more with MT than MP inference problems albeit without the interaction between the two strategies.

The belief revision choices for the medium-probability problems, however, did not follow the predictions of the mismatch principle (Hasson & Johnson-Laird, 2003; Johnson-Laird, 2006; Johnson-Laird & Byrne, 2002; Johnson-Laird, et al., 2004), disconfirming the second hypothesis. For both MP and MT problems, the conditional was

believed more. Several explanations can account for this finding. The conditionals that were meant to express either a high or low probability also did so quite clearly. Therefore it could have been that the participants picked up on this difference and reasoned that these were the only two probability levels. Since events expressed in the medium-probability problems were still about 50% likely to occur (as measured in a norming study), the participants could have judged them to belong more to the high probability problems and chose accordingly to believe the conditional more for these problems, irrespective of logic inference. Another line of argumentation for the finding, related to the first one, is that it can be argued that people embrace the conditional statement when they can imagine the event it conveys and at the same time judge the event reasonably likely to occur. In this regard, the role of familiarity comes into play.

The DT findings supported the third hypothesis in that belief revision took longer with medium-probability inference problems than with high- or low- probability inference problems. However, the findings for the high- and low- probability inference problems are in contrast to what was found in Experiment I. There they did differ significantly with longer DTs for the low- probability problems. A possible cause is one of a contextual kind, that of probabilistic variation. From the above finding, one can deduce that belief revision is more difficult with medium-probability than with high- and low-probability inference problems. In the face of the medium-probability problems, then, the low-probability problems might just not seemed as difficult as likely was the case in Experiment I. Another contextual cause is that the high increase in DTs for the medium-probability inference problems mirrors a switch from dealing with a more common set of inference problems with clear probability levels to a less common set of problems with less clear probability levels; the low and high probability inference problems occurred a total of 16 times, whereas the medium-probability problems occurred only 8 times.

## 3.5. Evaluation study II

### 3.5.1. Methods

#### 3.5.1.1. Participants

Twenty subjects (14 females, 6 males) aged 19-51 ($M$ = 23,60, $SD$ = 7,02) were recruited for the norming study. The students were from different faculties of the University of Giessen.

#### 3.5.1.2. Materials

Conditional statements with familiar and unfamiliar content were created for the experiment. The unfamiliar conditionals referred to the domains of micro-biology, physics, music instruments, and archeology. These domains were chosen because they are familiar and meaningful only to people who are experts in the respected fields but at the same time represent real-life if-then course of actions. The conditionals were checked for correctness by researchers working in the faculties of the above-mentioned domains at the University of Giessen. The familiar conditionals referred to everyday course of actions the participants would be familiar with, either by engaging in these activities themselves or knowing about them. They were similar to the conditional statements used in Experiment II representing the medium-probability statements. Both familiar and unfamiliar conditionals were designed to occur at or around a 50 % chance level. All conditionals referred to a single person performing a particular action with the person's name always appearing in the if-clause. A total of 48 conditionals were created; 24 familiar and 24 unfamiliar conditionals (6 of each unfamiliar domain).

Booklets were created consisting of these 48 conditionals, each printed on a separate page. Under each conditional statement was a rating scale from 0% (very unlikely) to 100% (very likely) with intervals of 10%. Also included under the rating scale were two check boxes, one for familiar and one for unfamiliar.

### 3.5.1.3. Procedure

The booklets were distributed among a group of 20 students who were instructed to rate for each conditional how probable they think it is true and if they were familiar with respect to the sentential content. The instructions on the first page were as follows, translated from German:

*"On each of the following pages, a statement is presented that is uttered by a person. Under each statement is printed a rating scale. Please rate how probable it is, that this person is speaking the truth. 0% means "very unlikely" and 100% means "very likely". Please also mark for each statement, whether you are familiar with the domain presented in the statement. Please rate each the statements in the order that they are presented to you. Please do not turn back to previous pages."*

With further verbal instructions, it was explained to the participants that with the term 'familiar' was meant understanding the content presented in the conditional either by personal experience or by use of prior world knowledge. Below is an example of a unfamiliar conditional in the booklet to show how the conditionals were resented to the subjects:

If Paul represses genes, then he mutates certain promoters

| 0% | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
|---|---|---|---|---|---|---|---|---|---|---|
| Very unlikely | | | | | | | | | | Very likely |

Familiar □ Not familiar □

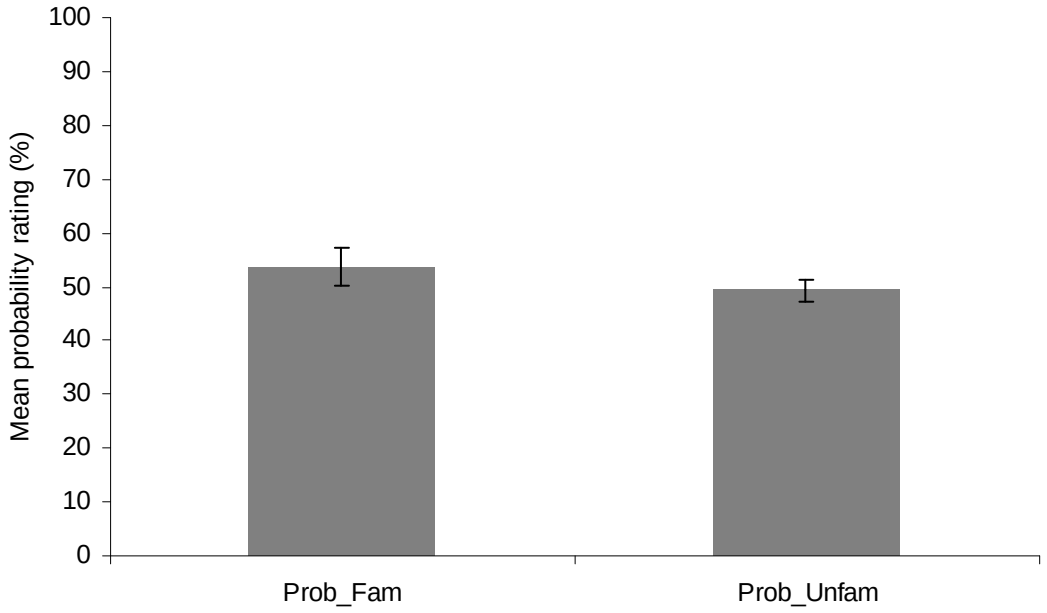
### 3.5.2. Results

Descriptive statistics were run to find eight conditionals rated closest to 50% from both the familiar and unfamiliar conditionals. Figure 9 shows the mean probability ratings for the groups of familiar and unfamiliar conditionals. The eight unfamiliar conditionals consisted of two conditionals of each of the four expertise domains. A

Friedman Test showed that the conditionals in each group did not significantly differ from each other with respect to probability, $\chi 2\,(7) = 8.701$, $p = .275$ and $\chi 2\,(7) = 3.498$, $p = .835$ for the eight familiar and eight unfamiliar conditionals, respectively.

A Wilcoxon's Signed Rank Test showed that the eight familiar conditionals were rated as significantly more familiar than the eight unfamiliar conditionals, $Z = -3.689$, $p<.0001$. Also, the two sets were rated equally probable (53.71% and 49.31% for the familiar and unfamiliar conditionals respectively), $Z = -1.065$, $p = .287$. Thus, the manipulation checks were successful.

***Figure 9.*** **Mean probability rating for the two groups of items.**



## 3.6. Experiment III: Belief revision as a function of mental models, and familiarity

The results from Experiment II motivated to explore whether familiarity or the variation of probabilities underlaid the lack of support for the mismatch principle. This gives the further possibility to investigate how familiarity affects the course of belief revision. For the current experiment we used familiar and unfamiliar MP and MT problems with medium-probability conditionals. We used a different approach to conceptualize familiar and unfamiliar statements than was done by Byrne and Walsh (2005), which will be explained in detail in the method section.

Hypothesis 1: If indeed, as was found in Experiment II, people have a tendency to believe the conditional more when it is familiar and reasonably likely to occur, then it is expected that the conditional will be chosen more with familiar sets of problems irrespective of inference problem. Likewise, if the problem content is unfamiliar, then the participants should show a tendency to shy away from believing the conditional and instead choose to believe the categorical statement. This would support Byrne and Walsh (2005). If, instead, familiarity is not an indicator of believability then in both cases people would revise their beliefs according to the mismatch principle.

Hypothesis 2: In case the mismatch principle guides belief revision, familiarity will influence the usage of the mismatch principle.

Hypothesis 3: Again, the DTs will be fastest in that case where the predictions from the two lines of theories lead to the same belief preference.

### 3.6.1. Methods

#### 3.6.1.1. Participants

The participants were 40 students (23 females, 17 males; aged 20 to 37, $M$ =23.03, $SD$ =3.47) recruited from different institutes of the University of Giessen. No psychology

students were included to ensure that the participants had no pre-knowledge of basic logic and reasoning. They received monetary incentives (6 Euro) for participation.

### 3.6.1.2. Design

A 2 x 2 within-subject factorial design was used. The independent variables were Inference problem (MP vs. MT) and Familiarity (Familiar vs. Unfamiliar). The dependent variables were revision choice and decision time (DT).

### 3.6.1.3. Materials

From both eight familiar and eight unfamiliar conditional statements, MP and MT sets of inference problems were created which resulted in a total of 32 problem sets for the experiment. Table 5 lists the conditionals used in Experiment III. Below are examples of a familiar MP problem and an unfamiliar MT problem:

*If Kerstin visits a friend, then she brings flowers*
*Kerstin visits a friend*
*Kerstin does not bring flowers*

*If Paul activates genes, then he mutates specific promoters*
*Paul does not mutate specific promoters*
*Paul activates genes*

*Table 5.* List of conditionals with used in Experiment III.

| *Familiar Conditional* |
| --- |
| If Katharina cooks in the evening, then she prepares a rice dish |
| If Volker wants to listen music, then he puts on classical music |
| If Kerstin visits a friend, then she takes flowers |
| If Christian is working in the office, then he leaves open the door |
| If Sophia is at home on the weekends, then she works in the garden |
| If Hannah stays on shore, then it is stormy weather |
| If Peter leaves his house, then he exists via the backdoor |
| If Heiko goes outside, then he puts on his sunglasses |

| *Unfamiliar Conditional* |
| --- |
| If Knut is learning to play the trumpet, then he starts with the short b. |
| If Lara is tuning the piano, then she inspects three tunes per note |
| If Martin detects low gas pressure in the detector, then he recalibrates the threshold |
| If Hendrik is working with the Silizium Vertex detector, then he does particle tracking |
| If Paul represses genes, then he mutates certain promoters |
| If Susan increases mutation rate, then she uses mutagens |
| If Wanda cleans the bottles, then she removes the planum |
| If Tim cleans Pollen, then he shakes them through a sieve |

### 3.6.1.4. Procedure

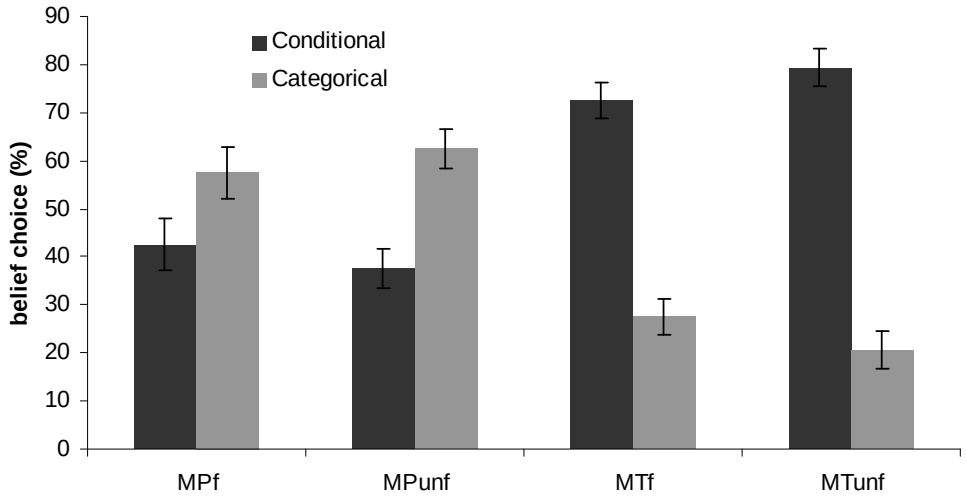
The current experiment followed the general procedure described in Experiment I with the exception that participant were asked, after the experiment, whether they were actually familiar with any of the 'unfamiliar' problems. This was not the case.

## 3.6.2. Results

### 3.6.2.1. Belief revision choices

The overall belief in the conditional was 57.97 % (*SD* =18.49). The mean preferences to believe the conditional or categorical statement more are presented in Figure 10. A repeated measures ANOVA revealed a main effect of Inference problem, $F$ (1, 39) =43.877, *MSE* = *.118,* p<.0001; the conditional was chosen at a much higher rate with MT problems (75.94%) than with MP problems (40.00%). No main effect appeared for Familiarity, $F$ (1, 39) <1, *MSE* = .026. However, an Inference problem x Familiarity interaction effect emerged, $F$ (1, 39) =6.973, *MSE* =.020*,* $p$ =.012. Post-hoc tests revealed that the difference in the percentage of believing the conditional with MP familiar problems ($M$ = 42.50%) was not significantly higher than with MP unfamiliar problems ($M$=37.50%), t (39) = 1.496, p = .143. For MT inference problems, the difference in believing the conditional with familiar ($M$ = 72.50%) and unfamiliar inference problems ($M$ = 79.38%) bordered significance, t (39) = 1.984, p = .054. Thus, with MT inference problems, a higher belief in the conditional was a more robust scenario with unfamiliar than with familiar problems.

***Figure 10.*** **Mean percentages of belief preferences as a function of familiarity.**



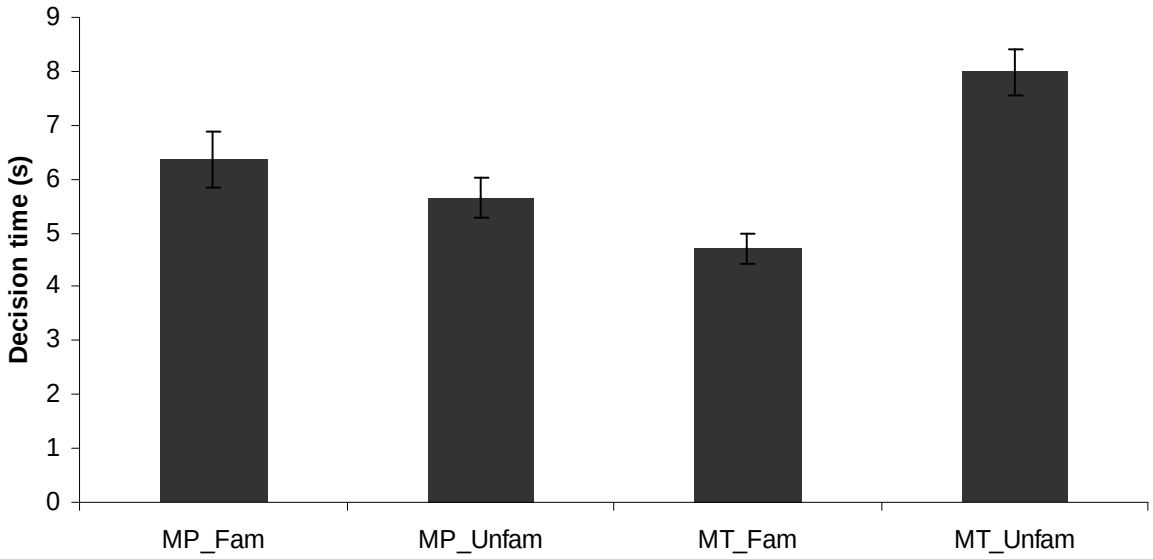

**Notes: MP_Fam = Modus ponens familiar inference problems; MP_Unfam = Modus ponens unfamiliar inference problems; MT_Fam = Modus tollens familiar inference problems; MT_Unfam = Modus tollens unfamiliar inference problems.**

## 3.6.2.2. Decision times

Figure 11 depicts the mean DTs for the 4 conditions. A Repeated Measures ANOVA did not reveal a main effect of Logic, $F(1, 39) = 2.333$, $MSE = 2.030$, $p = .135$. A main effect of Familiarity was detected such that it took subjects significantly longer to decide which statement to believe more with the unfamiliar inference problems ($M$=6.82 s) than with the familiar inference problems ($M$ = 5.54 s), $F(1,39) = 69.214$, $MSE = 0.949$, $p$=.001. However, a significant Familiarity x Inference problem interaction effect showed that this was only the case for MT problems, $F(1, 39) = 81.272$, $MSE$=1.962, $p$=.001. Post-hoc tests revealed that with MT problems it took subjects significantly longer to revise their

belief with unfamiliar problems ($M$ = 7.99 s) than with familiar problems ($M$ = 4.71 s), $t$ (39) = 11.03, $p$=.001. In contrast, with MP problems, the DTs were actually significantly higher with familiar ($M$ = 6.37 s) than with unfamiliar problems ($M$ = 5.65 s), $t$ (39) = 2.99, $p$ =.005.

*Figure 11.* **Mean decision times in seconds as a function of familiarity.**




## 3.5.3. Discussion

Mental models served as the overruling belief revision strategy in the absence of a clear probability; the categorical is believed more with MP problems and the conditional is believed more with MT problems. Based on this, the failure to find support for the mismatch principle in Experiment II was most likely due to the difference in probability levels. This further refutes the claim made by Byrne and Walsh (2005) that for familiar material belief in the conditional will win preference, independently of inference problem, because familiar conditional statements have a law-like nature. They found a preference to disbelief the categorical with familiar problems and to disbelief the

conditional with unfamiliar problems. However, their familiar conditionals were inherently high in probability and their unfamiliar problems referred to an imaginary world (similar to Byrne & Walsh, 2002; Elio & Pelletier, 1997; Politzer & Carles, 2001, Experiment 1). Supporting the second hypothesis, familiarity had a modulating influence on the usage of the mismatch principle. The pattern of using mental models to derive at the belief preference was more robust for the unfamiliar than for the familiar inference problems, which only strengthens support for the mismatch principle. Most likely, the content of unfamiliar problems has no informational value, with the consequence that people have no other choice than to use mental models to derive at their belief revision choice. This again shows a confluence of the two belief revision strategies. The highest percentage of belief was found with MT unfamiliar inference problems. The abstract nature of the conditional due to its unfamiliarity likely eased the reliance on mental models.

The DTs offer further interesting insights; the DTs did not parallel the revision choices. Intuitively you would expect unfamiliar problems to be more difficult than familiar problems. But based on the revision choices you would actually in this case expect the DTs to be faster for unfamiliar problems. However, this was the case only for MP unfamiliar problems. This demonstrates again the partnership of the two strategies. Although matching of mental models demonstrated to be the prevailing belief revision strategy, content of the inference problems still exerted its influence, albeit to a lesser degree. To explain, with MT unfamiliar problems, people would prefer to believe the conditional due to a clear match but nevertheless might find it a difficult task to perform because they possess no knowledge of the content. In contrast, the fast DTs for the MT familiar problems presumably result from the increased ease of choosing the conditional due to both its familiarity and the matching of mental models. The non-equivalent pattern between revision choice and the underlying DTs implies that the choice of revision does not coincide with ease of revision. In short, people use mental models as a guide to resolve an inconsistency between statements that lack clear probability, both with familiar and unfamiliar inference problems.

# Chapter 4

# The effects of task instruction and mental models on belief revision

## 4.1. Introduction

In experiments I-III, the mismatch principle was contrasted with the content factors probability and familiarity in the study of belief revision. It was tested whether they influence belief revision in a way similar to how their influence has been reported in deductive reasoning studies. This was found to be the case. Belief revision was guided mainly by the mismatch principle and probability. The results indicated that multiple belief revision strategies exist and that there was always one overruling strategy and another one modulating it. Probability and the use of mental models functioned as independent belief revision strategies. That is, people invested more belief in the conditional than in the categorical statement when the former contained a high probability. Similarly, when the conditional statement was of low probability, then people preferred to believe the categorical statement more. This finding is in line with the probability theories of human reasoning (Oaksford & Chater, 2001; Oaksford, et al., 2000). By contrast, when the conditional statement had a near 50% probability, then people converted to the use of mental models to resolve the inconsistency. That is, the categorical statement was believed more with MP problems and the conditional statement was believed more with MT problems, supporting the mental model theory (Johnson-Laird & Byrne, 2002; Johnson-Laird, et al., 2004). Familiarity did not lead to a distinctive belief revision strategies but rather modulated the strength of the utilization of mental models. That is, mental models were used to a greater extent with unfamiliar than with familiar inference problems. This was attributed to the fact that the unfamiliar material was entirely devoid of background information which makes people rely even more on the more abstract and general revision process of using mental models. Thus, people adjusted their belief in the conditional to the perceived probability of the conditional statement, leading to one of two belief revision strategies.

These former experiments showed a joint engagement of more specific knowledge-based revision processes based on probability and a general models-based process based on mental models, as has been shown in reasoning performance. However, content factors are only a minor part of the orbit of possible factors leading to specific belief revision strategies. To uncover further strategies, the study of belief revision should go

beyond linguistic content. The focus will turn now on context factors. Chapter IV reports on an experiment that tested the first of two context factors under investigation here, namely 'task instruction'. Task instructions have varied greatly in former belief revision studies (Byrne & Walsh, 2005; Elio, 1997; Elio & Pelletier, 1997; Dieussaert et al., 2000; Politzer & Carles, 2001; Revlin, et al., 2001; Revlin, et al., 2005). Findings from Experiment III, in which medium probability inference problems were used, supported the mismatch principle. However, Experiment III applied only one kind of task instruction: asking the participants to choose the statement they believe more. In Experiment IV, the participants are asked either to choose which statement they believe more or which they believe less. By this, we wanted to investigate what influence phrasing of task instructions might have on model construction and in turn believe revision. Of additional interest was to find out which cognitive process is easier to perform for people. To control for the factors 'probability' and 'familiarity', the current experiment employed familiar medium-probability inference problems.

**Hypothesis 1:** Based on the findings from Experiment III, it is expected that the belief revision is resolved mainly by the mismatch principle with an underlying influence of task instruction.

**Hypothesis 2:** Belief revision is an easier performance when one is asked what to believe more instead of less, because the former is closer to daily use.

## 4.2. Experiment IV: Belief revision as a function of mental models and task instruction

### 4.2.1. Methods

#### 4.2.1.1. Participants

Eighty participants (63 female, 17 men) aged 18 to 48 ($M$=22.64, $SD$ = 3.71) participated in the current experiment. All participants were recruited from the University of Giessen. All had no basic knowledge of logic. They received a small monetary incentive (4 Euro) for taking part in the experiment.

#### 4.2.1.2. Design

A 2 (Inference problem: MP vs. MT) by 2 (Task Instruction: 'believe more' vs. 'believe less') between- within subject design was used with Task instruction as the between-groups factor.

#### 4.2.1.3. Materials

The materials for the current experiment consisted of 12 conditionals with a near 50% probability of occurrence taken from Evaluation studies I and II, performed by students from the same population (see Table 6). The conditionals did not significantly differ from one another with respect to probability, $\chi 2$ (11) = 8.751, $p$ = .645. Of these 12 conditional statements, both MP and MT problems were constructed for the experiment. The participants were randomly assigned to one of two conditions; one group ($N$ = 40) received the instruction 'Which of the first two statements do you believe more?' (Hereafter called the 'more' condition) and the other group ($N$ = 40) the instruction 'Which of the first two statements do you believe less?' (Henceforth the 'less' condition).

*Table 6*. List of conditionals used in Experiment IV

| *Medium-probability familiar conditionals* |
|---|
| If Hendrik is sick, then he goes to work |
| If Hannah stays on shore, then it is stormy weather |
| If Hugo is allergic to fur-bearing animals, then he gets fish |
| If Katharina cooks in the evening, then she prepares a rice dish |
| If Sophia is at home on the weekend, then she works in the garden |
| If Peter leaves his home, then he exists through the backdoor |
| If Karl goes to work, then he takes the car |
| If Bruno is hungry, then he eats a cracker |
| If Volker wants to listen to music, then he puts on classical music |
| If Kerstin visits a friend, then she brings flowers |
| If Christina is working in the office, then she leaves the door open |
| If Heiko goes outside, then he puts on his sunglasses |

## 4.2.1.4. Procedure.

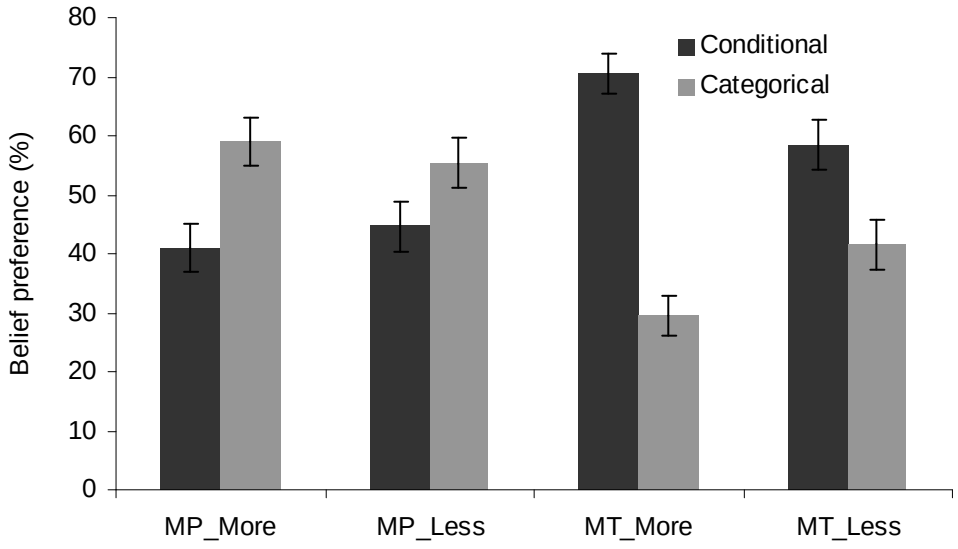
The procedure for the current experiment followed the general procedure as described for Experiment I-III with the exception that the task instruction was varied between two groups of participants (as outlined above).

## 4.2.2. Results

### 4.2.2.1. Belief revision choices

The overall belief in the conditional was 53.65 % (*SD* =20.56). Belief revision choices are depicted in Figure 12 as the mean preference for choosing the conditional or categorical statement. In the following, all effects are reported as significant at p<. 05. The revision choice data were submitted to a mixed between-within subjects ANOVA. This revealed a highly significant main effect of Inference problem, $F(1,78) = 41.073$, $MSE = 0{,}046$, $p < .0001$, reflecting again the higher endorsement of the conditional for the MT problems (64.48%) compared with the MP problems (42.81%). There was no main effect of Task instruction, $F(1, 78) = .905$, $MSE = 0.085$, $p = .344$; the conditional premise was believed to an equal extent in the 'more' condition (55.84%) as in the 'less' condition (51.46%). However, a significant Inference problem x Task instruction Interaction effect was obtained, $F(1, 78) = 5.484$, $MSE = 0{,}046$, $p = .022$. The interaction was further investigated using Independent Samples T-tests. With MT problems, the percentage of choosing to believe the conditional more was significantly higher in the 'more' condition ($M$ =70.62, $SD$ = 2.17) than in the 'less' condition ($M$ = 58.33, $SD$ = 2.70), $t(78) = 2.247$, $p$ = 027. Conversely, with MP problems the percentage of choosing to believe the conditional more was higher in the 'less' condition ($M$ = 44.58, $SD$ = 2.66) than in the 'more' condition ($M$ = 41.04, $SD$ = 2.65). However, this difference did not reach significance, $t(78) = 0.597$, $p = .553$.

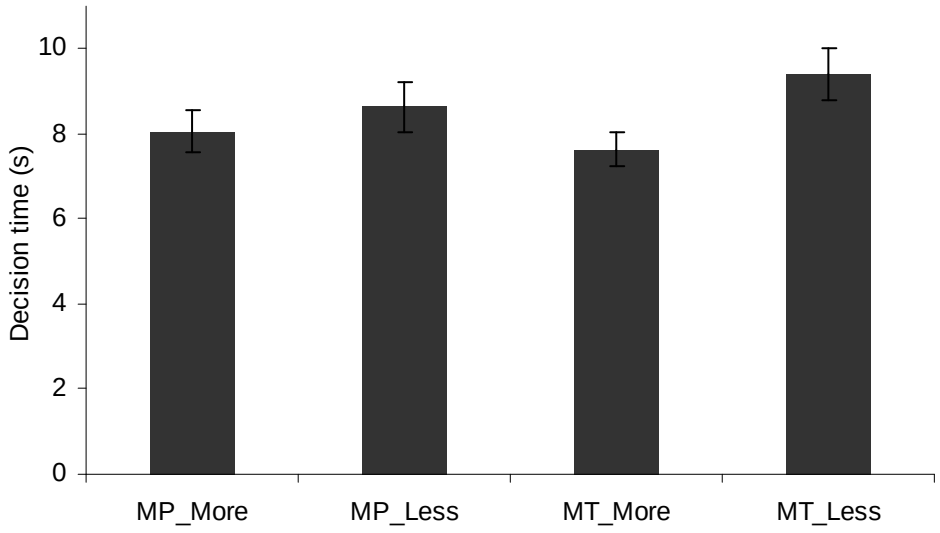
***Figure 12.*** **Mean percentages of belief preferences as a function of task instruction.**




## 4.2.2.2. Decision times

Figure 13 depicts the mean DTs for the four conditions. The analyses on the DT data did not elicit a main effect of Inference problem, $F(1, 78) = .517$, $MSE = 1.985$, $p = .474$; the DTs were similar for the MP ($M = 8.35$) and MT problems ($M = 8.51$). Also, no main effect appeared for Task instruction, $F(1, 78) = 2.686$, $MSE = 10.256$, $p = .105$. Being asked what to believe more resulted in comparable DTs ($M = 7.84$) as being asked what to believe less ($M = 9.02$). However, a significant Inference problem x Task instruction interaction emerged, $F(1, 78) = 6.850$, $MSE = 1.985$, $p = .011$. Post-hoc tests revealed that for the MT problems, significantly faster decision times were obtained in the 'more' ($M = 7.63$ s.) then in the 'less' condition ($M = 9.39$ s.), $t(78) = 2.384$, $p = .020$. With MP problems, participants also made faster decisions in the 'more' ($M = 8.05$ sec.) than in the 'less' condition ($M = 8.65$ sec.), however this difference did not achieve significance, $t(78) = .774$, $p = .441$.

***Figure 13.*** **Mean decision times in seconds as a function of task instruction.**




## 4.2.3. Discussion

In both task instruction conditions, the revision choice pattern resonates with the mismatch principle; the categorical was more often chosen with MP problems and the conditional more frequently with MT problems. The interaction effect further demonstrates that this effect was more robust in the 'more' condition than in the 'less' condition, where the belief revision choices hovered around 50%. This demonstrated that task instruction has its shared effect on belief revision. This supports the first hypothesis. A possible explanation for this might be that in real life one focuses more readily on what to believe more since that represents the statement one chooses to adopt; such thinking sets out a more straightforward cognitive path. The finding that in both task instruction conditions the participants applied the mismatch principle rules out the possibility that differences in task instruction explain the range of different belief revision responses in earlier research (Byrne & Walsh, 2005; Elio, 1997, Elio & Pelletier, 1997; Politzer & Carles, 2001).

Although the DTs in the 'less' condition were higher than those in the 'more' condition, this difference did not reach significance, which is not line with the second hypothesis.

However, the interaction effect can be explained in a similar fashion as was done for the DTs in earlier experiments. The fastest DTs when asked which statement one believes more with MT problems is most likely due to the fact that the question directs one's thinking toward the match. When asked which statement one believes less with MT problems, the thinking is diverted from the match and toward the mismatch which entails a more complex cognitive process. This additionally supports the first hypothesis. In short, the results from the Experiment IV suggest that directing people to choose the belief they favor instead of disfavor is a more solid approach.

# Chapter 5

# The effects of source trustworthiness and mental models on human belief revision

## 5.1. Introduction

Experiment VI tested the context factor 'task instruction' whereby the participants were asked to indicate which statement they believed more respectively less. Familiar inference problems with medium-probability were used. The results showed that in both conditions, the participants applied the mismatch principle, with task instruction modulating the strength of using mental models. Another promising context factor is the source of the utterance. Uncertainty can very well stem from the (social) context in which statements occur. Experiments V-VII examine to what extent the context factor 'source trustworthiness' affects the degree to which people revise their beliefs. A further advantage of manipulating the source of the utterance is that uncertainty is placed on both the conditional and categorical premise. In the former experiments and also in former belief revision research the manipulated factor was always imposed only on the conditional premise while the categorical premise stayed free of handling and in this way always conveyed a neutral statement. Completing the puzzle of human belief revision should include imposing the dependent factor on both the conditional and categorical statement. One way of doing so is manipulating the source of the utterances.

**Hypothesis 1:** Both source trustworthiness and mental models will affect belief revision. Source trustworthiness will have its effect as evidenced by its robust influence on persuasion and attitude. That is, belief in the conditional will be higher than in the categorical when it is coupled with a high trustworthy source and lower when it is coupled with a low trustworthy source. The appliance of the mismatch principle will be mirrored by a higher belief in the conditional for MT problems than for MP problems.

**Hypothesis 2:** DTs will be faster when the level of source trustworthiness is easily discriminable.

**Hypothesis 3:** A joint operation of belief revision strategies will appear. That is, beside a specific revision strategy guided by source trustworthiness, there would

be an influence of the more global strategy of mental models. We postulate that these strategies, although regarded as independent belief revision strategies, have a mutual engagement. We expect this to be evident in both the belief revision responses as well as the DTs.

## 5.2. Evaluation study III

A norming study was done to pre-test the trustworthiness of 96 occupations.

### 5.2.1. Methods

#### 5.2.1.1. Participants

The participants consisted of twenty students from different faculties of the University of Giessen (12 females, 8 males; aged 19 to 25 ($M$ = 22.47, $SD$ =1.61).

#### 5.2.1.2. Materials and Procedure

Booklets were created with general instructions on the front page and each of 96 occupations presented on a separate page. The instruction on the first page read as follows (translated from German):

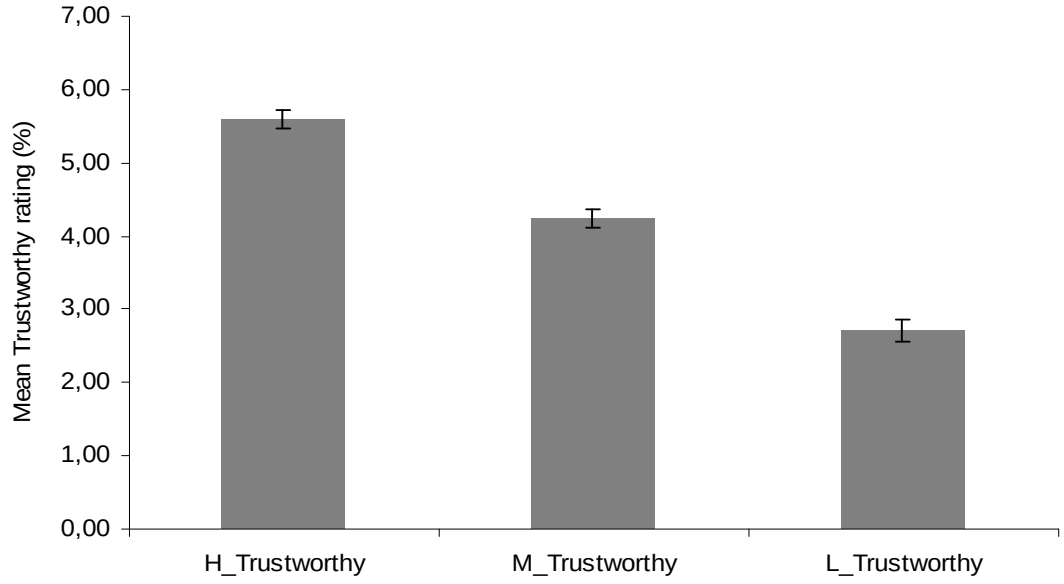
"*An occupation is presented on each of the following pages. A rating scale is presented under each occupation. On this scale, please rate how trustworthy you find a person with such an occupation. 1 means "very untrustworthy" and 7 means "very trustworthy". Please rate the occupations in the order they are being presented to you. Please do not go back to previous occupations. Many occupations are presented in male as well as female form. In case you find them differently trustworthy, simply indicate that by giving two different trustworthy ratings."*

The subjects rated each occupation on a 7-point Likert scale from 1 (*very untrustworthy*) to 7 (*very trustworthy*). Eight high judged-, eight low judged-, and 16 medium-judged trustworthy occupations were selected for the experiment.

## 5.2.2. Results

Wilcoxon Signed Ranks Tests showed that the high-trustworthy occupations were rated more trustworthy than the medium-trustworthy and the low-trustworthy occupations to a significantly extent, $Z = 3.920$, $p < .0001$, and $Z = 3.925$, $p < .0001$, respectively. In return, the medium-trustworthy occupations were rated significantly more trustworthy than the low-trustworthy ones, $Z = 3.921$, $p < .0001$.

***Figure 13.*** **Mean trustworthy rating for the three groups of items.**



## 5.3. Experiment V: The influence of an occupation's trustworthiness on belief revision

### 5.3.1. Methods

#### 5.3.1.1. Participants

Thirty participants from different faculties from the University of Giessen took part in the experiment (20 females, 10 males) aged 19 to 38 in exchange for a small monetary incentive (4 Euro). None of the participants had pre-knowledge of deductive reasoning and logic (this applies as well to Experiments VI and VII).

#### 5.3.1.2. Design

The experiment applied a 2 (Inference problem: MP vs. MT) x 3 (Source trustworthiness: High vs. Medium vs. Low) within-subjects design. The dependent variables were the percentage of choosing the conditional statement and the decision time (DT).

#### 5.3.1.3. Materials

The conditional statement always contained the abstract content "If A is the case, then B is the case". From the eight high- and eight low-trustworthy jobs, eight job matches were constructed whereby each high- and low-trustworthy job of each match preceded either the conditional or the categorical statement in a counterbalanced manner. Thus, this resulted in eight high/low and eight low/high trustworthy inference problems. In a similar fashion, eight job matches were formed from the 16 medium-trustworthy jobs (see Table 7 for an overview of the job matches; in each row the job on the left is matched with the job on the right). From each of these matches, both MP and MT problems were created. The abbreviation 'high/low' refers to the case where the conditional was preceded by a high-trustworthy job and the categorical by a low-trustworthy job. Similarly, the abbreviation 'low/high' signals the reversed order. And finally, medium/medium refers to the case whereby both premises are preceded by a medium-

trustworthy job, for each match. In total then, 48 inference problems were used for the experiment. An example of a high/low MP problem and a medium/medium MT problem are shown below.

*Spy: If A is the case, the B is the case*

*Judge: A is the case*

*Fact: B is not the case*

*Accountant: If A is the case, then B is the case*

*Businessman: B is not the case*

*Fact: A is the case*

***Table 7.*** **List of job matches used in Experiment V (with reported mean trustworthy ratings).**

| *High-trustworthy job* | *Low-trustworthy job* |
|---|---|
| Police officer (*M*=5.25) | Used car salesman (*M*= 2.00) |
| Judge (*M*=5.65) | Spy (*M*= 2.60) |
| Pilot (*M*= 5.75) | Salesperson (*M*= 3.30) |
| Veterinarian (*M*=5.65) | Real-estate agent (*M*= 2.79) |
| Educator (*M*=5.40) | Minister (*M*= 3.00) |
| Firefighter (*M*=5.78) | Insurance agent (*M*= 2.60) |
| Development aid worker (*M*=5.85) | Member of the Bundestag (German term for Congress) (*M*= 2.70) |
| Psychologist (*M*=5.45) | Stockbroker (*M*= 2.75) |
| *Medium-trustworthy job* | *Medium-trustworthy job* |
| Bookkeeper (*M*= 4.25) | Banker (*M*= 4.43) |
| Photographer (*M*= 4.60) | Hairdresser (*M*= 4.40) |
| Businessman (*M*= 3.70) | Accountant (*M*= 3.75) |
| Radio presenter (*M*= 4.25) | Travel agent (*M*= 4.35) |
| Caretaker (*M*= 4.75) | Student (*M*= 4.63) |
| Film producer (*M*=4.20) | Lawyer (*M*= 4.05) |
| Consultant (*M*= 3.8) | Salesperson (*M*= 3.75) |
| Hunter (*M*= 4.25) | Butcher (*M*= 4.65) |

### 5.3.1.4. Procedure

Participants were seated in a quiet laboratory room in front of a computer. They received verbal and written instructions. They were explained that they would repeatedly be presented with sets of three statements, one at a time, and that the first two statements were uttered by a person represented by his or her occupation and that the third statement would be presented as a fact. It was further stressed that the truth of the first two statements was uncertain but that the third statement, causing an inconsistency, was certainly true. Their task was then to resolve this inconsistency by choosing which of the first two statements they believed more. Four practice items preceded the actual experiment. The statements within each problem set were presented in successive order in the middle of a black screen. The participant could read the statements in a self-paced manner. The conditional and the categorical statement appeared in white font colour. The third statement appeared in red to signify its counterfactual nature and to aware the subjects they had to make a belief revision choice. The participant could read the statements in a self-paced manner and press a space-bar like button on a response box to run through the experiment. After the third statement, the participants made their choice by pressing a left or right button. The designation of the left and right button to the conditional and statements was counterbalanced across subjects. Furthermore, all 48 problem sets were randomized within subjects. The participants had the possibility to take a short break between the presentations of the problem sets.

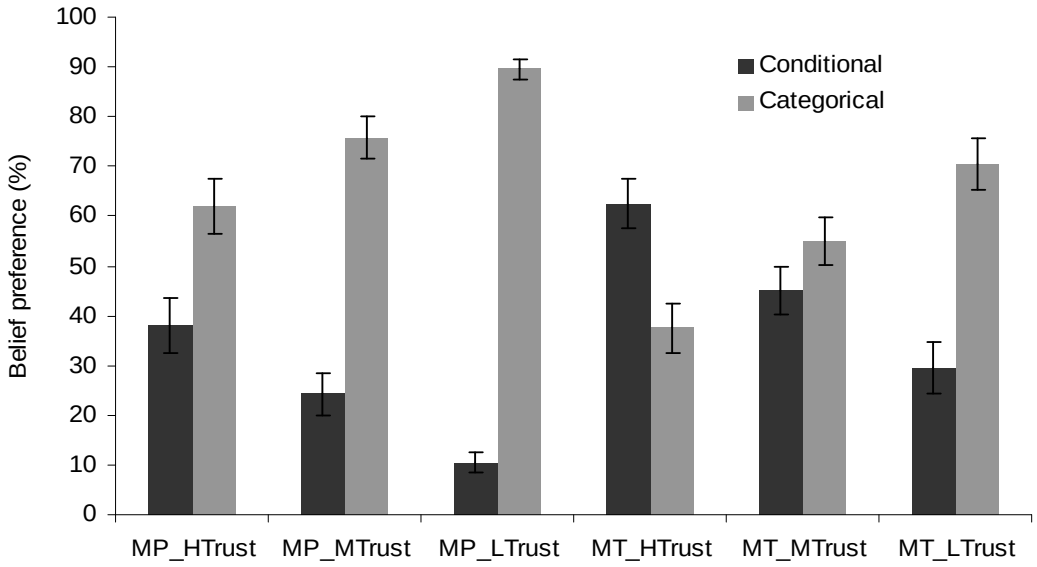
## 5.3.2. Results

### 5.3.2.1. Belief revision choices

The overall mean to believe the conditional more was 34.93% (*SD* = 14.95). Belief revision choices in the six conditions are depicted in Figure 14 as the mean preference for choosing the conditional or the categorical statement. In the following, all effects are reported as significant at $p < .05$. An ANOVA was performed with belief preference as

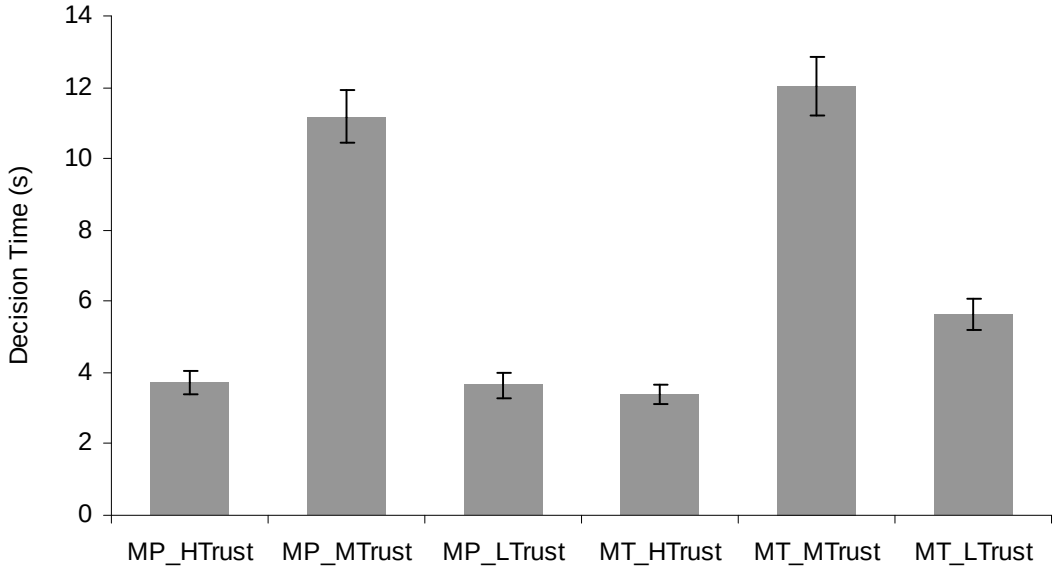
the dependent variable and Inference problem and Source trustworthiness as the independent variables. This revealed significant main effects of Inference problem, $F(1, 29) = 21.701$, $p < .0001$, and Source trustworthiness, $F(1.585, 45.956) = 24.708$, $p < .0001$. The conditional was believed more with MT ($M = 45.69$) than with MP problems ($M = 24.17$). With respect to Source trustworthiness, follow-up tests revealed that all three comparisons were significant: High/low-trustworthy ($M = 50.21$) vs. medium/medium-trustworthy ($M = 34.58$) condition, $t(29) = 3.717$, $p = .001$; High/low-trustworthy vs. low/high-trustworthy ($M = 20.00$) condition, $t(29) = 5.810$, $p < .0001$; Medium/medium-trustworthy vs. low/high-trustworthy condition, $t(29) = 4.455$, $p < .0001$. The Inference problem x Source trustworthiness interaction did not reach significance, $F(2, 58) < 1$.

***Figure 14***. **Mean percentages of belief preferences as a function of source trustworthiness.**




**Notes: MP_HTrust, MP_MTrust, MP_LTurst = modus ponens high-trustworthy, medium-trustworthy, and low-trustworthy problems, respectively. MT_HTrust, MT_MTrust, MT_LTurst = modus tollens high-trustworthy, medium-trustworthy, and low-trustworthy problems, respectively.**

***Figure 15***. **Mean decision times of belief preferences as a function of source trustworthiness.**




### 5.3.2.2. Decision times

The overall DT was 6.54 s. (*SD* = 1.80). Figure 15 depicts the mean DTs for the six conditions. A 2 (Inference problem: MP vs. MT) x 3 (Trustworthiness: High vs. Medium vs. Low) repeated measures ANOVA was carried out on the factor decision time. This revealed significant differences due to Inference problem, $F(1,29) = 27.330$, $p < .0001$ and Source trustworthiness, $F(1.136, 32.952) = 214.963$, $p < .0001$. The DTs were significantly higher for MT (*M* = 7.09) than for MP problems (*M* = 5.99). As for Source trustworthiness, all three comparisons were significant: High/low-trustworthy (*M* = 3.52) vs. medium/medium-trustworthy condition (*M* = 11.60), $t(29) = 16.246$, $p < .0001$; High/low-trustworthy vs. low/high-trustworthy condition (*M* = 4.52), $t(29) = 6.495$, $p < .0001$; Medium/medium-trustworthy vs. low/high-trustworthy condition, $t(29) = 13.613$, $p < .0001$. The interaction between Inference problem and Source trustworthiness was also significant, $F(1.457, 42.245) = 12.725$, $MSE = 2.393$, $p < .0001$. With MP problems, the DTs were higher when the conditional was coupled with a high-trustworthy occupation (*M* = 3.57) than with a low-trustworthy occupation (*M* = 3.36) but not to a significant extent, $t(29) = 1.190$, $p = .244$. However, the DTs of these two

aforementioned conditions were significantly lower than the DTs of the medium/medium-trustworthy condition ($M$ = 11.04), $t(29)$ = 14.031, $p$<.0001 and $t(29)$ = 12.529, $p$<.0001. In contrast, with MT problems, the DTs were significantly lower when the conditional was coupled with a high-trustworthy occupation ($M$ = 3.47) than with a low-trustworthy occupation ($M$ = 5.69), $t(29)$ = 8.877, $p$ <.0001. Also, the DTs in both high/low- and low/high-trustworthy conditions were significantly lower than in the medium/medium-trustworthy condition ($M$ = 12.14), $t(29)$ = 14.828, $p$<.0001 and $t(29)$ = 11.208, $p$<.0001.

### 5.3.3. Discussion

The belief revision results did not confirm the first hypothesis. Relying on source trustworthiness did not lead to a distinctive belief revision strategy; it was not the case that the conditional was believed more than the categorical when it was preceded by a high-trustworthy source and believed less than the categorical when preceded by a low-trustworthy source. However, source trustworthiness did have an overall effect in that the conditional was believed most with a high-trustworthy source, least with a low-trustworthy source, and at an intermediate level wih a medium-trustworthy source. Belief in the conditional was higher for MT problems than for MP problems, which attest to the mismatch principle. In line with the second prediction, the DTs showed that a clear contrast in levels of source trustworthiness eased performance; the DTs were clearly faster when the first two premises were preceded by sources of two contrasting trustworthiness each at the opposite ends of the continuum of trust than when both premises were uttered by persons with medium trustworthiness. Our third prediction was also confirmed in that both source trustworthiness and mental models influenced belief revision. The conditional gained more belief with MT than with MP problems, which attests to the prediction of the mismatch principle. For the belief revision choices, an interaction effect between Inference problem and Source trustworthiness was missing; the change in belief for the conditional from high-to low trustworthy source had a similar pattern for MP and MT problems. However, as can be seen in Figure 13, for MP inference problems the belief preference for the conditional remained low; in both conditions the preferred belief was the categorical statement. In contrast, for MT

problems the preferred belief was the conditional when it was preceded by a high-trustworthy source but this preference switched to the categorical when the conditional was preceded by a low-trustworthy source. This trend did not show in the analysis since it was performed only on the conditional statement. A change in design might reveal whether it is really the case that source trustworthiness seems to function as a belief revision strategy for the MT problems and as a modulating factor for the MP problems where applying mental models overruled. The findings of DTs, however, did demonstrate the interactive working of the two different belief revision strategies. Although, overall the DTs in the high/low- and low/high-trustworthy conditions were faster than in the medium/medium-trustworthy condition, a significant difference between the high/low- and low/high-trustworthy conditions only appeared for the MT problems, and not for their MP counterparts, as the interaction showed. We believe this is due to the occurrence of a match within the MT problems, that is between the contradicting fact's model 'p' and the first part of the first premise's model 'if p then q'. When in addition the conditional's source is highly trustworthy, than the two belief revision strategies work in harmony, easing the performance. No such clear match is evident within the MP problems. Thus, although the belief revision responses and DTs showed the working of two different belief revision strategies, only the DTs showed a confluence of these two strategies.

## 5.4. Evaluation study IV

Experiment V showed a clear effect of both source trustworthiness and mental model on belief revision. However, even though our aim of Experiment V was to investigate whether there is any effect of trustworthiness, using content-rich statements would more resemble real-life situations. In Experiment VI, statements were included that are likely to be uttered in a conversation between two persons that represent the two jobs of each match. A norming study was first conducted to create the material, which will be outlined here.

### 5.4.1. Methods

#### 5.4.1.1. Participants

Twenty participants from different faculties of the University of Giessen (12 females, 8 males; aged 20 to 29 ($M$ = 23.85, $SD$ = .58) took part in the norming study.

#### 5.4.1.2. Materials and Procedure

Four high-low/low-high and four medium-medium trustworthy job matches were taken from the pool of job matches used in Experiment V. For each match, six conditional statements were constructed. In addition, each of the six conditionals was preceded by each of the two jobs of each match. The resulting 96 conditionals were used for a norming study in which booklets were used with each statement appearing on a separate page, in random order. The participants' task was to rate under each statement on a rating scale ranging from 0% (*very unlikely*) to 100% (*very likely*) how likely they thought the statement was true. From each job match, the conditional that was rated around 50% with either job was taken for the norming study. That is, the one conditional from each of the eight job matches was taken twice, each one preceded by one of the two jobs. The instruction on the front page of the booklet read as follows (translated from German):

"*On each of the following pages you will be presented with a statement, which is uttered by a person presented by his or her occupation. A rating scale is presented under each statement. On this scale, please rate how possible it is that this person is speaking the truth. 0% means "very unlikely" and 100% means "very likely". Please rate the statements in the order they are being presented to you. Please do not go back to previous statements*."

### 5.4.2. Results

A Friedman Test showed that all 16 conditionals did not significantly differ from one another with respect to probability, $\chi$ = 16.663, $p$ = .339. Wilcoxon Signed Rank tests were performed to also ensure that the conditionals per match were of equal probability to be true. This gave the following results (conditionals are presented by the job they are preceded by): police man ($M$ = 53.00) vs. used car salesman ($M$ = 51), $Z$ = .401, $p$ =.688; Judge ($M$ = 55.79) vs. spy ($M$ = 59.00), $Z$ = .473, $p$=.636; Pilot ($M$ = 52.50) vs. salesperson ($M$ = 52.50), $Z$ = .058, $p$ = .954; Firefighter ($M$ = 44.00) vs. insurance agent ($M$ = 45.00), $Z$ =.247, $p$ = .804; Bookkeeper ($M$ = 43.68) vs. banker ($M$ = 49.00), $Z$ = 1.078, $p$ = .281; businessman ($M$ = 47.00) vs. accountant ($M$ = 49.00), $Z$ = .136, $p$ = .892; Film producer ($M$ = 55.00) vs. lawyer ($M$ = 52.00), $Z$ = .564, $p$ = .572; Hunter ($M$ = 60.00) vs. butcher ($M$ = 59.00), $Z$ = .434, $p$ = .664.

## 5.5. Experiment VI

In Experiment VI, again occupations were used to manipulate source trustworthiness. However, the content of the problems was meaningful and referred to a statement in a likely conversation between two sources.

**Hypothesis 1:** The first expectation was that source trustworthiness would affect belief revision more pronounced than in Experiment V, since the premises

contain meaningful information that are linked to the occupations. The conditional will be believed more than the categorical when coupled with a high-trustworthy source and believed less than the categorical when preceded by a low-trustworthy source. The appliance of the mismatch principle will be mirrored by a higher belief in the conditional for MT problems than for MP problems.

Hypothesis 2: DTs will be faster when the level of source trustworthiness was easily discriminable, as was found in Experiment V.

Hypothesis 3: it was expected to find again a joint operation of belief revision strategies. That is, beside a specific revision strategy guided by source trustworthiness we expected again the influence of the more global strategy of mental models. We postulate that these strategies, although regarded as independent belief revision strategies, to have a mutual engagement. We expect this to be evident in both the belief revision responses as well as the DTs.

## 5.5.1. Methods

### 5.5.1.1. Participants

Thirty-six participants from different faculties from the University of Giessen took part in the experiment (24 females, 12 males) aged 19 to 24 ($M$ = 23.19, $SD$ = 3.42) in exchange for a small monetary incentive (4 Euro).

### 5.5.1.2. Design

The experiment applied a 2 (Inference problem: MP vs. MT) x 3 (Source trustworthiness: High vs. Medium vs. Low) within-subjects design. The dependent variables were the percentage of choosing the conditional statement and decision time (DT).

### 5.5.1.3. Materials

For the experiment, MP and MT problems were created whereby the precedence of the conditional and categorical by a high- or low trustworthy job was counterbalanced. The medium-trustworthy job matches were similarly counterbalanced across conditional and categorical statement (medium1-medium2 and medium2-medium1 serve as abbreviations for the counterbalancing of the medium-trustworthy job matches). This resulted in 32 inference problems: four MP high/low and four MP low/high problems, four MT high/low and 4 MT low/high problems, four MP medium1/medium2 and four MP medium2/medium1 problems, and four MT medium1/medium2 and four MT medium2/medium1 problems. The design was identical to that of Experiment V. Table 8 lists the conditionals used with the accompanying job matches. Here follow 2 examples:

High/low trustworthy MP problem:

*Police officer: If Karl sells a damaged car, then he offers the car for a reduced price.*

*Used car salesman: Karl sells a damaged car.*

*Fact: Karl does not offer the car for a reduced price*

Medium1/medium2 trustworthy MP problem:

*Film producer: If Nils hires an actor for a film, then he pays the actor his share of the profit.*

*Lawyer: Nils does not pay the actor his share of the profit.*

*Fact: Nils hires an actor for a film*

*Table 8*. List of conditionals with their trustworthy job matches used in Experiment 2.

| *Conditional* | *Job matches HL/LH* |
|---|---|
| If Karl sells a damaged car, then he offers the car for a low prize | Policeman vs. Used car salesman |
| If Sophia enters a company, then she wears an access badge | Judge vs. spy |
| If Anne books a flight, then she chooses the flight company with the best reputation | Pilot vs. Salesperson |
| If Lukas owns a fire extinguisher, then he also has a smoke detector | Fire-fighter vs. Insurance agent |
| *Conditional* | *Job matches M1M2/M2M1* |
| If Laura buys an item, then she pays with credit card | Bookkeeper vs. Banker |
| If Paul seeks financial help, then he consults the yellow pages | Businessman vs. accountant |
| If Nils hires an actor for a movie, then he pays the actor his share of the profit | Film producer vs. lawyer |
| If Nicole owns a farm, then she gives her animals special diets | Hunter vs. butcher |

### 5.5.1.4. Procedure

The procedure was identical to that of Experiment V.

## 5.5.2. Results

### 5.5.2.1. Belief revision choices

Paired-samples t-tests were performed on the medium-trustworthy inference problems. For both MP and MT medium- trustworthy problems it made no difference which occupation of each match preceded the conditional or the categorical, $t$ (35) <1 and $t$(33) <1, respectively. For this reason, we collapsed over medium1/medium2 and medium2/medium1 inference problems and combined them, for both MP and MT problems. The overall preference for believing the conditional statement more was 42.27 % ($SD$ = 17.32). Mean percentages of belief revision choices in the six conditions are depicted in Figure 16. A repeated measures ANOVA revealed a main effect of Inference problem, $F$(1, 35) = 101.686, $MSE$= .087, $p$ <.0001, such that the conditional was endorsed significantly more with MT ($M$ = 62.62) than with MP problems ($M$ = 22.11). No main effect of Source trustworthiness emerged, $F$(2,70) <1. Also, no Inference problem x Source trustworthiness interaction was detected, $F$(2, 70) = 1.210, $p$ = .304.

*Figure 16.* Mean percentages of belief preferences as a function of source trustworthiness in Experiment 2.

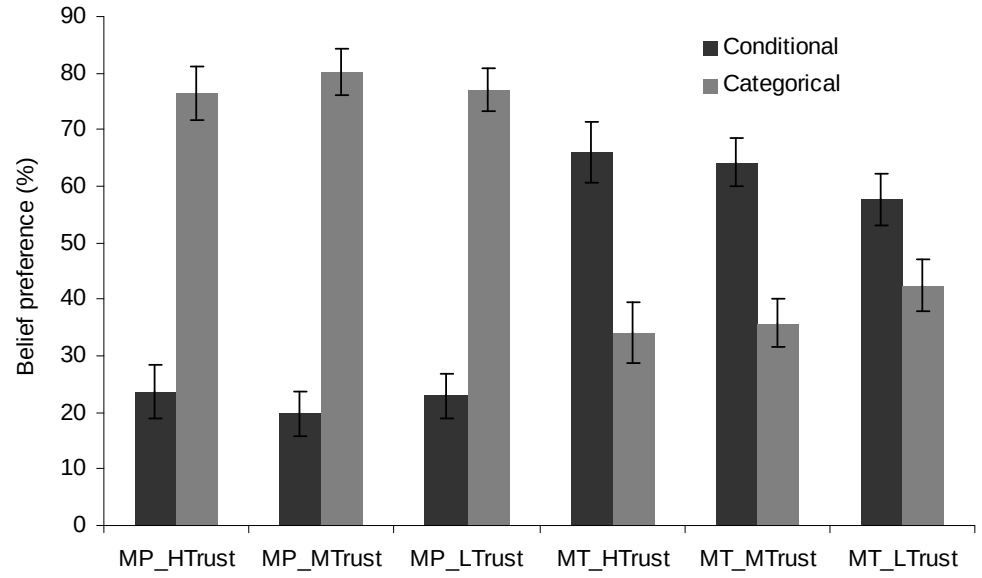
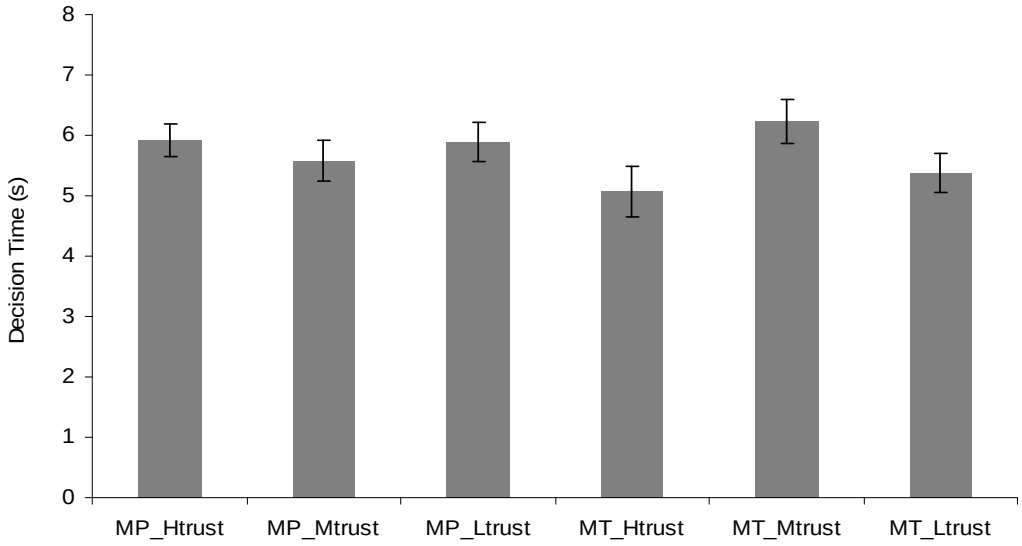



*Figure 17.* Mean decision times in seconds as a function of source trustworthiness in Experiment 2.



### 5.5.2.2. Decision times

Paired-sampled T-tests were performed on the two MP and MT medium-trustworthy problem conditions. For both MP and MT problems it made no significant difference which of the two medium-trustworthy occupations preceded which statement, $t(35) = 1.998$, $p = .056$ and $t(35) = .342$, $p = .734$, respectively. Therefore, the two conditions of each inference problem type were collapsed and from hereon referred to as MP medium-trustworthy and MT medium-trustworthy problems. The overall DT was 5.74 s. ($SD = 1.62$). Figure 17 shows the mean decision times for the six conditions. A repeated measures ANOVA showed no main effects of Inference problem, $F(1,35) = 1.456$, $MSE = 2.081$, $p = .236$, and Source trustworthiness, $F(2, 70) = 1.295$, $MSE = 2.441$, $p = .280$. However, an Inference problem x Source trustworthiness interaction emerged, $F(1.589, 55.627) = 6.118$, $MSE = 1.817$, $p = .007$. For MT problems, the DTs were significantly lower when the conditional was coupled with a high-trustworthy ($M = 5.08$) or a low-trustworthy occupation ($M = 5.37$) than with a medium one ($M = 6.23$), $t(35) = 3.370$, $p = .002$, and $t(35) = 2.239$, $p = .032$, respectively. In contrast to Experiment V, there was no significant difference when the conditional was coupled with a high trustworthy occupation rather than a low one, $t(35) > 1$. For MP problems, the DTs were the same regardless of whether the conditional was stated by a person with a high, low, or medium trustworthy occupation, MP high-trustworthy ($M = 5.92$) versus MP low-trustworthy ($M = 5.89$), $t(35) > 1$, MP high-trustworthy versus MP medium-trustworthy ($M = 5.58$), $t(35) = 1.082$, $p = .287$, and MP low-trustworthy versus MP medium-trustworthy, $t(35) = 1.184$, $p = .244$.

## 5.5.3. Discussion

The results of Experiment VI partly supported the first hypothesis and third hypothesis. The effect of source trustworthiness was not evident in the belief revision choices; it actually lost its effect when the premises contained meaningful content that could relate to a possible conversation. Instead, belief revision was solely a factor of

mental models, applied to both MP and MT problems. Only the DTs revealed a slight influence of source trustworthiness. The interaction effect found with the DTs did show that with MT inference problems, the effect of source trustworthiness led to faster DTs when the statements were presented by contrasting levels of the sources' s trust then when both statement were uttered by sources of equal trustworthiness. This also partly supports the second hypothesis.

A factor that could have 'occluded' the effect of source trustworthiness in the current experiment is expertise. Even though the high-trustworthy jobs of the high/low-low/high matches were rated significantly more trustworthy than their counterparts as was measured in the norming study described in the Experiment V section, the statements with which the jobs were combined could easily plausible be uttered by someone engaged in either jobs based on expertise. It is acknowledged that the set-up of the norming study for the current experiment was actually aimed at equal plausibility but it was expected that when both jobs would occur together as in the inference problems, and not separately as in the booklets, people would then compare the two jobs and prefer the statement, be it conditional or categorical, that is preceded by the more trustworthy job. Research in the domain of social psychology and social cognition has already shown that expertise has a stronger influence on people's decision making and persuasion than trustworthiness (Wiener & Mowen, 1986). A more recent study by Stevenson and Over (2001) already demonstrated that expertise of the source guides reasoning performance. Therefore, expertise seemed to have dominated trustworthiness in the current case. To control for expertise in future studies, the booklets might better include rating scales for both trustworthiness and expertise.

However, as stated in the introduction, the aim of the current investigation was to unravel whether the credibility factors 'source trustworthiness' plays a role in belief revision. Therefore, in Experiment VII another established social factor known to influence person perception and trustworthiness was investigated, namely personality description.

## 5.6. Experiment VII: Source Trustworthiness reflected in personality description: It's influence on belief revision

In Experiment VII, the statements were continuously uttered by two different people whose personality descriptions were given at the beginning of the experiment. One person was presented with a very trustworthy personality description and the other person with a very untrustworthy personality description. For the premises, statements were used that described ordinary daily life activities that have been rated to occur with a near 50% probability in Evaluation studies I and II.

Hypothesis 1: Personality description as the trustworthiness factor would have a similar influence on belief revision as job occupation had in Experiment V. Appliance of the mismatch principle would show itself by a greater belief in the conditional for MT problems than for MP problems.

Hypothesis 2: A confluence will appear of reliance on source trustworthiness and appliance of the mismatch in belief revision.

Hypothesis 3: The DTs will reveal which problem type is easiest to perform

### 5.6.1. Methods

#### 5.6.1.1. Participant

A total of 30 participants (21 females, 9 males) aged 19 to 30 ($M$ = 23.10, $SD$ = 3.08) from different faculties from the University of Giessen took part in the experiment in exchange for a small monetary incentive (6 Euro).

### 5.6.1.2. Design

The experiment utilized a 2 (Inference problem: MP vs. MT) x 2 (Trustworthiness of source: trustworthy vs. untrustworthy) within-subject design. The dependent variables were again belief preference and DTs.

### 5.6.1.3. Materials

Source trustworthiness was manipulated by presenting to the subject prior to the start of the experiment a person description of a trustworthy person, named Christian, and a person description of an untrustworthy person, named Mark, as follows.

> *Christian is known as a sincere, honest, reliable, responsible, and caring person. Christian is 45 years old and has had the same job for 20 years. He has been married now for 15 years. Christian is described as having a big heart and someone whom you can always rely on.*

> *Mark is known as a reserved, dishonest, unreliable, and uncaring person without any feeling of responsibility. Mark is 45 years old and has been fired from several jobs due to fraudulent activities. Mark's wife divorced him after five years of marriage, because he had repeatedly cheated on her. Mark is described as selfish and someone you cannot rely on.*

Six conditional statements were used representing ordinary daily-life actions that were rated around 50% probability in Evaluation studies I and II. A Friedman Test showed that these six conditionals did not differ from each other with respect to probability, $\chi$ = 5.397, $p$ = .369. A total of twenty-four inference problems were presented consisting of the following four conditions: Six high/low and six low/high MP problems, and six high/low and six low/high MT problems. High/low indicates that the conditional was

uttered by Christian and the categorical was uttered by Mark. Low/high means that the conditional was uttered by Mark and the categorical by Christian. The experiment consisted of six blocks, each containing four inference problems, one problem of each condition. Table 9 lists the conditional statements used for the experiment. Below follow examples of a low/high MP and a high/low MT problem:

*Mark: If Katharina cooks in the evening, then she prepares a rice dish*
*Christian: Katharina cooks in the evening*
*Fact: Katharina does not prepare a rice dish*

*Christian: If Karl goes to work, then he takes the car*
*Mark: Karl does not take the car*
*Fact: Karl goes to work*

*Table 9*. List of conditionals used in Experiment VII

| *Conditional* |
|---|
| If Katharina cooks in the evening, then she prepares a rice dish |
| If Hugo is allergic to fur-bearing animals, then he gets a fish |
| If Bruno is hungry, then he eats a cracker |
| If Hendrik is sick, then he will go to work |
| If Peter leaves his house, then he exists through the backdoor |
| If Karl goes to work, then he takes the car |

*Table 10*. **List of additional information concerning the sources used in Experiment VII.**

| Each of the following additional pairs of statements was preceded by the commentary: 'Now you receive the additional information' |
|---|
| Christian finds it important to respect norms and values |
| Mark plays people against each other for his own gain |
| Christian has received a medal for community service |
| Mark received a court summons due to violation of privacy |
| Christian is always eager to donate money to charity |
| Mark gambles money that he receives from friends |
| Christian tells his honest opinion without insulting anyone |
| Mark likes to lie to people for personal gain |
| Christian does his assignments at work very conscientiously, while |
| Mark often does his work assignments negligible |
| Christian commits to appointments and promises, while |
| Mark often cancels appointments and makes false promises |

The order of the problems was randomized per subject. To assure that the subjects kept the statements' sources in mind, additional information was offered concerning the trustworthiness of these two persons after each 4 problems (see Table 10). The order of these six pairs of statements was fixed across participants.

### 5.6.1.4. Procedure

The procedure was identical to that of Experiment V and VI except that at the end of the experiment the participants were asked to rate the trustworthiness of both Christian and Mark on a 7-point Likert scale.

## 5.6.2. Results

### 5.6.2.1. Belief revision choices

The overall preference for the conditional was 41.25% (*SD* = 14.16). The mean belief preferences per condition are shown in Figure 18. An ANOVA was performed with belief preference as the dependent variable and Inference problem and Source trustworthiness as independent variables. This 2 x 2 analysis showed a main effect of Inference Problem, $F$ (1, 29) = 34.701, *MSE* = .079, $p$ <.0001 such that the conditional statement was endorsed more with MT ($M$ = 56.39) than with MP problems ($M$ = 26.11). A main effect also showed up for source trustworthiness, $F$ (1, 29) = 29.034, *MSE* = .055, $p$ <.0001, in that the conditional was believed more when it was uttered by the high-trustworthy source ($M$ = 52.78) than by the low-trustworthy source ($M$ = 29.72). The Inference problem x Trustworthiness interaction also led to significance, $F$(1, 29) = 5.477, *MSE* = .022, $p$ = .026. With both MP and MT problems, belief in the conditional was significantly greater when it was preceded by a high-trustworthy than a low-trustworthy source. However, this difference was much greater with MT ($M$ = 71.11 vs. $M$ = 41.67; $t$(29)=5.227, $p$<.0001) than with MP problems ($M$ = 34.44 vs. $M$ = 17.78; $t$(29) = 3.746, $p$ = .001).

A paired-samples t-test was performed to check whether the two persons uttering the statements were also actually viewed as less or more trustworthy. This was the case; the high-trustworthy source (i.e. Christian) had a significantly higher mean

trustworthy rating of ($M$ = 5.65, $SD$ = 1.09) than the low-trustworthy source (i.e. Mark; $M$ = 2.60, $SD$ = 1.07), $t(29) = 8.311$, $p < .0001$. This shows that the overall higher belief in the conditional when it was preceded by a high- rather than low-trustworthy source is indeed attributable to the effect of source trustworthiness.

***Figure 18.*** **Mean percentages of belief preferences as a function of source trustworthiness in Experiment VII.**

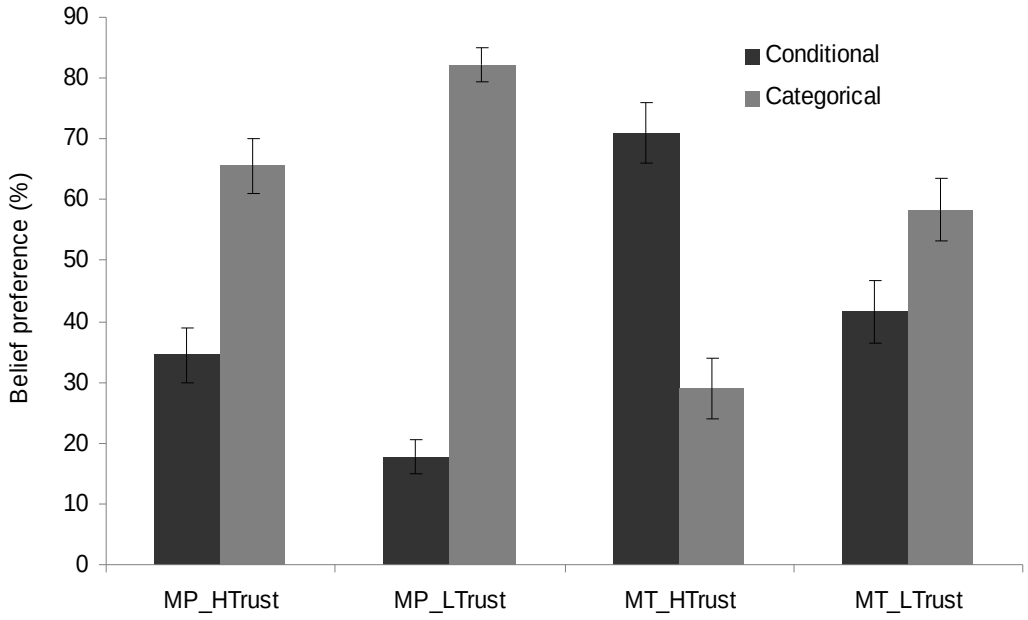
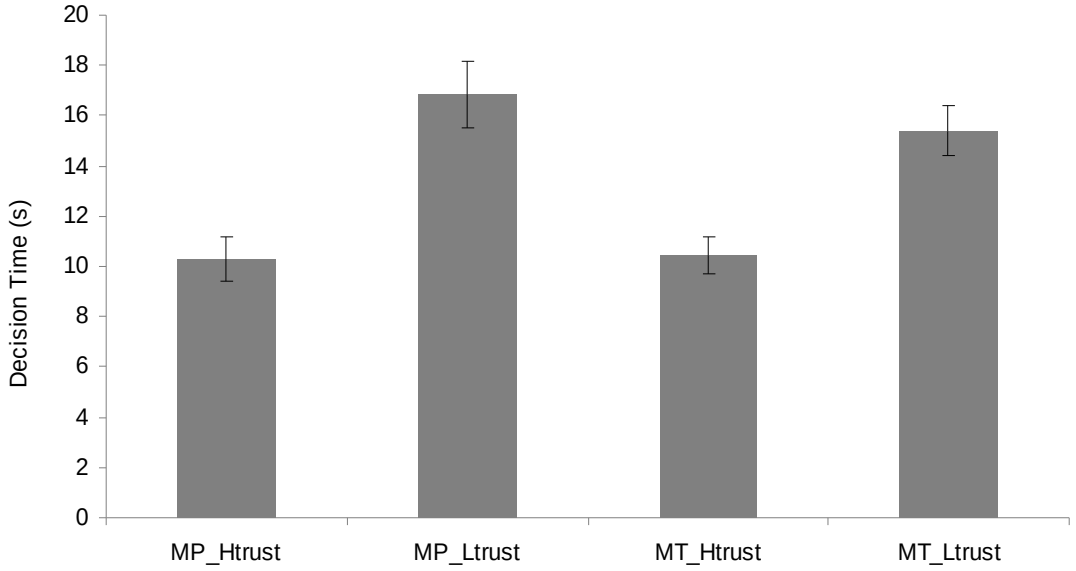


***Figure 19.*** **Mean decision times in seconds as a function of source trustworthiness in Experiment VII.**




## 5.6.2.2. Decision times

The overall DT was 13.23 s. (*SD* = 4.78). Figure 19 depicts the mean DTs for the 4 conditions. A 2 x 2 analysis of variance yielded one significant effect, that of Source trustworthiness, $F(1,29) = 62.503$, $MSE = 15.789$, $p < .0001$. Deciding which statement to believe more took longer when the conditional was uttered by the low-trustworthy ($M$ = 16.10) than by the high- trustworthy source ($M$ = 10.36). The effect of Inference problem and the Inference problem x Source trustworthiness interaction effect failed to reach significance, $F(1,29) = 1.783$, $MSE = 6.589$, $p = .192$ and $F(1,29) = 2.528$, $MSE = 7.743$, $p = .123$, respectively.

### 5.6.3. Discussion

The belief revision results from Experiment VII are similar to those of Experiment V. That is, trustworthiness had an influence on belief revision such that it profoundly lowered the belief in the conditional when this premise was coupled with a person of low rather than high trustworthiness. In line with hypothesis 3, the DTs also mirrored the influence of source trustworthiness in that belief preference for the conditional seemed an easier choice when it was coupled with a high- rather then a low-trustworthy source.

In addition, there was again an interplay of revision strategies in that beside source trustworthiness, the participants choices were also influenced by matching mental models of the premises. This confirmed the second hypothesis. In contrast to Experiment V, however, the mutual nature of these two revision strategies was evidenced not by the DTs but by the belief revision responses. The interaction effect showed that the effect of source trustworthiness was more robust for MT than for MP problems. As can be seen from Figure 17, reliance on source trustworthiness led to an independent belief revision strategy for MT problems, whereas it merely modulated the overruling strategy used for the MP problems, which was the mismatch principle. To illustrate, even though with MP problems the belief in the conditional lowered from 34.44% to 17.78% from high- to low trustworthiness, in both cases the categorical still won preference. In contrast, with MT problems the belief in the conditional dropped from 71.11% to 41%, meaning that in the first case the belief in the conditional was preferred but in the second case the belief preference switched to the categorical statement. Finally, the belief in the conditional in these last three experiments was lower than in Experiments I-IV. The most likely explanation is that people are more careful believing a conditional when it is explicitly coupled with a clear source.

# Chapter 6

# Discussion and Future Perspectives

## 6.1. General Discussion

This dissertation presented a series of experiments that served two main goals. The first goal was to find out to what extent belief revision processing can be explained by human reasoning theories rather than philosophical AI theories of belief revision. The involvement of these theories was attempted to be resolved through the comparison of the *mismatch principle* from the MMT with theories emphasizing the influence of context and context factors, with special emphasis on *probability theories*. To reiterate, the MMT posits that people think and reason in terms of mental models that capture the semantic meaning of the premises. Belief revision, then, is the result of retaining or investing more in that belief that has a matching mental model with that of the contradiction. The probability theory proposes, instead, that people take into consideration the probability of the premises, relate this to prior knowledge and belief and base their belief revision thereon. To test these theories the content factors probability, familiarity, and the context factors phrasing of questioning and source trustworthiness were explored. MP and MT inference problems were used to be able to test the mismatch principle. The factor 'probability', as inspired by the probability theory of reasoning, was investigated in experiments I, II, and III. The additional factor 'familiarity' was considered in experiment III. After having investigated these content factors, the focus turned to the context factors 'task instruction' in experiment IV and 'source trustworthiness' in experiments V, VI, and VII. In each of these experiments, the respective content or context factor was compared with the mismatch principle from the MMT. By doing so, the current work aimed to unravel whether there exists one unified strategy or rather a multiplicity of strategies in human belief revision. Second, decision times were investigated to take a deeper look into the cognitive processing mechanism behind belief revision.

Over the course of experiments, two main findings became evident that are of interest to the study of belief revision. First, both the mismatch principle as well as content and context factors significantly influenced the course of belief revision. Two main belief revision strategies were recognized, one based on probability and one based on mental models. Moreover, it was discovered that they did not operate independently

but rather interacted. This finding was supported by both the belief preference choices and the decision times. There was always one prevailing strategy of belief revision with an underlying influence of the other strategy. Second, conform to our second main hypothesis, the DT results offered insight into which inference problem was easiest to perform. In the case where the two theories predict the same revision choice, this elicited the fastest DTs. In a similar vein, the slowest DTs were found with those inference problems where the theories predicted opposite belief revision choices. The finding of two main belief revision strategies runs against the assumption of one general revision strategy as was concentrated on in former belief revision research.

The following sections will report in more detail on the exact influence of the mismatch principle and each of the content and context variables.

### 6.1.1. Mental models, probability and familiarity in belief revision

In experiments I, II and III, the factor probability was manipulated to explicitly test whether it guides belief revision or whether this process is a sole function of comparison of mental models. Support was found for both the MMT (Johnson-Laird & Byrne, 1991, 2002) and probability theories (Chater & Oaksford, 2001; Evans & Over, 2004; Evans, 2006; Evans, et al., 2003; Over & Evans, 2003; Oberauer & Wilhelm, 2003; Over et al., 2007). More precisely, probability guided belief revision when the conditional premise expressed a daily action that was either very plausible or very implausible. However, the matching of mental models showed to be the overruling strategy when the probability of the event displayed by the conditional was at or near chance level. The findings from both Experiments I and II demonstrated that when the inconsistent set of statements included a high probability conditional, the conditional was believed more than the categorical statement. Conversely, when the problem set included a low probability conditional, people preferred to believe the categorical statement instead. This was the case for both MP and MT inference problems. These findings show that people incorporate probability into their belief revision process. How then can the current results support for a probability approach be aligned with the probability account in deductive reasoning? Deductive reasoning tasks explore people's confidence in the conclusion. In belief revision tasks, on the other hand, both the conditional and the

categorical are understood as uncertain and the conclusion is assumed to be true as it is presented as a contradictive fact. However, this need not necessarily mean that the psychological process of probabilistic reasoning in belief revision is different from that in deductive reasoning. The results align with the Ramsey test (Ramsey, 1931/1990b; as stated in Oaksford & Chater, 2009). To explain, the psychological process behind the Ramsey test concentrates first on the general possibility of p. This possibility is then divided in 'p q' and 'p ¬q' possibilities (Evans et al., 2003; Oaksford & Chater, 2009). The reasoner then compares the possibility of 'p q' with the possibility of 'p ¬q'. In case 'p q' is judged as more probable than 'p ¬q', then the conditional 'if p then q' is assigned a high probability. If instead 'p ¬q' is judged as more probable than 'p q' then the conditional 'if p then q' receives a low probable value. To translate this to belief revision, if a person chooses to believe the conditional more, than one chooses for the first possibility p q. When instead someone chooses to believe the categorical more than the second possibility, p ¬q is opted for (Evans et al., 2003). Now, this is exactly what occurred with the current MP and MT problems. In the high-probability conditions, subjects believed the conditional more, that is they believed 'p q' to be more possible than 'p ¬q'. In the low-probability conditions, subjects believed the categorical more, that is they attached a higher belief to the possibility 'p ¬q' than to the possibility 'p q'. The role of mental models was evident in that the belief in the conditional was higher for MT problems than for MP problems. This aligns with the mismatch principle (Johnson-Laird, 2006; Johnson-Laird & Byrne, 2002). The confluence of the two strategies was portrayed by a prevalence of probability-guided belief revision with the underlying influence of mental models.

Experiment III tested the additional factor 'familiarity'. Familiarity did not lead to a distinctive belief revision strategy. Rather, the findings from Experiment III showed that with both familiar and unfamiliar problems the belief revision showed a pattern that strongly accords with the mismatch principle. That is, with MP inference problems the categorical was more endorsed and with MT problems the conditional was more endorsed. People, then, do not appear to revise their beliefs differently when confronted with contradicting premises of which they possess no or very little knowledge than when they are familiar with the content of the information. Rather, the picture emerges that

when the content of conditional premises is devoid of a clear high or low probability, people will revert to the use of mental models. However, familiarity did serve as a modulating factor. That is, the reliance on mental models became an even more preferred strategy when the inference problems were unfamiliar, but nevertheless known to occur in real-life, opposed to familiar. The findings do raise the question how it comes about that people are willing to claim belief in something that is not within their frame of knowledge. Of course, the participants had no other choice than to attach a stronger belief in one of the two statements since this was their task. In real life, however, people are often not "forced" to choose a statement and can therefore distance themselves from any set of statements and set their minds entirely free of belief. Still, even though they performed in line with the mismatch principle, the presence of a clear preference for the conditional with the MT problems indicates that they for some reason believed without understanding it. People's understanding of the tie between believing and comprehending already sparked discussions in ancient times, most famously between Descartes and Spinoza, described in a paper by Gilbert (1991). Gilbert describes that Descartes stated that believing and comprehending are independent where comprehension precedes belief. Spinoza, on the other hand, reasoned that comprehending and believing comprise the very same operation. Gilbert tested these two views against each other and found support for Spinoza's view. Findings of a study by Hasson, Simmons, and Todorov, 2005, however, supported Descartes' idea. The results indicate that in accordance with Hasson et al. (2005), believing and understanding are likely to be independent but in contrast to their findings, in experiment III believing seemed to precede understanding. This is line with the view by Recanati (1997) that people are capable to believe a statement that they do not fully understand. Recanati introduced the term 'quasi-belief' to indicate the process of 'acceptance without understanding'. He further discusses in his paper semi-propositional knowledge. This may include cognitive steps such as figuring out by certain words in a sentence what knowledge domain the statement might belong to. Even though the current unfamiliar premises did not match the participants' knowledge base they were willing to invest belief in one of the two premises (depending on the logical inference problem), because they knew it was truly existent and contains information which one can turn into

knowledge after learning and repeated exposure. The participants may have known that the conditionals referred to the domains of music instruments, micro-biology, physics, and archaeology and that these represent real-life knowledge domains in which human causal actions occur. This also explains the difference in belief revision choices found here and in earlier experiments on belief revision where science fictional material was used (e.g. Byrne & Walsh, 2002; Byrne & Walsh, 2005, Experiment 1; Elio & Pelletier, 1997; Politzer & Carles, 2001, Experiment 1). In those experiments the uniform finding was that in overall the conditional was believed less than the categorical. Based on the findings of Experiment III, it can be stated that it was not the unfamiliarity that made people shy away from believing the conditional but rather its non-existence (for referring to an imaginary world). This most likely prompts the reaction in people's minds that one cannot attach any level of belief to an action that simply does not exist.

### 6.1.2. Mental models and task instruction in belief revision

Experiment IV was inspired by the wide variety of response options utilized by the different studies of belief revision in the past (Byrne & Walsh, 2005; Elio, 1997; Elio & Pelletier, 1997; Dieussaert et al., 2000; Politzer & Carles, 2001; Revlin, et al., 2001; Revlin, et al., 2005). To rule out the possibility that the differences in task instruction might have caused the different results, Experiment VI manipulated the variable 'task instruction'. Familiar inference problem with medium probability were used and the participants were asked either which statement they believed *more* or which they believed *less*. The experiment brought about the interesting finding that phrasing of questioning had a significant effect on belief revision. In both instances, similar to the findings of experiment III, the participants used the mismatch strategy. However, this was more pronounced when asked to indicate the statement that was believed more. Earlier, it was already speculated this might be because asking oneself or others what to believe more, or using 'more' in any other evaluative situation, is a more frequent enterprise than using its opposite 'less'. To understand how, then, this change of one single word in a question that should result in the same evaluation can have such an effect we will enter here the field of psycholinguistics. Clark (1969) proposed a theory on the role of linguistic processes in (deductive) reasoning problems. The second of three

principles he proposed, the *principle of lexical marking*, is of interest here. According to this principle, the positive adjective (which he denoted "marked") of any pair of antonymous adjectives is stored in memory in a less complex manner than its negative counterpart (denoted "unmarked"). Examples given by Clark are *good-bad*, *long-short*, and *wide-narrow* (of which the first adjective of each pair is the positive one). A positive adjective is called 'unmarked' because it can occur in a noncommittal nominal sense (e.g. the sentence is 10 cm long) as well as in its more contrastive meaning (e.g. the long sentence) in which case it is compared with some implicit standard. A negative adjective, on the other hand, is termed 'marked' because it can take on only its contrastive meaning (e.g. the short sentence). The contrast between the two questions in Experiment IV can be similarly explained in the context of Clark's theory. The question '*Which of the two statements do you believe more*?' as well as its counter question '*Which of the two statements do you believe less*?' are both comparative in nature. However, the word 'more' is an unmarked adjective which implies that it can be interpreted in both its nominal and contrastive sense. The word 'less' is its marked counter-adjective; it can be interpreted only in its contrastive sense which makes it more complex and in turn more restrictive in its use. A nominative interpretation is less complex, easier to understand, and better retrieved. Therefore people will choose to interpret the 'more' questions in its nominal form. One might argue that believing versus disbelieving might be a better contrast because they are both believed to be unmarked (Gilbert, 1991). However, most recent studies on reasoning and belief revision show that people don't have an all-or-nothing belief (i.e. accept or reject) in a statement but rather a degree of belief or express doubt in a belief (e.g. Politzer & Carles, 2001, Experiment 1; Dieussaert et al., 2000). For this reason, the questioning that was used here more resembles real-life belief revision and belief update processes.

As in the first three experiments, the findings exposed the working of two cognitive mechanisms: the mismatch principle and a psycholinguistic device, although the latter is not considered here as a distinct belief revision strategy. The mismatch principle served here as the dominant belief revision strategy. The fact that there was found a significant difference between the two different task instructions only for the MT inference problems might imply that the resistance to not choose the match with MT is greater

than the resistance to not choose the non-contradicting mismatch with MP inference problems. The DT findings seem not only to reflect the lexical marking principle by Clark (1969) but also offer an adaptive view on the mismatch principle. Johnson-Laird and Byrne (2002) describe the mismatch principle as a process by which one searches for a mismatch between mental models. However, the results of Experiment IV rather portray the image that people more readily prefer to find a mental model within their belief set that matches instead of mismatches the mental model of the incontrovertible fact. This view is further strengthened by the DTs findings of the first three experiments with those MT problems where the different revision strategies played in harmony and both pointed toward believing the match; these consistently showed the fastest DTs of all the different conditions. Also, the finding in all four experiments that in overall the conditional was believed more than the categorical statement only further adds to the supposition that people might be more tuned into finding a match than a mismatch.

### 6.1.3. Mental models and source trustworthiness in belief revision

Experiments V, VI, and VII presented participants with MP and MT inference problems with varying degrees of source trustworthiness on which they had to perform belief revision. The first hypothesis was that source trustworthiness would have an effect on belief revision. This was for the most part supported. As with the medium-probability inference problems in experiment II, the prevailing strategy was the mismatch principle. The conditional was believed more with MT problems and the categorical was believed more with MP problems. The results from Experiment I and III clearly show that source trustworthiness influenced belief revision, both when the source was represented as an occupation and as a person description. This is in line with findings from the persuasion and attitude change literature (Eagly, et al., 1978; Hovland & Weiss, 1951; Kelman & Hovland, 1953; Mills & Jellison, 1967; O'Hara, et al., 1991; Priester & Petty, 1995; Rhine & Severance, 1970; Tormala & Clarkson, 2008; Tormala, et al., 2005; Wiener, & Mowen, 1986). More importantly, it aligns with theories stating that prior world knowledge is integrated in reasoning performance (Evans, 2008; Evans, et al., 1993). Results from experiments I-IV already showed that mental models, and content of the material, i.e.

probability and familiarity, had a profound effect on belief revision. Results from experiments V-VII show the further importance of social context effects. Notwithstanding a clear involvement of source trustworthiness in belief revision, it is however surprising that it did not serve as the prevailing belief revision strategy like probability did, testifying the immense impact source trustworthiness has on persuasion and attitude change in the social psychology literature (Eagly, et al., 1978; Hovland & Weiss, 1951; Kelman & Hovland, 1953; Mills & Jellison, 1967; O'Hara, et al., 1991; Priester & Petty, 1995; Rhine & Severance, 1970; Tormala & Clarkson, 2008; Tormala, et al., 2005; Wiener, & Mowen, 1986). The premise that was uttered by a high-trustworthy person did not always win preference over the premise uttered by the low-trustworthy source. One factor that may have waned the influence of source trustworthiness is that people do not want to come across as prejudice. However, this seems highly unlikely as expressing different levels of trust in occupational sources is not considered highly controversial. Nevertheless, to test whether the participants repressed their true preference, future studies might gain from asking their subjects what another person would belief more. Examining beliefs about other agents' beliefs is winning interest in the research domain of logic which has in turn been motivated by the literature on theory of mind in psychology (van Ditmarsch & Labuschagne, 2007). The effect of source trustworthiness may be stronger when people have to imagine and/or indicate what another might believe more. A more probable explanation, however, for the diminished effect of source trustworthiness is that, unlike in a real-life setting, the participants were not confronted with the source in person. Would one for example have standing in front of him or her a police officer in uniform, the direct perception of the authoritative nature of the officer would have a bigger impact than merely seeing the word 'police officer' on a computer screen.

Experiment VI failed to find an effect of source of trustworthiness. In the discussion section of Experiment VI it was already presumed expertise might have been inherent in the inference problems and displaying a stronger influence on belief revision than trustworthiness. It might therefore be that the participants did not even consider trustworthiness to be a factor in play but rather looked at expertise and decided thereupon that there was little difference in expertise. As already mentioned in the

introduction, the sole aim of the final three experiments was to investigate the effect of source trustworthiness. Nevertheless, as expertise has been shown to influence acceptance rate in deductive reasoning problems (Stevenson & Over, 2001), it presents a factor that should certainly not be neglected in further belief revision studies in order to complete the picture of source credibility.

The second expectation was that belief revision would take longest when the two uncertain premises were uttered by people of equal source trustworthiness. This seemed to be the case only when such a condition was contrasted with instances where the statements were preceded by sources of different level of trustworthiness. To illustrate, in Experiment V, the DTs for high/low and the low/high conditions were much lower, 3.52 s and 4.52 s respectively, than for the medium/medium condition which was 11.04 s. However, in Experiment VII, we found that the DTs for the low/high condition was much higher, 16.10 s, than for the high/low condition, 10.36 s. This illustrates that the ease of relying on source information in belief revision is much dependent on the social context. This finding relates to studies showing that human judgment is dependent on contrastive context effects. Perceived knowledge about or attitudes towards an object or person is influenced by information or attitude of other objects or persons encountered in close temporal proximity (Markman & McMullen, 2003; Mussweiler, 2003; Tormala & Petty, 2007). The applied standard of comparison in Experiment 1 was that of a 'clear difference' (High/low and Low/high) versus 'no clear difference' (medium/medium). In contrast, the standard of comparison in Experiment III was 'High' versus 'Low'. This finding is important because people rarely ever receive information in isolation but rather from multiple sources within a single context.

Similar to previous findings in the belief revision literature (Dieussaert, et al., 2000; Elio, 1997; Elio & Pelletier, 1997; Politzer & Carles, 2001; Revlin, et al., 2001), all three experiments showed a lower overall endorsement of the conditional over the categorical statement. However, in these former experiments this lower preference for the conditional applied to both MP and MT inferences. We found, in contrast, that this effect accounted only for MP inference problems. Content factors, such as probability and familiarity, and context factors, such as task instruction and trustworthiness (current experiments) influence the relative rate of belief in the two premises. As we speculated

earlier, the lower rate of belief in the conditional in earlier experiment was most likely due to the non-existent nature of the content of the material. The current finding that the belief for the conditional was lower in Experiment I (34.93 %), in which abstract material was used, than in Experiments II (42.27 %) and III (41.25 %), where the materials' content related to real-life events, strengthens our speculation. The material was abstract which could have triggered this resistance.

Again, there was a shared operation of belief revision strategies. There was a joint involvement of source trustworthiness and the appliance of mental models in belief revision. Belief revision was a confluence of a more specific strategy based on source trustworthiness and a more global strategy of comparing mental models, as shown by the revision choices in Experiment I and the DTs in Experiment III. Although the preference for the conditional dropped significantly from the high/low to the low/high condition, in all three experiments preference for the conditional was found to have its highest percentage with the MT relative to the MP inference problems. This finding is in line with the predictions of the mismatch principle (Johnson-Laird & Byrne, 2002). These different strategies are a direct result from different processing styles. However, a drawback of the experiments is that participants might have already had a preference for one over the other premise even before the contradictive conclusion was presented. A control study should be added whereby only the two premises are presented without the contradictive conclusion whereby the participants are similarly asked which statement or source they believe more.

## 6.2. Dual-process account of belief revision

The results across the seven experiments pointed to a shared operation of belief revision strategies. Both revision choice and DT showed a unified trend in that they reflected a partnership of applied belief revision strategies. Belief revision was a confluence of a more a specific strategy based on content and context factors and a more global strategy of comparing mental models. This is a new and interesting addition to the research of belief revision. This finding adheres to the segmentation of human reasoning in domain-general and domain-specific reasoning (e.g. Thompson, 1994). In overall, people adjusted their belief in the conditional to the perceived probability of the conditional statement,

leading to one of two belief revision strategies. This section proposes how this finding can likely compose a theory of human belief revision.

In Experiment I, the prevailing strategy of belief revision was based on probability but with an underlying influence of mental models. In the remaining experiments, belief revision was dominated by the manipulation of mental models on which the content and context factors modulated. In Experiment III, the manipulation of mental models was stronger the case with unfamiliar than with familiar inference problems. Results from experiment IV exposed the mutual working of the mismatch principle and a psycholinguistic device; the context factor 'task instruction' influenced the strength of applying the mismatch principle. The experiments on 'source trustworthiness' found a shared operation of this context factor together with the appliance of mental models operating belief revision. This was shown by the revision choices in Experiment V and the DTs in Experiment VII. Although the preference for the conditional dropped significantly from the high/low to the low/high condition, in all three experiments preference for the conditional was found to have its highest percentage with the MT relative to the MP inference problems. This finding is in line with the predictions of the mismatch principle (Johnson-Laird & Byrne, 2002). The DTs results across all experiments add to this that in the case where the two theories predict the same revision choice, this elicited the fastest DTs. In a similar vein, the slowest DTs were found with those inference problems where the theories predicted opposite belief revision choices. This joint involvement of prior knowledge and belief and the appliance of MM/belief revision strategies adds further strengths against the assumption of one general belief revision strategy (e.g. Elio & Pelletier, 1997).

How would the pattern of findings across the seven experiments, then, fit into a theory of human belief revision performance? We argue that the current findings share some commonalities with existing dual-process theories on thinking and reasoning. Leading theories of its kind in the domain of thinking and reasoning are the dual process theory of reasoning and judgment developed by Evans and Over (1996) and the dual-process model by Sloman (1996). These theories are very close cousins of the dual-process theories developed in social psychology, of which the two most influential and most pertinent to the current research actually come from the field of persuasion and attitude,

the *elaboration likelihood model* (ELM) (Petty & Cacioppo, 1981, 1986) and the *heuristic-systematic model* (HSM) (Chaiken, 1980; Chaiken, Liberman, & Eagly, 1989). These theories have in common that they segregate between a heuristic style of thinking and an abstract rule- or model- based thinking. In the thinking and reasoning arena they are also called System 1 and System 2 processes, respectively. System 1 processes are characterized as fast, automatic, and unconscious and System 2 processes as slow, conscious, and controlled. When approaching a problem based on heuristic-style thinking, one uses preexisting knowledge and belief (Evans and Over, 1996; Sloman, 1996). By this route, one approaches a problem by using simple, well-learned and readily accessible associations from memory about similar encounters in the past (e.g. Chaiken, 1980; referred to as a mental shortcut by Petty & Cacioppo, 1981, 1986). With heuristic processing in charge, agreement with a message might stem from an association between positive feelings and the source of the message when this source is attractive, likeable, or trustworthy. Rule-based thinking is believed to serve decontextualized abstract thinking, involving logical norms and leading to deliberate scrutinization of a problem. Rule-based thinking or reasoning is analytic of nature and involves an effortful scrutiny of a message. It performs a validity test on the problem by accessing stored knowledge to 'upload' relevant facts from related problems.

Some have argued that mental model-based thinking is more akin to an analytic type of thinking and probabilistic reasoning resembling more a heuristic thinking style. In the past, the widely shared viewpoint was that heuristic or peripheral type of processing constituted the faulty one and susceptible to bias and the systematic or central type of processing was the accurate logic-based thinking (e.g. Evans and Over, 1996; Oaksford and Chater, 2001; Verscheuren, Schaeken, & d'Ydewalle, 2004). Most dual-process theories have undergone this connotation of an evaluative distinction (Smith & DeCoster, 2000). However, it is not the intention at this point to attach different processing connotations to the belief revision strategies, for several reasons. First, Verscheuren et al. (2004) and Evans (2008) postulate that the construction of the initial explicit model is an automatic process. What makes the mental model process analytic is the fleshing out of the remaining former implicit models; the effortful part that requires working memory. However, there was no indication that the participants searched for

counterexamples. Rather, the appliance of the mismatch principle was based on the initial explicit model only. Second, System 2 type of thinking requires more processing and in turn longer decision time than System 1 thinking (Evans, 2003). However, the DTs here did not show that reaching a belief revision decision by applying the mismatch principle took longer than when it was derived at by prior knowledge and belief (this, however, could be directly related to sticking to the one explicit mental model). Third, the assumption of decontextualization from background knowledge of System 2 processing is quite ironical when applied to the current context; model-based belief revision could not have operated fully decontextualized because one should first determine that the probability of the conditional statement is near 50% before converting to the mismatch principle. Therefore, it is not a matter of *whether* prior knowledge and believe is invoked, but more so whether one's knowledge base will function as the chosen instrument with which to derive at the preferred belief revision choice. Finally, System 1 type of reasoning is believed to result in erroneous reasoning performance because the participant is distracted by an erroneous feature of a deductive performance task. Evans (2002) outlined in a recent review that for many decades, researchers believed that logic underlies rational human thought and reasoning. Logical thinking should be a process inherent in every human mind that is devoid of prior knowledge and beliefs. Therefore pragmatic rich contents should lead to biases and wrongful performance. System 1 is constructed from a similar line of thinking. However, belief revision tasks differ from deductive reasoning performance settings, to which this critic was directed at, in that there is no correct answer. Even so, we do not believe that using prior knowledge and belief in reaching a favoured belief is a faulty belief revision strategy. For example, from an evolutionary standpoint it is advantageous to think flexible. Evolution bequeathed us with flexible minds with which we can easily adapt to the environment. In this regard, relying on experience and taking context into account at the cognitive intersection of belief revision seems a feasible strategy to guide our future actions. In more recent days, the viewpoint that heuristic style of thinking can actually be very helpful is winning popularity.

Apart from these differences, what is shared by the current findings and one major postulation of the dual-process theories is that the thinking styles or strategies do not

function independently. The dual-process models from both social and cognitive psychology take the stand that the two different processing styles do not rule each other out, but rather work simultaneously in cognitive operations; their relative contribution differ in strength and can operate in addition or opposition. The belief revision strategies function in a similar fashion. At each point in time, two contradictory belief revision processes operate. They may lead to divergent belief preferences or may converge to the same belief. As Sloman (1996) puts it, responses are guided by different processing modes whereby each has its own subjective 'pull'.

As a final note, the factors familiarity, task instruction, and source trustworthiness were set apart as independent factors different from probability in the development of the experiments. However, this is not to say that a probabilistic approach to belief revision does not cover other factors other than merely probability. The other content and context factors can be viewed as a subpart of probabilistic reasoning because they induce uncertainty which is the core idea of probabilistic reasoning. When viewing so, then two major belief revision strategies exist, one by which statements are represented as mental models and another by which statements are represented as probabilities, either directly or resulting from another uncertainty factor.

## 6.3. Can all the results be explained solely by the MMT or by the probability theory?

The results across the seven experiments have been interpreted as providing support for both the mental model theory and the probability account. In the former part, a kind of preliminary dual-process account was proposed to explain the involvement of both a logic-based belief revision process and probabilistic-based belief revision process and the interaction between the two. However, one should consider the possibility that all the results can be explained either solely by the mental model theory or by the probability theory. This possibility will be discussed in this section. However, more in-depth discussion will be devoted to the probability theory as it is in the process of becoming the dominant theory of reasoning.

The MMT states that a conditional captures a core semantics that allows the three possibilities pq, ¬pq, and ¬p¬q. These possibilities very well fit the concept of truth-functionality as dictated by logic. Logic gives truth-functional values to propositions, which means that they can be either true or false. The MMT was build mainly to explain reasoning with basic conditionals (Evans, Over & Handley, 2005). Johnson-Laird and Byrne (2002) define basic conditionals as "those with a neutral content that is independent as possible from context and background knowledge, and which have an antecedent and consequent that are semantically independent apart from their occurrence in the same conditional" (p.648). However, one can simply not deny that prior knowledge affects how conditionals are understood by people and how they draw inferences from them (Evans, et al., 2005). To date, the MMT is the only theory on human reasoning to propose a mechanism driving human belief revision. However, the mismatch principle does not clarify how content is dealt with. Johnson-Laird and colleagues extended the MMT at a later stage by incorporating a theory of probabilistic reasoning which they termed 'extensional reasoning' to try to account for the effects of context and prior knowledge in reasoning (Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999; Girotto & Johnson-Laird, 2004). However, their extensional reasoning account cannot adequately explain the probability effect we found in Experiments I and II. Johnson-Laird et al. (1999) take extensional to mean 'inferring the probability of an event from the different possible ways in which it could occur' (p.63). By 'different possible ways' they mean the possible mental models that can be constructed. By default, they posit, people assign equiprobable possibilities to each mental model which they coined the *equiprobability principle*. Johnson-Laird et al. do however assume that the explicit model will be assigned a higher probability then the implicit models. Johnson-Laird and Byrne (2002) further state that in case of non-basic conditionals, the context of these will trigger general knowledge and also more specific knowledge concerning the utterance by means of *pragmatic modulation*. This modulated mechanism works as follows: "This context is normally represented in explicit models […] and aid the process of constructing fully explicit models" (p. 659). Applying the pragmatic modulation to belief revision, a conditional expressing a low probability should trigger people to fully flesh out their mental models of the conditional statement to search for a more suitable

model, which in this case would be *p & and not-q.* This should convince people to engage more belief in the conditional as was found with the high-probability problems in Experiments I and II. However, it was found that people retract from believing the conditional statement when it expresses a low probability. This indicates that people construct only the explicit mental model *p & q*. Therefore, the findings are difficult for the theory to explain, since they cannot be attributed to the operation of corresponding principles of mental models. For this, it is advocated that reasoning with probabilities is of a different kind than is proposed by the MMT.

In the above it is outlined why the MMT faces difficulty explaining the results that did not coincide with the predictions of the mismatch principle. However, there still exists the possibility that the results that seemingly fit the principles of the mismatch principle can be explained by the probability theory. Recall that on medium-probability MP problems, the categorical was believed more and that on medium-probability MT problems the conditional was believed more. The mismatch principle clarifies the results as follows: with MT problems, there is a clear match between the contradiction's model 'p' and the antecedent of the conditional's model 'p  q'. For this reason, people will discard the categorical and express a preference to believe the conditional. With MP, there is a clear conflict and mismatch between the mental model of the contradiction 'not-q' and that of consequent of the conditional 'p q'. People will then choose to discard the conditional and believe the categorical more (Johnson-Laird, 2006; Johnson-Laird & Byrne, 2002; Johnson-Laird, et al., 2004). Now lets consider how the probability theory might explain the findings with medium-probability inference problems. The probability theory takes the stand that everyday reasoning rests on uncertainty, where a premise can be judged ranging from most unlikely true to most likely true. This mirrors a discrepancy in the meaning and interpretation that people give to the premises. What this dissertation found is that when a conditional's content is unlikely plausible, it receives a low belief value and when it is very likely plausible it receives a high belief value. The mismatch principle was given credit for the results on medium-probability problems. Now it will be considered how the probability theory would most likely explain the results on the medium-probability problems given its assumptions laid out in the introduction. When it comes to a premise with medium certainty, the conditional

probability theory states that the probability of the conditional should equal the probability of the conclusion (e.g. Oaksford & Chater, 2009). This would then explain the results on the medium-probability inference problems. With respect to MP problems, in the case where subjects judge "if p then q" to be of low or medium probability, they will be easily convinced of the contradictive fact and place little confidence in the conditional: This would imply the following for the inference problems:

> Low- or medium probability MP problem: Asserting with certainty that the conclusion $P$(not-q) = 1 implies that $P$(q) = 0 and so $P(q \mid p) = P$(if p then q) = 0.

With MT inference problems, being convinced of the contradiction would imply that people place high confidence in the conditional.

> Low- or medium probability MT problem: Asserting p with certainty makes P(p) = 1 which implies that $P(q \mid p) = P$(if p then q) = 1.

However, directly applying the thought process behind probabilistic thinking in consistent reasoning problems to the inconsistent reasoning problems that ask for belief revision would be a rather blunt "hypothesis". Below, it will be outlined how the unique process of probabilistic reasoning in belief revision applies to inference problems with medium-probability conditionals.

The belief revision tasks reflect the belief state in a person's mind. In belief revision, a person holds a belief with a certain level of certainty about a causal relationship. Here follows an example of a conditional with medium-certainty (taken from Experiment II): "If Christian is working in his office, then he leaves the door open". A person then receives the information that Christian is working in his office. From her causal belief she then infers that Christian has his door open. The person then learns as a fact, for example, by first-person observation or from a recent photo that Christian is indeed in an office, but the door is closed. The person then can revise her believe that when "If Christian is working in his office, then he leaves the door open" or she can believe that

the office she sees Christian in is not Christian's office but his college's office. Perhaps they are discussing important information. The person that gave her the information that Christian is in his office then was mistaken. To put this in modus ponens form:

| | |
|---|---|
| If Christian is working in his office, then he leaves the door open | $P$(if p then q) |
| Christian is working in his office | $P$(p) |
| Christian does not leave the door open | $P$(¬q) |

Herein lies the difference with ordinary (consistent) reasoning problems. In the case above, even though the contradictive conclusion as being a fact should receive a probability of 1.0, this does not automatically imply that the conditional statement also receives a 1.0 probability. Inferring from the above-sketched thinking process, there are two possible scenarios. Because the person has only a 50% belief in the conditional, probability will not help her decide whether Christian always has his door open or whether he is simply not in his office. Because the antecedent clause, i.e. the categorical premise, is uncertain, the presumed 1.0 probability of the contradictive statement might, as a consequence, not be attached to the conditional. The door referred to in the third statement could be another person's office door. The belief revision process with MT inference problems follows the same line of reasoning:

| | |
|---|---|
| If Christian is working in his office, then he leaves the door open | $P$(if p then q) |
| Christian does not leave the door open | $P$(¬q) |
| Christian is working in his office | $P$(p) |

The conditional consequence that Christian leaves the door open when he is in the office, has 50% probability. But the categorical statement that the door is not open is also left

uncertain. It could be that Christian does not always leave the door open when he is in his office or it could be that the door that is referred to in the categorical statement is not his door. The problem's medium-probability makes that probability cannot be relied on when deciding which statement to believe more. The problems might be approached in terms of probability, but if probability is not high or low enough to provide sufficient informational value, people will go by resolving the contradiction by other means. In the case of medium-probability problems, people can resolve the conflict by reverting to a more global belief revision strategy, that of comparing mental models.

There is a second point that needs consideration. Across the experiments it was found that the conditional gained more belief with MT than with MP problems. In all the experiments, preference for the conditional was found to have a higher percentage with the MT relative to the MP problems. This finding is in line with the predictions of the mismatch principle (Johnson-Laird & Byrne, 2002). Because there will be high confidence in the truth of the contradiction with medium-probability problems, proponents of the conditional probability hypothesis might argue that the belief in the conditional for MP is low and therefore always lower than it is for MT (unless not-q is certain). However, the results of several current experiments challenge the conviction of the probability theory that all content-rich reasoning problems are approached and resolved merely in terms of probability. The best way to illustrate this is with inference problems that had medium-probability content but were manipulated on another content or context factor and showed different results.

The first example is experiment III that used medium-probability inference problems but manipulated the factor 'familiarity'. With MT problems, the conditional gained higher belief with familiar than with unfamiliar problems. Also, the decision time was faster for familiar than for unfamiliar problems. Likewise, with MP problems the DTs were higher with unfamiliar than with familiar problems. If the probability of the conditional should equal the probability of the conclusion, then no difference between the rate of belief in the conditional and DTs should show up across the two familiarity conditions. Even more challenging to explain for the probability theory are the results from experiment IV. Again, medium-probability inference problems were used but 'task instruction' was manipulated. What makes it even harder for the probability theory to

explain is that the content of the MP and MT problems across the two task instruction conditions was identical, only task instruction differed. With MT problems, the percentage of choosing to believe the conditional more was significantly higher in the 'more' condition than in the 'less' condition. The DTs with these inference problems were also significantly faster in the 'more' than in the 'less' condition. Task instruction should not have led to an interaction effect if it were the case that people approached these problems with subjective probability estimates.

In short, the probability theory postulates that belief in the conditional should equal that of the conclusion. This fits the idea of conditionalization, in which case the categorical assumption is assigned the status of truth. However, emphasis was put here on the Ramsey test, not the Equation. This dissertation supports the probability theory to the end that people interpret reasoning problems in the context of prior knowledge and belief and that they assign degrees of beliefs to statements. The Ramsey test, on which the probability theory is founded, finds support in the results of the current experiments. However, whether the probability of the conditional should exactly match the probability of the contradiction (i.e. involvement of the Equation) was not the aim of this dissertation and would need further investigation to gain support.

In conclusion, only by means of a confluence of two strategies can the current results be adequately explained.

## 6.4. Linking the findings to AI belief revision theories

The current investigation was purposely framed within the realism of psychological theories of human thinking, detached from terms such as 'epistemic entrenchment' (but see e.g. Byrne & Walsh, 2005; Elio & Pelletier, 1997). Nevertheless, it would be undeserving toward the AI and philosophy arena to not attempt to link the current theorizing of the findings to the original belief revision theories since they lay the foundation on which psychologists build their belief revision research. Also, the goal state in both cases is to develop a theory of how knowledge is represented and updated in the face of new and often contradictory information. Although numerous theories on belief revision saw the light of day in AI (e.g. AGM by Alchourrón, Gärdenfors, and

Makinson, 1985; epistemic entrenchment by Gärdenfors, 1988) and in philosophy (e.g. Harman, 1986), we will restrict the discussion here to the notion of "epistemic entrenchment" (Gärdenfors, 1988) and the "minimal change principle" (Harman, 1986). In AI (Gärdenfors, 1988), epistemic entrenchment is a function of preference ordering on the possible changes of belief revision. Statements will be ordered by their explanatory power or usefulness with entrenchment of that statement that is most deserving to be maintained. Elio & Pelletier (1997) expatiated upon how epistemic entrenchment might function in human belief revision. They stated that epistemic entrenchment can be interpreted in two ways. The first one is the idea that conditionals contain a certain preference ordering, assuming a priori that conditionals have a 'law-like' nature and often express regularities about the world that are hard to refute. The other perspective they take on is that categorical statements inherent more entrenchment because they express observations, data, or straight-out evidence. This, however seems a rather restricted definition of conditional and categorical statements. The current results showed that it's not about the form of the statement per se, it is about what it expresses, its' content. When the conditional premise contains information that conveys believability, then preference ordering seems to be based on exactly that. Probability is such a believability factor. In this context, the conditional wins entrenchment when it conveys information with a high probability which in turn gives it more explanatory power. On the hand, the categorical is more deserving of entrenchment when the conditional expresses a low probability event. In this regard, probability-based belief revision fits the epistemic entrenchment notion by Gärdenfors (1992). According to Gärdenfors, belief revision is not syntax- but knowledge-based which implies that sentences that are logically equivalent but have different syntactical structure should result in identical revisions. The MP and MT problems are logically equivalent but with different syntactical structures. With probability-based belief revision, it was indeed the case that the same premise was believed across the two inference problems. Therefore, probability-based reasoning is more akin to epistemic entrenchment as set out by Gärdenfors.

When the content of the material displays no sign of believability, then people convert to a process of matching and mismatching of mental models. This seems to fit

the more global view on minimal change as proposed by Harmann (1986). In the AI domain, minimal change is the process of engaging in belief revision by making as little change as possible in the belief state. The driving force behind the minimal change principle is simplicity. The Simple Measure of Change in View takes into consideration only the numbers of changed beliefs. It was also found, for example that the measure has a preference for simple hypotheses over more complicated hypotheses (Harman, 1986). Preferring to invest more belief in a statement that has a mental model that matches instead of mismatches the mental model of the counterfact seems more akin to the minimal change approach. Also, the cognitive demand that is added when one is asked to attend to the statement that is believed less (that is, the mismatching model) further supports the view that the MMT parallels the minimal change principle.

## 6.5. Limitations of current research and future directions

The findings of the seven experiments have accumulated to a proposal of dual processing of belief revision strategies: a global logic-based strategy based on mental models and a specific probability-based strategy. Before it can be afforded the status of a theory of human belief revision, one should speculate on an agenda for future works. For example, the current investigation does suffer from a few methodological shortcomings that shall be pointed out with suggestions to overcome them. Furthermore, future directions are offered to increase knowledge about human belief revision. For example, deeper investigation in the current results is needed. The research should expand to involve more diverse complex thinking areas. Also, the study on belief revision should “step out” the laboratory and involve more ecologically valid belief revision encounters.

## 6.5.1. Better realignment of former studies with current studies on belief revision

### 6.5.1.1. Science fiction versus real

People are capable of reasoning not only about the real world, but also about a hypothetical or fictional world (Garnham & Oakhill, 1994). The belief values assigned to propositions expressing these different situations, however, may differ greatly.

It was argued here that the unfamiliar inference problems are inherently different from unfamiliar inference problems in former research (Byrne & Walsh, 2005; Elio & Pelletier, 1997; Politzer & Carles, 2001) in that the former are realistic and the latter science fictional. The unfamiliar material in Experiment IV pertained to realistic everyday actions. Despite its unfamiliarity, it was clear to the participants that the unfamiliar problems expressed situations that do really occur. In overall, the majority of the current experiments showed a higher belief in the conditional than in the categorical. In studies that applied science fiction material, the categorical was more favourable (Byrne & Walsh, 2005; Elio & Pelletier, 1997; Politzer & Carles, 2001). Results from Experiment V, which utilized abstract material, showed a lower belief in the conditional than in all other experiments reported here. These abstract conditionals would be more similar to the situations described in science-fiction material, in that in both cases one cannot imagine the situation. Nevertheless, a more direct comparison should clarify whether people recognize the difference between unfamiliar but real-life conditionals and science-fiction-based conditionals and adjust their belief revision accordingly. A future study might compare MP and MT problems with three conditions: science-fiction-problems, unfamiliar but realistic problems and familiar problems.

### 6.5.1.2.. From the particular to the general

In past research on belief revision, most often general statements were used and most experiment used science-fiction material. The common finding was that the categorical

was believed more than the conditional statement. The current set of experiments always applied very particular statements referring to a single person. The recurrent finding in these experiments, in contrast, was that the conditional favoured more belief than the categorical statement. In contrast to most former studies, daily life situations were used. This most likely influenced the difference. However, another plausible factor in play is the difference between generality and particularity. Revlis et al (1971), for example, found that verification of statements with particular quantifiers is easier than of statements with universal quantifiers. Since a general statement has to be true in all instances, it is more likely to be erroneous than a particular statement. People might pick up on this logic and will therefore be less prone to invest belief in a general statement. However, a direct test of this hypothesis has not been done yet. One way to look into this is to perform an experiment in which the conditionals of the inference problems are either general statements or particular statements. The inference problems can consist of medium-probability familiar MP and MT problems. A 'general' conditional will refer to people in general (denoted by 'a person'). A 'particular' conditional will refer to a single person denoted by its first name. It is expected that the mismatch principle will be applied in both conditions. However, the conditional will be believed more to a significant extent in the 'particular' condition than in the 'general' condition.

The effect of generality should also be examined in the case where another belief revision strategy was used other than mental models, namely plausibility. Material can consist of MP and MT problems whereby the conditional statement expresses either a very plausible or a very implausible daily action. It is expected that in both conditions, belief revision is guided by plausibility. That is, when the conditional expresses a high plausible action, then people will belief the conditional more. Conversely, when the conditional contains a low plausible action, people will belief the categorical statement more. However, the factor generality will cause an interaction in that the difference in believing the conditional between plausible and implausible inference problems will be significantly greater for general inference problems than for particular inference problems.

If it is the case that generality has a mediating factor in the former two experiments, then a follow-experiment can be performed wherein the factor 'generality' will be coupled

with an already explored mediating factor, namely familiarity. This would results in the following conditions: MP general-familiar, MP general-unfamiliar, MP particular-familiar, MP particular-unfamiliar, MT general-familiar, MT general-unfamiliar, MT particular-familiar, MT particular-unfamiliar problems. An interaction effect is expected to show up, such that generality will have a different influence on MP and MT problems when combined with familiarity. For MP problems, the mediating factors familiarity and generality will move in the same direction. That is, the mismatch principle will reveal itself strongest with unfamiliar-general inference problems and the least strong with familiar-particular inference problems. For MT problems, the mismatch principle will reveal itself strongest in the familiar-particular condition and the least strong in the unfamiliar-general condition. Such a series of experiments will also reveal possible interaction effects, additive effects, and attenuation effects. In the current experiments, each content and context factor was independently contrasted with the mismatch principle. It would be of interest to compare those content and context factors against each other.

## 6.5.2. Deeper investigation into the involvement of the two strategies in belief revision.

### 6.5.2.1. Involvement of mental models

Methodologically, the support for the mental model theory needs to be further scrutinized. It did not serve the status of an independent variable as it was not operationalized as the factors probability, familiarity, and source trustworthiness were. Instead, it serves an abstract theoretical construct. However, from the point of view that the MP and MT arguments are theoretically motivated by the MM theory, further testing could bring closer "prove" for the role of the mismatch principle in belief revision. A consideration that needs to be taken into account is the final mental model that people constructed in the case where there seemed to be support for the mismatch principle. The principle of truth in the mental model theory implies that the 'p & not-q' case is a

false possibility and is not even constructed in people's minds (Johnson-Laird & Byrne, 1991; 2002; Evans & Over, 2003). But it was found here that when the categorical was believed more than the conditional, then people actually do accept 'p & not-q' cases; with MP inference problems the categorical 'p' is then believed together with the fact 'not-q' and with MT the categorical not-q is believed together with the fact 'p'. One way to resolve this, is to present the belief revision problems on paper and ask subjects to write down arguments for their belief preference choices. Another option is to include more belief revision options. Besides the options "belief the conditional more" and "belief the categorical more", subjects should have the possibility to have no belief preference. An experiment in which people receive a third answer option, that of 'my degree of belief in either statement is very low', should in the case of medium probability be more often chosen.

A second way to resolve the confusion is to include AC and DC inference problems. In both cases, if the categorical is believed more, then the final mental model would be 'not-p and q' which would coincide with one of the three true possibilities. If people actually do consider this latter possibility as a true one and 'p and not-q' as a false possibility, then the rate for choosing the categorical on AC and DA problems should be significantly higher than on MP and MT problems. To test that, they should be incorporated in a within-subjects design with medium-probability problems.

### 6.5.2.2. Involvement of the probability theory

The design of the experiments allowed to refute the claim made by the MMT that 'p & not-q' cases are not taken into account by people. The design did not allow, however, the possibility to refute the proposition made by probability theories that 'not-p' states are irrelevant in that people do not consider them (Evans, et al., 2003; Oberauer & Wilhelm, 2003; Over & Evans, 2003; Over et al., 2007). However, Evans et al. (2003) demonstrated that when the frequency of ¬p increased, rating of the probability of the conditional decreased. To test whether people consider ¬p cases in belief revision tasks, DA and AC problems should also be tested, in a similar as they could be used for testing the mental model theory more deeply. With typical belief revision tasks, the contradiction with AC would be not-p and for DA the categorical statement would be the not-p case. If the

probability theories were correct in that people do not consider not-p cases, then people are expected to invest more belief in the conditional statement with both AC and DA problems (whereby probability is controlled) than with MP and MT problems.

Finally, the probability account was tested in the current investigation under the assumption of the Ramsey test which evaluates a conditional sentence as what the probability is for q given p. According to the Ramsey test, the antecedent is first entered into the existing belief state. A brief 'belief check' is first performed on the antecedent before introducing the consequent to the belief base. If the antecedent contradicts existing beliefs, a minor change as possible is made to the antecedent to reach consistency. Only then one adds the consequence to his or her belief state and considers whether the full statement can be accepted. In the current investigation only so-called open conditionals were used (Gärdenfors, 1992) where the antecedent p readily fits with a person's knowledge stock; the antecedent on its own contained little or no uncertainty. By having done so, in theory only the second step of the Ramsey test was tested. In experiments V-VII, the source induced uncertainty. However, the mismatch principle played the lead role in these inference problems. Moreover, the interest lies with problems whose content differs in level of probability. To fully test the Ramsey test that underlies the probability account, a feasible follow-up investigation would be to induce uncertainty in the antecedent by giving the antecedent an improbable action such as:

If Mark takes the hot-air balloon to work, then he takes along his briefcase
Mark takes the hot-air balloon to work
Mark does not take along his briefcase.

Another way to fully test the Ramsey Test is to examine counterfactuals, which are conditionals whereby the antecedent contradicts what is in the belief state (Gärdenfors, 1992; Thompson & Byrne, 2002; Walsh & Byrne, 2004)). By using counterfactuals, the plausibility of the antecedent can be varied. This way, both antecedent and consequent seem very likely or very unlikely.

### 6.5.3. Testing the dual-process account further

A dual-process account was offered to attempt to fit the results in a possible belief revision theory. Although many aspects need further consideration to lay a solid basis for a dual-process theory of belief revision, special attention will be devoted here to the commonly reported heuristic/analytic demarcation between two types of reasoning (Chaiken, 1980; Chaiken, et al., 1989; Evans & Over, 1996; Petty & Cacioppo, 1981, 1986; Sloman, 1996).

#### 6.5.3.1. Testing a possible heuristic/analytic connotation of a dual-process account of belief revision

The dual-process theories view abstract rule or model-based thinking as analytic characterized by slow, conscious, and controlled processing. Heuristic thinking is thought to be devoid of deliberate thinking and is characterized by fast, automatic, and unconscious processing (e.g. Evans & Over, 1996). Even though some commonalities between the belief revision strategies and the dual-process modes are pointed out, it is not the intention at this point to attach different connotations of processing to the two belief revision strategies. To find out whether this connotation of an evaluative segregation (Smith & DeCoster, 2000) also holds for belief revision strategies, further research is needed.

Although the different means by which people revise their belief are considered strategies in their own right, the possibility of relying on prior knowledge and belief as a shallow type of processing should be ruled out. Claiming that people reason on the grounds of belief, experience, and knowledge can explain why they make errors on logically constructed deductive inference problems. However, this does not automatically imply that reasoning based on probability is not free of biased thinking, which is a heuristic style of thinking that does not always properly suit the occasion or situation. A good candidate to test the possibility of heuristic processing is an experiment that uses the context factor 'source trustworthiness'. The content used in experiments V-VII had no personal relevance to the people tested. A direct test of whether trustworthiness constitutes a heuristic style of thinking is to use material varied on the scale of personal

relevance uttered by persons of varying levels of trustworthiness. When the statement with the high-trustworthy source is preferred irrespective of personal relevance of the statement, then this would support the notion that basing message preference on source trustworthiness is more of a heuristic kind of processing. Another factor to look into in this regard is testing belief revision with counterfactual reasoning problems. Counterfactual reasoning is regarded as a function of the rule-based system (Clark, 1997, as cited in Smith & DeCoster, 2000).

Another way to test heuristic thinking is to borrow work from the heuristics and biases approach by Kahneman, Slovic, and Tversky (1982). This includes a representativeness heuristic that leads to the conjunction fallacy. With the *representativeness heuristic* is meant that " people often evaluate the probability of an uncertain event or a sample by the degree to which it is (i) similar in essential properties to its parents population, and (ii) reflects the salient features of the process by which it is generated" (Tversky & Kahneman, 1982, pp.84-85). By conjunction effect is meant that a conjunctive event is judged as more probable than either one of its components. Kahneman and colleagues tested this by presenting participants with a short personality sketch followed by eight possible statements, reporting the person's possible occupation, which had to be ranked by their probability. These included a representative outcome, an unrepresentative outcome, and the conjunction of the two (Kahneman & Tversky, 1982). The compound statement was judged as more probable than the critical simple statement. Such a statistical reasoning problem can be applied and modified to a belief revision problem to find out if people's probabilistic thinking relates more to realistic probabilities or to illusory probabilities. The role of illusory correlations in belief revision can be tested by contrasting real-life probability-based belief revision problems with illusory probability- based belief revision problems. In an evaluation study, subjects can be asked to judge for example how likely two traits match together. Based on that, inference problems can be constructed based on either correct high-probabilistic information, intuitively compelling but incorrectly high probabilistic information, or correct but medium-probabilistic information. An example of an intuitively compelling but incorrectly high-probability inference problem is 'If Linda is concerned with issues of

discrimination, then she is active in the feminist movement, Linda is concerned with issues of discrimination, Linda is not active in the feminist movement'.

### 6.5.4. Rationality of belief revision: From belief revision to action/goals

An important issue that should not be neglected is the role of beliefs in goal processing. The purpose of the principles of valid inference of logic is to dictate what must follow from premises that are assumed to be true. Logic does not go beyond this; it is an end in itself. In everyday practice, by contrast, valid inferences guide people to achieve a further end or goal, e.g. constructing a theory in science (Evans & Over, 1996). To this extent, the literature makes a distinction between two kinds of rationality, as follows (taken from Evans & Over, 1996, pp. 8):

> *Rationality1:* Thinking, speaking, reasoning, making a decision, or acting in a way that is generally reliable and efficient for achieving one's goals.
>
> *Rationality2:* Thinking, speaking, reasoning, making a decision, or acting when one has a reason for what one does sanctioned by a normative theory.

This distinction between rationality 1 and rationality 2 is reflected in the current findings. When belief revision was a function of the mismatch principle (e.g. Johnson-Laird & Byrne, 2002), then the belief preference was different for MT and MP inference problems, even though both problems contained exactly the same information. When probability was the guiding factor for belief revision, then the same premise was believed across MP and MT problems.

Belief revision determines goal processing and intention revision. In order to achieve a goal one has to modify the appropriate beliefs. However, the contribution of beliefs to one's behaviour depends heavily on the goals one is assessing or pursuing. Goals of the kind mentioned here are called "achievement goals" opposed to "maintenance goals". The former concern goals about state of interests that have not yet been realized whereas the latter are goals directed at keeping things the way they are. Furthermore, both relevant

and irrelevant beliefs are in play in the cognitive regulation of action (Castefranchi & Paglieri, 2007). In general, probability-based belief revision is more readily activated by relevant beliefs leading to achievement goals. Relevant belief might have been in play as the participants coupled the conditional's content to their own knowledge and belief, However, as the material in the current investigation was not controlled for personal relevance, the belief revision problems may not have triggered achievement goals. Since belief revision is rarely detached from goals and desires, it would be of interest to examine the role of beliefs in goal processing. A series of MP and MT medium-probability problems can be presented whereby each single inference problems contains one consecutive step of a final goal presented in the last inference problem. These inference problems should be presented in chronological order intervened by irrelevant inference problems (not pertaining to the goal) which will be randomly presented across participant. It is expected that the belief revision on the first few 'goal-directed' inference problems will be solved by means of the mismatch principle but that as the steps increase more belief will be invested in the conditional as the actions in the conditional should lead to the goal. A control group should receive both the irrelevant and the goal-directed inference problems in random order. A follow-up experiment can contrast goal-directed inference problems of personal relevance with goal-directed inference problems of no personal-relevance.

In the current framework of goal processing and action in belief revision, it would be also interesting to unravel whether there is a sustainable effect of source trustworthiness on belief revision. Numerous studies in social psychology have identified a so-called 'sleeper-effect' in people's message processing (Hovland & Weiss, 1951; Kelman & Hovland, 1953; Kumkale & Albarracín, 2004; Pratkanis, Greenwald, Leippe, & Baumgardner, 1988). The sleeper-effect refers to the phenomenon that people who were initially dissuaded by a message due to an untrustworthy source will after a lapse of time show 'liking' in the message. This is because people presumably remember the statement or message but forget the source of the message. The time that people belief one statement over the other is of valuable interest because in real-life one endorses a newly gained belief for a longer period of time. Future studies might look into the

duration of endorsement and rejection of one newly encountered piece of information over another.

### 6.5.5. Other factors that might influence belief revision

There are still a plethora of factors that are of interest to the study of human belief revision. Two factors that are of direct interest with respect to the current studies will be outline here.

#### 6.5.5.1. Personal relevance: relevant and irrelevant beliefs in the cognitive regulation of action.

In experiments I and II it was found that the level of plausibility has great influence on how people revise their beliefs. That is, when the conditional statement in a deductive inference problem is either very plausible or very implausible, the factor plausibility guides people's belief revision. When the material is near-50%, then people will convert to the use of mental models to reach a belief revision choice. However, many factors have not yet been looked into that might dampen this latter finding. One such factor is personal relevance. Research from the area of social cognition has repeatedly shown that when faced with new information, people are (differentially) influenced by this information via different routes depending on whether the information is personally relevant (e.g. Petty & Cacioppo, 1986). In the current experiments the participants learned about the two uncertain statements shortly before being confronted with the new fact. By doing so, the belief state was in a way artificially created in the person's mind. It would be interesting also to tap into belief states already vastly held by people and find out how these belief states are updated in the face of contravening information.

Pre-attitude on different topics can first be tested with an evaluation study in which students are presented with different topics which they have to rate for personal relevance. From these general topics the most relevant and the most irrelevant topic are chosen for a second evaluation study. From both topics, statements are constructed that

relate to the respected topic. The second evaluation study will consist of these statements that again are rated for relevance by a group of students. The statements are also rated for plausibility. Six or eight most relevant and irrelevant statements are chosen for the experiment and MP and MT problems are constructed from both groups of statements. The expectation is that with both relevant and irrelevant inference problems the mismatch principle will be applied. However, in overall, the conditional will be believed more with relevant than irrelevant inference problems to a significant extent.

### 6.5.5.2. Expertise

Source expertise and source trustworthiness are both considered to fall under the umbrella of source credibility. Experiments V-VII focused on the effect of source trustworthiness on belief revision. In experiments V and VI, source trustworthiness was manipulated such that it represented occupations. Each premise within each inference problem was uttered by someone who was presented with either a high or a low trustworthy occupation. In experiment V, abstract material was used in that the conditional read 'If A is the case, then B is the case'. In experiment VI, the abstract material was replaced by statements that could likely be said in a possible conversation between the two persons. In experiment VII, source trustworthiness was manipulated by using character descriptions. The two persons uttering either one of the statements with each inference problems, presented by their first name, were introduced prior to the start of the experiment by either a low or a high trustworthy personality description. Only in the first and the third experiment, a main effect of source trustworthiness showed up. The overall belief revision strategy was the mismatch principle, but the conditional was believed less when it was uttered by a low-trustworthy person then when it was uttered by a high-trustworthy person. This effect failed to show in the second experiment. A possible confound might have been source expertise. Source expertise is an important factor that should not be ruled out.

An evaluation study can be performed where occupations are coupled with statements that fit the occupation. A group of students will rate for each statement how possible it is and the expertise level of the person with respect to the statement. The experiment

should consist of the following three conditions: high expertise/low expertise, low expertise/high expertise, medium expertise/medium expertise. Based on findings in the literature (Stevenson & Over, 2001; Wiener & Mowen, 1986), it is expected that expertise will dominate the belief revision process. The factor 'Source trustworthiness' can be added in a follow-up experiment. It is expected that this latter factor will serve an underlying role to expertise.

### 6.5.6. External validity

Belief revision was investigated in the context of deductive inference problems. The research of belief revision should go beyond the confines of deductive reasoning. It is not certain whether the factors that were found here to influence belief revision also apply outside the laboratory in more real-life contexts. By using deductive inference problems, the setup of the contradicting statements permitted binary choices only, which restricts the generality of the findings to everyday conversational contexts. In many circumstances, contradictions occur between multiple statements uttered by either two or more people. However, in the current developmental stage of the existing human deduction and belief revision theories it is difficult to test otherwise. Nevertheless, it is in any case an interesting aspect to look into further in upcoming research. Investigating belief revision in more complex conversational contexts could extend these theories or plant the seeds from which additional theories grow.

Moreover, since content and context factors substantially impact mode of belief revision it would be worthwhile to test people who vary more on the demographic scale, for example people of different age groups or of varying cultures.

## 6.6. Main Conclusion

In conclusion, the current set of experiments showed that people have different belief revision strategies at their disposal. The propensity to belief one statement over the other varied as a function of probability as well as mental models. Support for probability theories was found in that people assign probability estimates to the

conditionals and this transforms into a belief revision strategy when the conditional statement expresses a very plausible or implausible action. The mismatch principle, on the other hand, is applied when the conditional statement conveys a near-chance level plausibility. The conditional was believed more with MT problems and the categorical was favoured with MP problems. The factors familiarity, task instruction, and source trustworthiness served as mediating factors. The DTs also mirrored the different belief revision strategies. Thus, there was the presence of mutually exclusive belief revision strategies. People integrate into their belief revision process domain-specific content and context factors as well as domain-general mental models. The DTs offered the further interesting insight that people not only employ different belief revision strategies but that these, depending on the material, can facilitate but also hamper the process of belief revision. This creates the picture that although the mental-model approach and the probability approach are theoretically two distinct belief revision strategies, psychologically they are characterized by two independent but also interacting cognitive operations. That is, they both reside in the human mind whereby they can aid but also conflict with one another. Equally interesting, the DT trend showed that no strategy was in particular easier than the other. These findings corroborate with theories emphasizing psychological processes as well as theories that focus on abstract processes in reasoning and belief revision. Even though it is to early at this point in time to establish a firm theory of belief revision, the findings do lead into the direction of a possible dual-process account of belief revision.

A derivation of the research is that the study on human belief revision would profit from diverting from AI theories. Instead, a more fruitful approach would be to integrate the gathered knowledge from the different domains of human psychology, such as cognitive and social psychology, into the research of human belief revision. Moreover, for further development of a human belief revision theory, scholars in the field of human reasoning might profit from disengaging from debating what single belief revision strategy people use and instead turn to identifying the number of possible different strategies that exist and what factors influence which one strategy dominates in any particular situation. So far we have demonstrated that probability, mental models, familiarity, task instruction, and source trustworthiness influence belief revision. What

orbit of other factors exist is an interesting research question awaiting further investigation. As a final note, since there is no correct or logic answer possible with belief revision tasks, the choice patterns people adopt on these tasks will not only help to increase knowledge of human belief revision but can be placed in a broader context by unfolding new reasoning strategies that can in turn be integrated in research on judgment and decision making and problem solving.

# Chapter 7

# References

Alchourrón, C.E., Gärdenfors, P., & Makinson, D. (1985). On the logic of theory change: Partial meet contraction and revision functions. *The Journal of Symbolic Logic, 50*, 510-530.

Areces, C., & Becher, V. (2001). Iterable AGM functions. In M.-A. Williams & H. Rott (Eds.), *Frontiers in belief revision*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Bochman, A. (2001). Contraction of epistemic states: A general theory. In M.-A. Williams & H. Rott (Eds.), *Frontiers in belief revision*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Braine, M.D.S. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review, 85*, 1-21.

Braine, M.D.S., & O'Brien, D.P. (1991). A theory of If: Lexical entry, reasoning program, and pragmatic principles. *Psychological Review, 98*, 182-203.

Byrne (1989). Suppressing valid inferences with conditionals. *Cognition, 31*, 61-83.

Byrne, R.M.J., & Walsh, C.R. (2002). Contradictions and counterfactuals: generating belief revisions in conditional inference. In W.D. Gray & C.D. Schunn (Eds). Proceedings

of the 24th Annual Conference of the Cognitive Science Society (pp. 160-165). Mahwah, NJ:Erlbaum.

Byrne, R.M.J., & Walsh, C.R. (2005). Resolving contradictions. In V. Girotto &, P.N. Johnson-Laird (Eds.), *The shape of reason: essays in honour of Paolo Legrenzi*. Hove: Psychology Press.

Castelfranchi, C., & Paglieri, F. (2007). The role of beliefs in goal dynamics: Prolegomena to a constructive theory of intentions. *Synthese, 155*, 237-263.

Cheng, P.W., & Holyoak, K.J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology, 17*, 391-416.

Cheng, P.W., & Holyoak, K.J., Nisbet, R.E., & Oliver, L.M. (1986). Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology, 18*, 293-328.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39*, 752-66.

Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic information processing within and beyond the persuasion context. In J. S. Uleman, & J. A. Bargh (Eds.), *Unintended thought* (pp. 212-252). Guilford, New York.

Chater, N., & Oaksford, M. (2001). Human rationality and the psychology of reasoning: Where do we go from here? *British Journal of Psychology, 92,* 193-216.

Chun, K.-T., Campbell, J. B., & Yoo, J.-H. (1975). Perceived trustworthiness of occupations. Personality effects and cross-cultural generalizability. *Journal of Cross-Cultural Psychology, 6*, 430-443.

Clark, H.H. (1969). Linguistic processes in deductive reasoning. *Psychological Review, 76*, 387-404.

Cummins, D.D. (1995). Naive theories and causal deduction. *Memory & Cognition, 23*, 646-658.

Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory and Cognition, 19,* 274-282.

Dieussaert, K., Schaeken, & d'Ydewalle, G. (2002). The relative contribution of content and context factors on the interpretation of conditionals. *Experimental Psychology, 49*, 181-195.

Dieussaert, K., Schaeken, W., Neys, W. De, & d'Ydewalle, G. (2000). Initial belief state as a predictor of belief revision. *Current Psychology of Cognition, 19*, 277-286.

Ditmarsch, H. van, & Labuschagne, W. (2007). My beliefs about your beliefs: A study in theory of mind and epistemic logic. *Synthese, 155*, 191-209.

Doyle, J. (1979). A truth maintenance system. *Artificial Intelligence, 12*, 231-272.

Doyle, J. (1992). Reason maintenance and belief revision. Foundations vs. coherence approach. In P. Gärdenfors (Ed.), *Belief revision* (pp. 29-51). Cambridge: Cambridge University.

Dragoni, A.F.,& Giorgini, P. (2001). Revising beliefs received from multiple sources. In M.-A., Williams & H. Rott (Eds.), *Frontiers in belief revision*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Dubois, D. (2008). Three scenarios for the revision of epistemic states. *Journal of Logic and Computation, 18*, 721-738.

Dubois, D., and Prade, H. (1992). Belief change and possibility theory. In: *Belief Revision*, P. Gärdenfors, ed., pp. 142-182. Cambridge University Press, Cambridge UK.

Dubois, D., & Prade, H. (1993). Possibility theory, belief revision and nonmonotonic logic. *Proceeedings of the Workshop on Fuzzy Logic in Artificial Intelligence*, Vol. 847. Lecture Notes in Computer Science. (pp. 51-61). London, UK: Springer-Verlag.

Eagly, A. H., Wood, W., & Chaiken, S. (1978). Causal inferences about communicators and their effect on opinion change. *Journal of Personality and Social Psychology, 36*, 424-435.

Elio, R. (1997). What to believe when inferences are contradicted: The impact of knowledge type and inference rule. *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (pp. 211-216). Mahwah, NJ: Lawrence Erlbaum Associates.

Elio, R., & Pelletier, F.J. (1997). Belief change as propositional update. *Cognitive Science, 21,* 419-460.

Evans, J.St.B.T. (1991). Theories of human reasoning: The fragmented state of the art. *Theory & Psychology, 1*, 83-105.

Evans, J.St.B.T. (2002). Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin, 128*, 978-996.

Evans, J.St.B.T. (2003). In two minds: Dual-process accounts of reasoning. *TRENDS in Cognitive Sciences, 7,* 454-459.

Evans, J.St.B.T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review, 13*, 378-395.

Evans, J.St.B.T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*, 255-278.

Evans, J.St.B.T., Handley, S.J., Harper, C.N.J., & Johnson-Laird,, P.N. (1999). Reasoning about necessity and possibility: A test of the mental model theory of deduction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1495-1513.

Evans, J.St.B.T, Handley, S.J., & Over, D.E. (2003). Conditionals and Conditional Probability. *Journal of Experimental psychology: Learning, Memory, and Cognition, 29*, 321-335.

Evans, J.St.B.T., Newstead, S.E., Allen, J.L. & Pollard, P. (1994). Debiasing by instruction: The case of belief bias. *European Journal of Cognitive Psychology, 6,* 263-285.

Evans, J.St.B.T., Newstead, S. and Byrne, R.M.J. (1993). Human reasoning: The Psychology of Deduction. Hove, UK, and Hillsdale, NJ: Lawrence Erlbaum Associates. pp.310.

Evans, J. St. B. T., & Over, D. E. (1996). *Rationality and Reasoning*. Psychology Press, Hove.

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

Evans, J.St.B.T., Over, D.E., & Handley, S.J. (2005). Suppositions, extensionality, and eonditionals: A critique of the mental model theory of Johnson-Laird and Byrne (2002). *Psychological Review, 112, 1040-1052.*

Evans, J. St, B., T., Over, D.E., and Manktelow, K. I. (1993). Reasoning, decision making, and rationality. *Cognition, 49*, 165-187.

Garnham, A. & Oakhill, J. (1994). *Thinking and Reasoning*. Oxford, UK: Blackwell Publishers Ltd.

Gärdenfors, P. (1988). *Knowledge in flux*. Cambridge, MA: MIT Press.

Gärdenfors, P. (1992). Belief Revision: An introduction. In P. Gärdenfors (Ed.), *Belief revision* (pp. 1-28). Cambridge: Cambridge University Press.

George, C. (1997). Reasoning from uncertain premises. *Thinking and reasoning, 3,* 161-190.

George, C. (1999). Evaluation of the plausibility of a conclusion derivable from several arguments with uncertain premises. *Thinking and reasoning, 5,* 245-281.

Gilbert, D.T. (1991). How mental systems believe. *American Psychologist, 46,* 107-119.

Girotto, V., & Johnson-Laird, P.N. (2004). The probability of conditionals. *Psychologia, 47*, 207-225.

Girotto, V., & Johnson-Laird, P.N. (2004). The probability of conditionals. *Psychologia, 47*, 207-225.

Girotto, V., Johnson-Laird, P.N., Legrenzi, P., and Sonino, M. (2000). Reasoning to consistency: How people resolve logical inconsistencies. In Garcia-Madruga, J., Carriedo, M, and Gonzalez-Labra, M. J. (Eds.) *Mental Models in Reasoning*. Madrid: UNED. Pp. 83-97.

Harman, G. (1986). *Change in view*. Cambridge, MA: MIT Press.

Hasson, U., and Johnson-Laird, P. N. (2003). Why believability cannot explain belief revision. In Alterman, R., and Kirsh, D. (Eds.), *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Hasson, U., Simmons, J.P., & Todorov, A. (2005). Believe it or not. On the possibility of suspending belief. *Psychological Science, 16,* 566-571.

Holyoak, K.J., & Spellman, B.A. (1993). Thinking. *Annual Review of Psychology, 44*, 265-315.

Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly, 15*, 635-650.

Johnson-Laird, P.N. (1972). The three term series problems. *Cognition, 1,* 58-82.

Johnson-Laird, P.N. (2006). Beliefs, heresies, and changes in mind. In P.N. Johnson-Laird (Ed.), *How we reason*. New York: Oxford University Press.

Johnson-Laird, P.N., & Byrne, R.M.J. (1991). *Deduction.* Hove, UK: Erlbaum.

Johnson-Laird, P.N., & Byrne, R.M.J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review, 109*, 211-228.

Johnson-Laird, P.N., Byrne, R.M.J., & Schaeken, W. (1992). Propositional reasoning by model. *Psychological Review, 99*, 418-439.

Johnson-Laird, P.N., Girotto, V., Legrenzi, P. (2004). Reasoning from inconsistency to consistency. *Psychological Review, 111*, 640-661.

Johnson-Laird, P. N., & Hasson, U. (2003). Counterexamples in sentential reasoning. *Memory & Cognition, 31*, 1105-1113.

Johnson-Laird, P. N., Legrenzi, P., & Girotto, V. (2004). How we detect logical inconsistencies. *Current Directions in Psychological Science, 13*, 41-45.

Johnson-Laird, P.N., Legrenzi, P., Girotto, V., & Legrenzi, M.S. (2000). Illusions in reasoning about consistency. *Science, 288*, 531-532.

Johnson-Laird, P.N., Legrenzi, P., Girotto, V., Legrenzi, M.S., and Caverni, J-P. (1999). Naive probability: a mental model theory of extensional reasoning. *Psychological Review*, 106, 62-88.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

Kelman, H. C., & Hovland, C. I. (1953). "Reinstatement" of the communicator in delayed measurement of opinion change. *Journal of Abnormal and Social Psychology, 48*, 327-335.

Kern-Isberner, G. (2001). Revising and updating probabilistic beliefs. In M.-A. Williams & H. Rott (Eds.), *Frontiers in belief revision*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Kumkale, G. T., & Albarracín, D. (2004). The sleeper effect in persuasion. *Psychological Bulletin, 130*, 143-172.

Kyburg, H.E.Jr. (1983). Rational belief. *Brain and behavioral Sciences, 6*, 231-273.

Kyburg, H.E.Jr. (1984). Believing on the basis of evidence. *Computational Intelligence, 10*, 3-20.

Legrenzi, P., Girotto, V., & Johnson-Laird, P.N. (2003). Models of inconsistency. *Psychological Science, 14,* 131-137.

Liu, I.-M., Lo, K.-C, & Wu, J.-T. (1996). A probabilistic interpretation of "If-Then". *The Quarterly Journal of Experimental Psychology, 49A*, 828-844.

Marcus, S.L., & Rips, L.J. (1979). Conditional reasoning. *Journal of Verbal Learning and Verbal Behaviour, 18*, 199-223.

Manktelow, K. (2004). *Reasoning and Thinking*. UK: psychology Press Ltd.

Markman, K. D., & McMullen, M. N. (2003). A reflection and evaluation model of comparative thinking. *Personality and Social Psychology Review, 7*, 244-267.

Markovits, H. (1986). Familiarity effects in conditional reasoning. *Journal of Educational Psychology, 78*, 492-494.

Mills, J., & Jellison, J. M. (1967). Effect on opinion change of how desirable the communication is to the audience the communicator addressed. *Journal of Personality and Social Psychology, 6*, 98-101.

Mussweiler, T. (2003). Comparison processes in social judgment: Mechanisms and consequences. *Psychological Review, 110*, 472-489.

Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. *Trends in Cognitive Science, 5*, 349-357.

Oaksford, M., & Chater, N. (2009). Précis of bayesian rationality: The probabilistic approach to human reasoning. *Behavioral and Brain Sciences, 32*, 69-120.

Oaksford, M., Chater, N., & Hahn, U. (2008). Human reasoning and argumentation: The probabilistic approach. In J. Adler, & L. Rips (Eds.), *Reasoning: Studies of human inference and its foundations*. Cambridge: Cambridge 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*, 883-899.

Oberauer, K., & Wilhelm, O. (2003). The meaning(s) of conditionals: Conditional probabilities, mental models, and personal utilities. *Journal of Experimental psychology: Learning, Memory, and Cognition, 29*, 680-693.

O'Hara, B. S., Netemeyer, R. G., & Burton, S. (1991). An examination of the relative effects of source expertise, trustworthiness, and likability. *Social Behavior and Personality, 19*, 305-314.

Over, D.E., Evans J. St. B.T. (2003). The probability of conditionals: The psychological evidence. *Mind & Language, 18*, 340-358.

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

Petty, R. E., & Cacioppo, J. T. (1981). *Attitudes and persuasion: Classic and contemporary approaches*. W. C. Brown, Dubuque, IA.

Petty, R. E., & Cacioppo, J. T. (1986). The Elaboration Likelihood Model of persuasion. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology*. New York, New York: Academic Press.

Politzer, G. (2005). Reasoning, judgment, and pragmatics. In I. Noveck, & D. Sperber (Eds.), *Experimental Pragmatics* (pp. 94-115). London: Palgrave.

Politzer, G., & Bourmond, G. (2002). Deductive reasoning from uncertain conditionals. *British Journal of Psychology, 93,* 345-381.

Politzer, G., & Carles, L. (2001). Belief revision and uncertain reasoning. *Thinking and reasoning, 7,* 217-234.

Pratkanis, A. R., Greenwald, A. G., Leippe, M. R., & Baumgardner, M. H. (1988). In search of reliable persuasion effects: III. The sleeper effect is dead: Long live the sleeper effect. *Journal of Personality and Social Psychology, 54,* 203-218.

Priester, J. R., & Petty, R. E. (1995). Source attributions and persuasion: Perceived honesty as a determinant of message scrutiny. *Personality and Social Psychology Bulletin, 21*, 637-654.

Rhine, R. J., & Severance, L. J. (1970). Ego-involvement, discrepancy, source credibility, and attitude change. *Journal of Personality and Social Psychology, 16*, 175-190.

Rich, E. (2001). What is artificial intelligence? In E. Rich (Ed.), *Artificial intelligence*. Singapore: McGraw-Hill, Inc.

Rich, E. (2001). Knowledge representation using other logics. In E. Rich (Ed.), *Artificial intelligence*. Singapore: McGraw-Hill, Inc.

Richardson, J., & Ormerod, T.C. (1997). Rephrasing between disjunctive and conditionals: Mental models and the effects of thematic content. *The Quarterly Journal of Experimental Psychology, 50A,* 358-385.

Recanati, F. (1997). Can we believe what we do not understand? *Mind and Language, 12,* 84-100.

Revlin, R., Calvillo, D.P., & Ballard, S. (2005). Counterfactual reasoning: Resolving inconsistency before your eyes. *Psychologica Belgica, 45-1,* 47-56.

Revlin, R., Cate, C.L., & Rouss, T.S. (2001). Reasoning counterfactually: Combining and rending. *Memory & Cognition, 29*, 1196-1208.

Revlis, R. (1974). Prevarication: Reasoning from false assumptions. *Memory & Cognition, 2*, 87-95.

Revlis, R, & Hayes, J.R. (1972). The primacy of generalities in hypothetical reasoning. *Cognitive Psychology, 3*, 268-290.

Revlis, R., Lipkin, S.G., & Hayes, J.R. (1971). The importance of universal quantifiers in a hypothetical reasoning task. *Journal of Verbal Learning and Verbal Behaviour, 10*, 86-91.

Rips, L.J. (1983). Cognitive processes in propositional reasoning. *Psychological Review, 90*, 38-71.

Rott, H. (2000). Two dogma's of belief revision. *Journal of Philosophy, 97*, 503-522.

Rotter, J. B., & Stein, D. K. (1971*).* Public attitudes toward the trustworthiness, competence, and altruism of twenty selected occupations. *Journal of Applied Social Psychology, 1*, 334-343.

Ryckman, R. M., & Sherman, M. F. (1974*)*. Locus of control and attitudes of workers and college students toward members of selected occupations. *Journal of Applied Social Psychology, 4*, 351-364.

Segerenberg, K. (2001). The basic dynamic doxastic logic of AGM. In M.-A. Williams & H. Rott (Eds.), *Frontiers in belief revision*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*, 3-22.

Smith, E. R., & DeCoster, J. (2002). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4,* 108-131.

Staudenmayer, H. (1975). Understanding conditional reasoning with meaningful propositions. In R.J. Falmagne (ed.), reasoning representation and process in children and adults. Oxford: Lawrence Erlbaum Associates.

Stevenson, R.J., & Over, D.E. (1995). Deduction from uncertain premises. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 48A*, 613-643.

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

Thompson, V.A. (1994). Interpretational factors in conditional reasoning. *Memory & Cognition, 22*, 742-758.

Thompson, V. A. (2000). The task-specific nature of domain-general reasoning. *Cognition, 76,* 209-268.

Thompson, V. A., & Byrne, R.M. (2002). Reasoning counterfactually: Making inferences about things that didn't happen. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 1154-1170.

Tormala, Z. L., Briñol, P., & Petty, R. E. (2006). When credibility attacks: The reverse impact of source credibility on persuasion. *Journal of Experimental Social Psychology, 42*, 684-691.

Tormala, Z. L., & Clarkson, J. J. (2008). Source trustworthiness and information processing in multiple message situations: A contextual analysis. *Social Cognition, 26*, 357-367.

Tormala, Z. L., & Petty, R. E. (2007). Contextual contrast and perceived knowledge: Exploring the implications for persuasion. *Journal of Experimental Social Psychology, 43*, 17-30.

Tversky, A., & Kahneman, D. (1982). Judgments of and by representativeness. *Science*, 185, 1124-1131. Reprinted in part in D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

Van der Henst, J.B. (2002). Mental model theory versus the inference rule approach in relational Reasoning. *Thinking and Reasoning, 8*, 193-203.

Verscheuren, N., Schaeken, W., & d'Ydewalle (2004). A dual-process theory on everyday conditional reasoning. Technical Report Nº307, Laboratory of Experimental Psychology, Leuven: Belgium.

Walsh, C.R., & Byrne, R.M.J. (2004). Counterfactual thinking: The temporal order effect. *Memory & Cognition, 32*, 369-378.

Wason, P.C. (1966). Reasoning. In B.M. Foss (Ed.), *New horizons in psychology*, Vol. I Harmondsworth: Penguin.

Wiener, J. L., & Mowen, J. C. (1986). Source credibility: On the independent effects of trust and expertise. In R. J. Lutz (Ed.), *Advances in Consumer Research* (pp.306-310).

# Eidesstattliche Erklärung

„Ich erkläre: Ich habe die vorgelegte Dissertation selbständig und ohne unerlaubte fremde Hilfe und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis“ niedergelegt sind, eingehalten.“

Gießen, den 12. Dezember 2010

_____

(Ann Gabriella Wolf)