Should the person be punished?
Defeating conclusions from legal conditionals

Inaugural-Dissertation
zur Erlangung des akademischen Grades
Doctor rerum naturalium (Dr. rer. nat.)

an der

Justus-Liebig-Universität Gießen
Fachbereich 06:
Psychologie und Sportwissenschaften
Otto-Behaghel-Strasse 10F
35394 Gießen

vorgelegt von

Dipl.-Psych. Lupita Estefania Gazzo Castañeda
geboren am 08.03.1986
in Lima, Perú

März, 2016

Supervisor and First Reviewer:

Prof. Dr. Markus Knauff

Second Reviewer:

PD Dr. Dr. Marco Ragni

Declarations

Parts of this thesis have been published previously in different peer-reviewed journals and presented at peer-reviewed conferences. To facilitate readability I decided to embed and adapt these publications as chapters of a monograph, instead of a cumulative thesis. In the following I declare which material has been published previously and where:

- Chapter 3 has been published in a slightly different form in:

  Gazzo Castañeda, L. E., & Knauff, M. (2016). Defeasible reasoning with legal conditionals. *Memory & Cognition, 44*, 499-517. doi: 10.3758/s13421-015-0574-7 (with permission of Springer)

- Chapter 5 has been published in a slightly different form in:

  Experiment 6-7:

  Gazzo Castañeda, L. E., & Knauff, M. (2016). When *will* is not the same as *should*: The role of modals in reasoning with legal conditionals. *The Quarterly Journal of Experimental Psychology, 69*, 1480-1497. doi: 10.1080/17470218.2015.1085067

  Experiment 8:

  Gazzo Castañeda, L. E., & Knauff, M. (2015). Defeasible reasoning with quantifiers. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 770-775). Austin, TX: Cognitive Science Society.

- Chapter 6 has been published in a slightly different form in:

  Experiment 9:

  Gazzo Castañeda, L. E., Richter, B., & Knauff, M. (2016). Negativity bias in defeasible reasoning. *Thinking & Reasoning, 22*, 209-220. doi: 10.1080/13546783.2015.1117988

As a consequence of my embedding these publications in a monograph, there is also some overlap of this thesis' Aims and Hypothesis (Chapter 2) and General Discussion (Chapter 7) and the introductions and discussions of the respective publications.

Acknowledgments

First, I want to thank my supervisor Prof. Dr. Markus Knauff for giving me the possibility to investigate such an interesting topic were I could combine my two favorite psychological topics: cognition and legal psychology. Thank you also for your support and all the interesting and productive discussions. I want also to thank Dr. jur. Carsten Bäcker and Emma Harms for their advice on legal subjects. Next, I want to thank all my student assistants Lena Dienelt, Jessica Ewerhardy and Mohammad Nasiri for helping me collecting data, German proofreading, and interesting ideas. Further I want to thank all my colleagues for being there whenever I had questions or when I needed a break by going to the Mensa. And finally I also want to thank my friends, my family, and Victor for all the psychological support, love and the necessary times of distraction.

This research was supported by DFG grant KN 465/10-1 and KN 465/10-2 to Markus Knauff within the DFG Priority Program “New Frameworks of Rationality”.

Zusammenfassung

Anfechtbares Denken (d*efeasible reasoning*) beschreibt die Fähigkeit von Menschen, zuvor gezogene Schlüsse im Lichte neuer Information zu revidieren. Es ist besonders in der Rechtsprechung wichtig, weil dort strafausschließende Umstände dazu führen können, dass Richter schlussfolgern, dass eine strafbare Handlung nicht bestraft werden soll. Das Ziel dieser Arbeit ist es daher zu untersuchen, wie Menschen Schlussfolgerungen von rechtlichen Regeln revidieren. In einer Reihe von Experimenten wurden rechtliche Regeln als Konditionale präsentiert (z.B. „Wenn eine Person einen Menschen tötet, dann soll die Person wegen Totschlags bestraft werden“) und in Inferenzaufgaben zusammen mit potenziell strafausschließenden Umständen (z.B. Notwehr) eingebettet. Strafausschließende Umstände wurden entweder explizit als eine dritte Prämisse präsentiert (Experimente 1, 2, 4, 5) oder durch Vorstudien implizit erfasst (Experimente 6-8). Die Versuchsteilnehmer sollten entscheiden, ob der in der Inferenzaufgabe beschriebene Täter bestraft werden soll (Experimente 1, 2, 4-8) oder bestraft wird (Experimente 6-7). In Experiment 3 wurden die Versuchsteilnehmer aufgefordert strafausschließende Umstände selbst zu generieren. In allen Experimenten hatten die Versuchsteilnehmer kein rechtliches Vorwissen (d.h. Laien), aber in den Experimenten 1-3 wurden auch Juristen (d.h. fortgeschrittene Jura-Studierende oder Jura-Absolventen) getestet. Während Juristen beim Schließen den Regeln des Strafgesetzbuches folgten, hatte das Gerechtigkeitsempfinden von Laien einen Einfluss auf ihre Schlussfolgerungen. Wenn Laien gefragt wurden, ob ein Täter bestraft werden soll und die Straftat moralisch empörend war, ignorierten sie oft potentielle strafausschließende Umstände. In solchen Fällen hatten Laien sogar Schwierigkeiten selbst strafausschließende Umstände zu generieren. Nur wenn Laien danach gefragt wurden, ob ein Straftäter bestraft *wird*, konnten sie Ausnahmen für moralisch besonders verwerfliche Straftaten berücksichtigen. Des Weiteren konnte gezeigt werden, dass abhängig von den Einstellungen und Präferenzen der Teilnehmer manchmal rechtlich strafbare Taten nicht bestraft wurden. Zwei weitere Experimente (Experimente 9-10) zeigen, dass Menschen auch in Alltagssituationen oft Ausnahmen für emotional geladene Ereignisse ignorieren. Die Befunde sind für die kognitive Psychologie relevant, weil sie die Wichtigkeit von Vorwissen, Einstellungen und Präferenzen beim Denken zeigen. Außerdem sind sie für die Rechtswissenschaften, Sozialpsychologie und unserer Gesellschaft bedeutsam: die Ergebnisse zeigen, dass Paradigmen der kognitiven Psychologie verwendet werden können, um sozial relevante Konstrukte aus der Rechtstheorie und Sozialpsychologie zu testen.

Abstract

Defeasible reasoning is people's ability to withdraw previously drawn conclusions in light of new evidence. Defeasible reasoning is therefore especially important in law, where exculpatory evidence can bring judges to conclude that an offence should not be punished after all. The aim of this thesis was thus to investigate how people withdraw conclusions from legal rules. In a series of experiments, legal rules were presented as legal conditionals (e.g., "If a person kills another human, than the person should be punished for manslaughter") and embedded in inference tasks together with potentially exculpatory circumstances (e.g., self-defense). Exculpatory circumstances were presented either explicitly as a third premise (Experiments 1, 2, 4, 5), or captured implicitly via preliminary studies (Experiments 6-8). Participants had to decide whether the offender described in the inference task should (Experiments 1, 2, 4-8) or will be punished (Experiments 6-7). In Experiment 3 participants were asked to generate exculpatory evidence. Participants in all experiments were people without legal education (i.e., laypeople), but in Experiments 1-3 lawyers (i.e., advanced law students and graduated lawyers) were also tested. Whereas lawyers' defeasible reasoning adhered to the rules of penal code, the results showed that laypeople's defeasible reasoning depended on their own sense of justice. When asked whether an offender should be punished, laypeople ignored potential exculpatory evidence when the offence was highly morally outraging. In these cases, laypeople even had difficulties in retrieving exculpatory evidence from memory. Only when laypeople were asked whether an offender *will* be punished, were they more willing to also consider exceptions for highly morally outraging offences. Moreover, depending on people's attitudes about offences and offenders, sometimes legally punishable actions were not punished. Two additional experiments (Experiments 9-10) suggested that people are also prone to ignore exceptions for emotionally-charged events in everyday scenarios. The findings are relevant for cognitive psychology because they show the importance of considering domain knowledge and the reasoners personal attitudes and preferences when predicting inferences. Moreover, the results also have implications for law, social psychology, and society: they show how cognitive paradigms can be applied to test socially relevant constructs from legal theory and social psychology.

Table of Contents

Introduction **1**

Chapter 1: Theoretical Background **5**

1.1. Conditional Reasoning 5

1.1.1. Defeasible Reasoning 11

1.1.1.1. The Content of Conditionals 12

1.1.1.2. The Context of Conditionals 18

1.1.2. Theories on Conditional Reasoning 20

1.1.2.1. Mental Rules Theories 21

1.1.2.2. Mental Models Theory 24

1.1.2.3. Probabilistic Theories 28

1.1.2.4. Dual Process Theories 32

1.1.2.5. Theories on the Utilities of Conditionals 35

1.2. Legal Reasoning 38

1.2.1. Legal Reasoning and Conditionals in Legal Theory 38

1.2.2. Legal Rules and Justice in Social Psychology 43

1.2.3. Empirical Findings on Legal Reasoning 45

Chapter 2: Combining Conditional and Legal Reasoning - Aims and Hypotheses **48**

Chapter 3: Legal Rules, Moral Outrage and Domain Knowledge **55**

3.1. Experiment 1: The Legal Conditional Reasoning Paradigm 55

3.1.1. Methods 56

3.1.1.1. Participants 56

3.1.1.2. Materials and Design 57

3.1.1.3. Procedure 59
3.1.2. Results 59
3.1.2.1. Perceived Severity Ratings (Manipulation Check) 60
3.1.2.2. Defeated Conclusions (no-punishment) 61
3.1.2.3. Decision Times and Certainty Ratings 61
3.1.3. Discussion 63
3.2. Experiment 2: Legal Conditional Reasoning and Moral Outrage 64
3.2.1. Methods 65
3.2.1.1. Participants 65
3.2.1.2. Materials and Design 65
3.2.1.3. Procedure 68
3.2.2. Results 68
3.2.2.1. Perceived Severity Ratings (Manipulation Check) 68
3.2.2.2. Certainty Ratings 69
3.2.2.3. Defeated Conclusions (no-punishment) 70
3.2.2.4. Decision Times 72
3.2.3. Discussion 73
3.3. Experiment 3: Generating Exculpatory Circumstances 75
3.3.1. Methods 77
3.3.1.1. Participants 77
3.3.1.2. Material and Design 77
3.3.1.3. Procedure 78
3.3.2. Results 78
3.3.3. Discussion 79
3.4. Summary and Implications of Chapter 3 80
Chapter 4: Legal Conditional Reasoning, Religiosity, and Culture 82

4.1. Experiment 4: Religiosity 82
4.1.1. Methods 84
4.1.1.1. Participants 84
4.1.1.2. Materials, Procedure, and Design 85
4.1.2. Results 85
4.1.2.1. Perceived Severity 85
4.1.2.2. Defeated Conclusions (no-punishment) 87
4.1.2.3. Decision Times 89
4.1.2.4. Certainty 91
4.1.3. Discussion 91
4.2. Experiment 5: Culture 93
4.2.1. Methods 96
4.2.1.1. Participants 96
4.2.1.2. Material, Procedure, and Design 96
4.2.2. Results 99
4.2.2.1. No Circumstance Information 99
4.2.2.2. Vigilante and Exculpatory Circumstances 101
4.2.3. Discussion 104
4.3. Summary and Implications of Chapter 4 107

Chapter 5: Linguistic Factors in Legal Conditional Reasoning 109

5.1. Experiment 6: The Modal Auxiliary 110
5.1.1. Methods 112
5.1.1.1. Participants 112
5.1.1.2. Material and Design 112
5.1.1.3. Procedure 116
5.1.2. Results 116

5.1.2.1. Inference Task 116
5.1.2.2. Generation Task 118
5.1.2.3. Moral Outrage 119
5.1.3. Discussion 119
5.2. Experiment 7: The Modal Auxiliary and Closeness with the Offender 122
5.2.1. Methods 123
5.2.1.1. Participants 123
5.2.1.2. Material, Design, and Procedure 123
5.2.2. Results 123
5.2.2.1. Inference Task 123
5.2.2.2. Generation Task 126
5.2.2.3. Moral Outrage 126
5.2.3. Discussion 126
5.3. Experiment 8: Quantifier 128
5.3.1. Methods 129
5.3.1.1. Participants 129
5.3.1.2. Materials and Design 129
5.3.2. Results 132
5.3.3. Discussion 133
5.4. Summary and Implications of Chapter 5 135
Chapter 6: The Defeasibility of Negative Emotions in Everyday Scenarios 138
6.1. Experiment 9: Defeating Strong Negative Emotions 139
6.1.1. Methods 140
6.1.1.1. Participants 140
6.1.1.2. Material and Design 140
6.1.1.3. Procedure 142

6.1.2. Results 142
6.1.2.1. Control Problems with no Additional Information 142
6.1.2.2. Problems with Additional Positive or Negative Information 143
6.1.3. Discussion 144
6.2. Experiment 10: Defeating Mild Negative Emotions 144
6.2.1. Methods 145
6.2.1.1. Participants 145
6.2.1.2. Material, Design, and Procedure 145
6.2.2. Results 146
6.2.2.1. Control Problems with no Additional Information 146
6.2.2.2. Problems with Additional Positive or Negative Information 146
6.2.3. Discussion 148
6.3. Summary and Implications of Chapter 6 149
Chapter 7: General Discussion 151
7.1. Consequences for Cognitive Psychology 153
7.1.1. Defeasibility, Domain Knowledge, and Decision Making 153
7.1.2. Emotions in Defeasible Reasoning 157
7.1.3. Mental rules, Mental Models, Probabilities, or Utilities? 161
7.1.4. Legal Reasoning and Deontic Reasoning 167
7.1.5. What does it mean to be rational? 171
7.2. Consequences for Law 174
7.3. Consequences for Social Psychology 176
7.4. Consequences for Society 179
7.5. Limitations and Further Perspectives 181
7.5.1. Measuring Moral Outrage 181

7.5.2. Improving the Experimental Paradigm 184

7.5.3. Different Legal Conditionals 185

7.5.4. Expertise and Longitudinal Studies 188

7.5.5. External Validity 189

7.6. Conclusion 189

Chapter 8: References 191

Introduction

Imagine you are a judge presiding over a dispute about bodily injury. The case is very clear: a man *A*, 25 years old, beat his neighbor *B*, 30 years old, during a dispute that originated because *B* destroyed *A*'s flower bed. You know that the penal code includes a rule stating that *if a person beats another human, then the person should be punished for bodily injury*. How will you decide?

This is a simple case. The facts coincide perfectly with the legal requirements of the rule, so we can apply this rule and conclude that *A* should be punished for bodily injury. Such reasoning from if-then rules is known as conditional reasoning, and conditionals describing legal rules can be called *legal conditionals*.

Conditional reasoning is not unique to the legal domain. We actually reason from conditionals constantly in our day to day lives, often without even noticing it. For instance, when we see that it is sunny, we might consider using sunscreen because we know that *if the sun shines, then one can get sunburn*. Or, when we try arriving on time to work, it might be because our boss once told us that *if we come late, then he will fire us*. Conditional reasoning is so central to our daily lives that – not surprisingly – many theories have been developed to understand how human beings reason with conditionals (e.g., Braine & O'Brien, 1991; Johnson-Laird & Byrne, 1991; Oaksford & Chater, 2007). Many of these theories measure human reasoning performance against the assumptions of classical logic. Classical logic prescribes that whenever we have a conditional rule of the form "if *p* then *q*", and a fact stating *p*, *q*, ¬*p* or ¬*q*, we can infer logically what follows. All three examples presented in this chapter followed this structure: we had a conditional "if *p*, then *q*" and the fact *p*. As in classical logic, we assumed that *p* is sufficient for *q* to happen and thus concluded that if *p* is the case, *q* follows. In classical logic such a conclusion can only be true or false, nothing in between. Further, no additional information can make such a conclusion false, because as long as the premises (i.e., the rule and the fact) are true, the conclusion is necessarily true. This is the property of monotonicity of classical logic. But do we reason monotonically in our daily lives?

Imagine again you are a judge. Once more you are confronted with a case of bodily injury. At first glance the case is very similar to the one you solved before. A man *C*, 25 years old, beat this neighbor *D*, 30 years old, in an dispute that originated after *D* destroyed *C*'s flower bed. As in the previous case, you might consider applying the rule against bodily

injury and conclude that *C* should be punished. However, unexpectedly *C*'s lawyer comes up with *exculpatory evidence* showing that C was actually defending himself. In the dispute about the flower bed, *D* had actually attacked *C* first so that *C* had no alternative but to attack back. How would you decide?

In law, exculpatory evidence can make judges change their conclusions. Judges may, in light of exculpatory evidence, refrain from punishing offenders, even though a punishable act was committed. Also in our daily lives we draw conclusions which we then withdraw in light of new information. For example, even though the sun is shining, we might refrain from concluding that we should use sunscreen if we know that we will be at home the whole day. Or, even though our boss told us to arrive on time for work; we know that our boss will not fire us if we have a written permission to arrive late once. This flexibility of human reasoning cannot be captured by monotonic classical logic. This is why in recent years psychologists' attention has turned to the investigation of *defeasible reasoning* – a kind of reasoning where conclusions can be defeated by subsequent information (e.g. Oaksford & Chater, 1995a; 2013; Pfeifer & Douven, 2014; Pollock, 1987; Stenning & van Lambalgen, 2005).

Defeasible reasoning has a long tradition at the interface between philosophy, logic and artificial intelligence (e.g., Brewka, 1991; Delgrande, 1987; McCarthy & Hayes, 1969; Pelletier & Elio, 1997). In those fields researchers have long proposed non-monotonic logics to capture human defeasibility (e.g., *System P* by Kraus, Lehmann & Magidor, 1990; or *close-world reasoning* by Stenning & van Lambalgen, 2008). In psychology, the term "defeasible reasoning" is not encountered often. Still, there are many psychological studies showing that people defeat initially drawn conclusions in light of new information, dating back to the late eighties (e.g., Byrne, 1989; Cummins, Lubart, Alksnis, & Rist, 1991; De Neys, Schaeken, & d'Ydewalle, 2003a; Oaksford & Chater, 1991; 1995a). Psychologists' main focus of research in this area has been to find out which factors affect the withdrawal of conclusions. For this, the majority of studies employed conditionals describing everyday situations, such as the sunburn example above. But what about conditionals describing legal cases, such as in the bodily injury example? How human beings defeat conclusions from conditionals describing *legal rules* has hardly been investigated.

The aim of the current thesis is to combine the fields of legal reasoning and conditional reasoning to investigate defeasibility in law experimentally. Understanding how human beings reason with legal conditionals is important from a psychological, but also from a legal point of view. Most societies have a penal code and other legal stipulates that prescribe which actions deserve punishment and which do not. However, the particulars of this

knowledge are mostly in the hands of legally educated people. Laypeople usually do not know all regulations of the penal code. We therefore cannot know how people without legal education reason with legal rules. It may be that their defeasible reasoning is similar to that of lawyers and judges. However, it could also be that laypeople have problems with accepting exculpatory evidence. Often people complain about courts' and judges' decisions, for example when a suspect is believed to be guilty but is released because of lack of evidence – as in the lawsuit against Jörg Kachelmann in Germany (see Jüttner, 2011; "Geteiltes Echo auf," 2011). Complaints also arise when offenders commit morally reprehensible acts but exculpatory evidence – such as psychological disorders – exonerates them or reduces their sentences (e.g., Bloechl, Vitacco, Neumann, & Erickson, 2007; Hans, 1986; cf. Tyler & Boeckmann, 1997). These reactions from laypeople to offenders suggest that laypeople's own sense of justice may influence their legal reasoning. But how? Does laypeople's reluctance to forgive offenders also lead to less consideration of exculpatory evidence? This thesis aims to shed light on this open question by investigating whether and how people's own sense of justice affects their defeasible reasoning with legal rules, and how this defeasibility can be modulated.

This thesis starts with an introduction on conditional reasoning. It is explained how conditional reasoning is usually investigated in psychology and under which circumstances people's reasoning is defeasible. The cognitive mechanisms behind defeasible reasoning are further described by presenting the most important theories on human reasoning. Afterwards, the literature on legal reasoning is presented; first from the perspective of legal theory. It is explained how penal codes implement defeasibility and how conditionals can be used to describe legal reasoning. Then, legal reasoning is explained from the perspective of social psychology by presenting studies on people's sense of justice. Next, empirical findings on how humans reason with conditionals describing punishable or undesired acts are presented. After having provided the theoretical overview on conditional and legal reasoning, the two domains are brought together to draw the main hypotheses of this thesis. It is proposed that contrary to lawyers, people without legal education use their own sense of justice to reason with legal conditionals. Laypeople's withdrawal of conclusions drawn from legal rules will therefore depend on how morally wrong they perceive an offence to be. These hypotheses are tested in three experimental blocks. The first block consists of three experiments and investigates differences between laypeople's and lawyers' legal reasoning. For this, legal rules describing offences were phrased as conditionals and presented together with exculpatory

evidence. It is investigated whether the moral wrongfulness of offences predicts laypeople's acceptance of exculpatory evidence and the withdrawal of conclusions. The second block investigates how behavioral rules affect laypeople's sense of justice, and in turn, their defeasible reasoning. The effect of behavioral rules is investigated by testing the role of religiosity and cultural background on legal reasoning. The third block consists of three experiments, which explore how linguistic factors, such as the way in which legal rules are phrased, influence legal reasoning and the withdrawal of conclusions. Finally, two last experiments are presented. In these two experiments the domain of legal reasoning is abandoned to test whether the findings from the previous experiments on the defeasibility of legal conditionals have equivalents in everyday scenarios. The thesis ends with a General Discussion, where the main findings are recapitulated and their theoretical and practical implications for cognitive psychology, social psychology, law, and society are discussed.

Chapter 1: Theoretical Background

1.1. Conditional Reasoning

The experimental investigation of defeasible reasoning in law requires the combination of two fields: cognitive psychology and legal theory. Therefore, I start with reviewing the psychological literature on human reasoning in general. Introducing current reasoning theories and the concepts of deduction and defeasibility enables the reader to understand how defeasibility can be investigated experimentally and applied to legal theory.

The study of reasoning in psychology originates from classical logic, which is taken to provide standards for how deductive inferences should be drawn (Evans, 2002). In deduction people reason from the general to the particular; such as judges do if they have to decide whether the general rule of bodily injury applies to a specific case. Deduction differs therefore from inductive reasoning, where people reason from the particular to the general (e.g., inferring that all thieves wear masks after having observed some thieves wearing masks). Besides syllogistic (i.e., reasoning with quantifiers) and relational reasoning (i.e., reasoning about relations), *conditional reasoning* has been one of the most investigated forms of deductive reasoning in psychology. In conditional reasoning, reasoners draw inferences from a conditional rule of the form "if *p*, then *q*". The if-part of a conditional is called the antecedent (*p*) and the then-part the consequent (*q*), both are atomic propositions. Each atomic proposition can have the truth value true or false. Depending on these truth values, the sentence "if *p*, then *q*" is either true or false. Propositional classical logic captures this dependence by the following truth table for the connective "if" (also called *material implication*):

Table 1

Truth table for material implication.

p	*q*	If *p* then *q*
true	true	true
true	false	false
false	true	true
false	false	true

In the first row of Table 1, *p* and *q* are true, and therefore the statement "if *p*, then *q*" is also true. In the second row, *p* is true, but *q* is false, so that the statement "if *p*, then *q*" is false. However, if *p* is false, irrespective of the truth value of *q*, the compound "if *p*, then *q*" is true. It is worth noting that in propositional logic, the connective *if* has a different meaning than in everyday language. "If *p*, then *q*" does not mean that *q* is only true if *p* is true. Instead, material implication states the truth value of "if *p*, then *q*" depending on whether the atomic propositions are either true or false.

Psychologists have employed different paradigms to investigate conditional reasoning; the most widely used being the *truth table task*, the *selection task*, and the *conditional inference task*. Because of its relevance for legal reasoning (see Section 1.2.1) the experiments in this thesis dealt with the conditional inference task. In the conditional inference task participants usually are confronted with three statements, for instance:

If the sun is shining (*p*), then Jack puts sunscreen on his face (*q*).
The sun is shining (*p*).

Does Jack put sunscreen on his face (*q*)?

The two statements above the line are called *premises*, the first of them being a conditional, and the second a *fact* (also called *categorical statement*) stating either *p*, *q*, ¬*p* (not *p*), or ¬*q* (not *q*). The participant's task is usually to decide whether the statement below the line, the *conclusion*, follows necessarily from the previous premises. The conclusion can either be phrased as a question (as in the above example) or as a concrete statement (i.e., Jack puts sunscreen on his face). According to classical logic a conclusion is always true if the premises are true and the reasoning is correct. Additional information can never alter a valid conclusion. This property is referred to as *monotonicity* (cf. Stenning & van Lambalgen, 2008) and results in the four possible inferences depicted in Table 2.

Table 2

The four inferences in classic propositional logic. Table based on Knauff (2006).

Inference	Validity	Example
Modus Ponens (MP)		
If *p*, then *q*	valid	If the sun is shining, then Jack puts sunscreen on his face
p		The sun is shining
q		Jack puts sunscreen on his face
Modus Tollens (MT)		
If *p*, then *q*	valid	If the sun is shining, then Jack puts sunscreen on his face
¬*q*		Jack does not put sunscreen on his face
¬*p*		The sun is not shining
Affirmation of the Consequent (AC)		
If *p*, then *q*	invalid	If the sun is shining, then Jack puts sunscreen on his face
q		Jack puts sunscreen on his face
p		The sun is shining
Denial of the Antecedent (DA)		
If *p*, then *q*	invalid	If the sun is shining, then Jack puts sunscreen on his face
¬*p*		The sun is not shining
¬*q*		Jack does not put sunscreen on his face

According to material implication, Modus Ponens (MP) and Modus Tollens (MT) are valid inferences, but Affirmation of the Consequent (AC) and Denial of the Antecedent (DA) are not (Table 2). In classical logic, the validity of inferences depends only on the structure of the inference. The actual content of the conditional is irrelevant. Nonetheless, I will illustrate why some inferences are valid and others not with the conditional from the initial example.

Given the conditional "If the sun is shining, then Jack puts sunscreen on his face" we can conclude that every time the sun is shining Jack will put sunscreen on his face (MP) and

that if he is not putting sunscreen on his face, it is because the sun is not shining (MT). The validity of MP and MT inferences can be actually read off from the truth table of material implication, assuming that the conditional "If *p*, then *q*" is true. In MP inferences *p* is given as true. Accordingly, in the first row of Table 1 we see that when *p* is true, *q* is also true. Similarly, in MT inferences ¬*q* is given. Accordingly, in the second row of Table 1 we can see that if *q* is false, *p* is also false. However, it is wrong to conclude that if Jack puts sunscreen on his face, it is because the sun is shining (AC), and similarly that if the sun is not shining, Jack does not put sunscreen on his face (DA). AC and DA are fallacies according to material implication because there are alternative reasons why Jack might put sunscreen on his face although the sun is not shining (e.g., testing how tolerant his skin is for a new sunscreen, or putting sunscreen as a preventative measure). Also the truth table shows why AC and DA are fallacies. In AC inferences *q* is given, and reasoners are asked whether *p* follows. However, Table 1 shows that if *q* is true, *p* can be true (first row) or false (third row); there is thus no certain conclusion that can be drawn. The same applies for DA. In DA inferences *p* is negated (i.e., ¬*p* is true). But Table 1 shows that if *p* is false, then *q* can be either true or false (see the last two rows). So again, we can infer nothing from ¬*p*. Because of this relationship between *p* and *q*, some researchers argue that *p* is sufficient but not necessary for the consequent *q*, and the consequent *q* necessary for its antecedent *p* (e.g., Hilton, Jaspars, & Clarke, 1990; Thompson, 1994; 1995).

In the past 60 years many studies have been conducted to investigate human conditional reasoning (Evans, 2002), some studies even date back to the beginnings of the 20th century (see Störring, 1908). In most of them, participants were asked to assume the premises as true and to infer what necessarily follows from them. Sometimes abstract material (e.g., Evans, Clibbens, & Rood, 1995; Marcus & Rips, 1979), but also content rich conditionals such as the conditional about Jack and the sunscreen (e.g., Byrne, 1989; Markovits, 1986; Stevenson & Over, 1995) were used. Based on which conclusions reasoners drew, researchers aimed to discover something about human rationality. But, contrary to their expectations, researchers found that the inferences people draw – especially the ones with content rich material – do not always comply with classical logic. As can be seen in Table 3, people have difficulties in recognizing the validity of some inferences. For instance, AC and DA inferences are often erroneously classified as valid, in some studies in over 50% of the cases. Also, MT inferences are difficult. While in some experiments reasoners manage to draw correct MT inferences, in others they reject MT inferences in almost half of the cases. Though less frequently than with MT inferences, in some experiments people also have difficulties

with MP, i.e., concluding that *q* follows from *p* (see Table 3, up to 14% errors in some studies). Researchers have tried to find explanations to these deviations from classical logic.

Some researchers see these deviations from classical logic as errors in a strict sense, i.e., cases of wrong reasoning. Wrong reasoning can happen because people understood the task wrongly, because of limitations of working memory, or because reasoners are biased by the content of the conditional. For instance, errors in accepting AC and DA inferences are often explained by assuming that people interpret conditionals not as material implication, but as biconditionals (Evans, Newstead, & Byrne, 1993). In a biconditional interpretation the conditional is understood as "If and only if *p*, then *q*", implying that *q* only happens if *p* is the case. Therefore, if reasoners hear that *p* is false they assume wrongly that also *q* is false; or when they hear that *q* is false they assume that also *p* is false. Such erroneous interpretations are often the result of conversational processes (Grice, 1975). For instance, humans usually assume that speakers only communicate what the hearer needs to know. So, when confronted with the conditional "If the sun is shining, then Jack puts sunscreen on his face", reasoners assume that all other instances where Jack might use sunscreen are not relevant for the speaker and are consequently not mentioned. Another explanation for errors in conditional reasoning are problems in processing negations. Negations are often thought to be the primary reason for problems with MT inferences (De Neys et al., 2003a). However, errors in MT inferences can be also explained by the amount of mental steps necessary to arrive at the conclusion (see Section 1.1.2.2) or the amount of mental rules that have to be applied to infer that ¬*p* follows from ¬*q* (see Section 1.1.2.1).

Other researchers, however, argue that deviations from classical logic are not actually errors. Instead, they argue that people's conclusions do not always follow the rules of classical logic because classical logic is likely not the correct norm for describing human reasoning (Bonnefon & Vautier, 2010; Evans, 2002; 2012; Oaksford & Chater, 2009). In classical logic, premises can be only true or false, nothing in between. But in our daily lives things are rarely that certain. The information we get is only true to a certain degree. If somebody tells us that *if the sun shines, Jack uses sunscreen*, we know that this is not an absolute truth. Instead, we know this rule only describes something which is usually the case, with some degree of uncertainty. Consequently, the conclusions we draw from such uncertain premises are also only more or less probable, and not definitely true or false as in classical logic (e.g., Evans, 2012; Oaksford & Chater, 2013; Pfeifer, 2013). Particularly the monotonicity of classical logic does not apply to many situations in our daily lives. Many conclusions we draw can be defeated by subsequent information. For instance, although we

might have concluded initially that Jack will put sunscreen on his face, we might withdraw this conclusion when we hear that Jack will be at home the whole day. Everyday reasoning is therefore *non-monotonic* and *defeasible*: additional information can make people reject previous valid conclusions (e.g., Oaksford & Chater, 1995a; 2013; Pfeifer & Douven, 2014; Politzer, 2007; Pollock, 1987; Stenning & van Lambalgen, 2005).

Table 3

Frequency (%) of acceptance of MP, MT, AC, and DA inferences in the literature up to 1990. Table adapted and extended from Evans, Newstead, and Byrne (1993).

Study	n	MP	MT	AC	DA
Taplin (1971)	56	92	63	57	52
Evans (1977)	16	100	75	75	69
Marcus and Rips (1979)					
Experiment 1 (2 choice)	36	99	62	29	31
Experiment 2	24	98	52	33	21
Rumain, Connell, and Braine (1983)					
Experiment 1 [a]	24	98	81	27	28
Experiment 2 [a]	24	100	63	28	17
Markovits (1988)					
Trial 1 (average between both conditions)	76	100	59	42	52
Byrne (1989)					
Experiment 1 (simple arguments)	24	96	92	71	46
Markovits and Vachon (1990)					
Experiment 2 (concrete material)[ab]	150	86	70	15	8
Experiment 2 (abstract material)[ab]	150	90	70	28	34

[a] Children were also tested in this experiment, but only data from the adult control group is reported here

[b] Averaged across the two presentation orders

1.1.1. Defeasible Reasoning

In defeasible reasoning previously drawn conclusions can be withdrawn (i.e., *defeated*) in light of new information. Defeasible reasoning is very important in argumentation (e.g., Pollock, 1987), but also in law, e.g., where exculpatory evidence can bring judges to conclude that an offence should not be punished (e.g., Prakken & Sartor, 2004). In fact, focusing on the defeasibility of human reasoning has a long tradition in philosophy and artificial intelligence (e.g., Brewka, 1991; Delgrande, 1987; McCarthy & Hayes, 1969; Pelletier & Elio, 1997). In psychology, however, researchers have only recently started to investigate defeasible reasoning experimentally – even though the term "defeasible reasoning" is not used often.

One of the first psychological papers showing that people defeat previously drawn conclusions was written by Byrne (1989). In a seminal paper Byrne (1989) confronted participants with two different kinds of inference tasks. Half of her problems had the structure of typical conditional inference tasks, describing a conditional "If *p*, then *q*":

If Ann has an essay to write (*p*), then she will study late in the library (*q*).
Ann has an essay to write (*p*).

Ann will study late in the library (*q*).

She found that in around 96% of the cases participants correctly made the MP inference. However, the other half of the problems contained an additional premise "If *r*, then *q*", describing an additional requirement *r* for *q*:

If Ann has an essay to write (*p*), then she will study late in the library (*q*).
If the library stays open (*r*), then Ann will study late in the library (*q*).
Ann has an essay to write (*p*).

Ann will study late in the library (*q*).

Now, participants made the MP inference in only 38% of the cases. Byrne initially interpreted her findings as evidence showing that people do not apply mental rules during reasoning (see Section 1.1.2.1). However, Byrne's findings are understood nowadays as evidence for people's ability to defeat initially drawn conclusions by additional information (Da Silva

Neves, Bonnefon, & Raufaste, 2002; Stenning & van Lambalgen, 2005).[1] Since Byrne's study, many other experiments have been carried out to measure human defeasibility. Some researchers even argue that the central paradigm of cognitive psychology has changed (Evans, 2012). Instead of testing people's deductive reasoning abilities against the rational norms of classical logic, nowadays – in the so called *new psychology of reasoning* – researchers are more concerned with people's everyday reasoning: how background knowledge, preferences, and experiences affect the conclusions people draw – and withdraw. In the following passages I describe how the content (Section 1.1.1.1) and the context of conditionals (Section 1.1.1.2) affect the degree to which people engage in defeasible reasoning.

1.1.1.1. The Content of Conditionals

One of the main reasons why reasoning is defeasible stems from people's background knowledge about the content of conditionals (e.g. De Neys, Schaeken, & d'Ydewalle, 2003a, 2003b; Dieussaert, De Neys, & Schaeken, 2005; Evans & Over, 2004; Johnson-Laird & Byrne, 2002; Oaksford & Chater, 1995a; 2003a; 2009). Consider again the sunscreen example:

If the sun is shining (*p*), then Jack puts sunscreen on his face (*q*).
The sun is shining (*p*).

Does Jack put sunscreen on his face (*q*)?

At first sight, the MP conclusion that Jack puts sunscreen on his face given the sun is shining seems plausible. However, people may refuse to draw MP inferences if they think there are reasons why people would not put sunscreen on their faces although the sun is shining, for example if somebody had already a deep tan, if somebody did not care about skin cancer, if someone wanted to get tanned faster, if someone knew he would be inside all day, if they had run out of sunscreen, if someone developed an allergy against sunscreen components, etc. …

[1] One could argue that Byrnes' (1989) findings can be still explained within the monotonicity framework if all possible instances that break the link between *p* and *q* are inserted as part of the antecedent (If *p* and not *r*, *s*, *t*…, then *q*). However, this explanation is psychologically and computationally implausible because it is impossible to enumerate exhaustively all possible defeaters that may exist (cf. Dennett, 1987; Fodor, 1987; see Chapter 1.2.1 for how this problem applies to legal reasoning).

When people consider conditions that prevent *q* even though *p* is given, they will reject or have problems with accepting MP conclusions (e.g., Byrne, 1989; Cummins et al., 1991; De Neys et al., 2003a, 2003b; Dieussaert et al., 2005; Johnson-Laird & Byrne, 2002; Markovits & Potvin, 2001; Oaksford & Chater, 2001). These *disabling conditions*, or more general, *defeaters*, do not only influence MP, but also MT inferences. For example:

If the sun is shining (*p*), then Jack puts sunscreen on his face (*q*).
Jack does not put sunscreen on his face (¬*q*).

Is the sun shining (*p*)?

As in the MP example, defeaters such as the fact that the sunscreen has run out, or that Jack is already tanned, or an unexpected allergy etc. can lead people to reject the valid MT conclusion and instead conclude that the sun is actually shining. And note that contrary to Byrne's (1989) experiment, it is not necessary to explicitly present possible defeaters as part of the conditional inference task. Even if defeaters are not explicitly presented in inference tasks, people can still consider them because they have background knowledge about the content of the conditional (cf. Cummins, 1995; Cummins et al., 1991; De Neys, Schaeken, & d'Ydewalle, 2002; De Neys et al., 2003a; Thompson, 1994).

Defeaters are also relevant for the fallacies AC and DA. In this case, the literature refers to them as *alternatives* (e.g., Cummins et al., 1991; Markovits & Quinn, 2002). Consider the following examples:

If the sun is shining (*p*), then Jack puts sunscreen on his face (*q*).
Jack puts sunscreen on his face (¬*q*).

Is the sun shining (*p*)?

If the sun is shining (*p*), then Jack puts sunscreen on his face (*q*).
The sun is not shining (¬*p*).

Does Jack put sunscreen on his face (*p*)?

Here, background knowledge can also help people to generate conditions that would make Jack put sunscreen on his face even though the sun is not shining, such as to test how a sunscreen feels on the skin, or if it is the only lotion available to moisturize the skin, or applying it preventively, etc. However, note that the consideration of alternatives does not

lead people to reject logically valid conclusions. On the contrary, considering such alternatives makes people aware that AC and DA are fallacies.

The likelihood that defeaters are considered depends on how many disabling and alternative conditions actually exist for a given conditional. For instance, the conditional about Jack and the sunscreen has many disabling conditions, but other conditionals like for example "If Jack cuts his finger, then he will bleed." do not. One of the first researchers studying the effect of the amount of defeaters on reasoning was Cummins (1995; Cummins et al., 1991). In a preliminary study, she created conditionals whose amount of disabling conditions and alternatives was varied orthogonally. She therefore created conditionals for which people can generate many disabling and many alternative conditions, conditionals for which people can generate many disabling conditions but few alternatives, conditionals for which people can generate few disabling conditions but many alternatives, and conditionals for which participants can generate few disabling conditions and few alternatives. She embedded these conditionals in MP, MT, AC and DA inferences and asked participants to evaluate the conclusion. Although defeaters were actually never presented as part of the inference task, Cummins found that participants considered them anyway. MP and MT inferences from conditionals with many disabling conditions were accepted less frequently than inferences from conditionals with few disabling conditions. Analogically, AC and DA inferences from conditionals with many alternatives were accepted less frequently than inferences from conditionals with few alternatives. Cummins' findings have been replicated several times (e.g., Byrne, Espino, & Santamaria, 1999; Cummins, 1995; De Neys et al., 2002; 2003a; 2003b; Dieusseart et al., 2005; Thompson, 1994, 1995).

Many psychologists assume that defeaters affect inferences because people usually activate their semantic knowledge about the content of conditionals during reasoning tasks (De Neys et al., 2003a; 2003b; Janveau-Brennman & Markovits, 1999; Markovits, 2000; Vadeboncoeur & Markovits, 1999; see also Markovits, Fleury, Quinn, & Venet, 1998). In this way, the more disabling conditions or alternatives people have stored in their memory for a given conditional, the more probable it is that at least one of them will be retrieved and therefore considered during reasoning. Initially this was thought to be an "all-or-nothing phenomenon" (De Neys et al., 2003a, p. 582): as soon as a person retrieves one disabling condition or one alternative he or she rejects the corresponding conclusion; if they do not find one, they accept the conclusion (see Markovits, 2000; Vadeboncoeur & Markovits, 1999). However, in a series of experiments De Neys and colleagues (2003a) showed that the consideration of defeaters is gradual, and that every additional disabling condition or

alternative has an impact on the degree to which people accept a conclusion. In their experiments, De Neys and colleagues embedded conditionals with many disabling conditions and many alternatives in conditional inference tasks, and, additionally to the usual premises, presented either one, two, three, or four defeaters. They found that people's acceptance of the conclusion was related in a linear fashion to the amount of disabling conditions or alternatives additionally presented.

A potential mechanism by which the amount of defeaters affects conclusions is via the influence they exert on the perceived sufficiency and necessity relation between *p* and *q*. In classical logic *p* is sufficient, but not necessary for *q* to happen (e.g., Hilton et al., 1990; Thompson, 1994; 1995). However, the existence of disabling conditions questions this sufficiency of *p*. If there are conditions which prevent *q* from happening even though *p* is the case (such as when the sunscreen bottle is empty), then *p* is not sufficient for *q* anymore. In other words, the more disabling conditions exist, the less sufficient one perceives *p* to be for *q*. The same applies for alternative causes. The more alternatives one can generate, the less *p* is perceived as necessary for *q*, and the fewer AC or DA fallacies reasoners will make (see Thompson, 1994; 1995).

Additionally to their amount, also the associative strength or relative salience of defeaters influences inferences (De Neys, et al., 2002; De Neys et al., 2003b; Chan & Chua, 1994; see also Markovits & Potvin, 2001). If we go back to our initial example of Jack and the sunscreen, one can generate different kinds of disabling conditions. For instance, one disabling condition could be that the bottle of sunscreen is empty, and another one can be that Jack developed an allergy against some component of the sunscreen preventing him from using it. Now imagine you are confronted with this conditional in a reasoning task and have to solve a MP inference. Which of both disabling conditions will pop up more readily in to your mind? Although both are valid disabling conditions, the former may occur to people more easily than the latter. According to De Neys and colleagues (2003b) this happens because some disabling conditions are more strongly associated with one's semantic knowledge of how to prevent *q* from happening (see also Markovits et al., 1998; Quinn & Markovits, 1998). Analogously some alternative causes are more strongly associated to one's semantic knowledge of how to cause *q*. Strongly associated defeaters are considered more readily than those which are not. As a consequence, one strongly associated disabling condition can have a higher impact on withdrawing from MP or MT inferences than many weakly associated disabling conditions; and one strongly associated alternative cause can have a higher impact on rejecting AC and DA inferences than many weakly associated alternatives (see Quinn &

Markovits, 1998). A similar account was proposed by Chan and Chua (1994), but only for disabling conditions. They propose that disabling conditions differ in their relative salience, i.e., how important they are perceived to be based on the prior knowledge of the reasoner. Depending on this salience, potential disabling conditions are accepted more or less strongly. Also Manktelow and Fairley (2000) proposed that not all defeaters are the same. They say that disabling conditions and alternatives differ in their relevance in respect to *superordinate principles*, such as in our example "ways in preventing sunburns" (cf. Manktelow & Fairley, 2000).

One direct consequence of considering defeaters is that they lower the believability of the conditional. For instance, the conditional "If Anna eats lots of candies, then she will get cavities" is probably not believed very much, because it lays at hand that by brushing the teeth and regular visits to the dentist one can still eat candies but not get cavities. Experiments on the believability of conditionals on defeasible reasoning have mainly focused on the role of believability on the rejection of MP and MT inferences. George (1997), for instance, conducted experiments where conditional statements where either presented traditionally as "if *p*, then *q*" (e.g., If Pierre is in the kitchen, then Marie is in the garden), or with an additional "very probably" or "not very probable" in front of the conditional (e.g., It is very probable that if Pierre is in the kitchen then Marie is in the garden). He found that the uncertainty of the conditional statement influenced participant's belief in the conclusion. However, in another experiment George (1995) found that the believability of conditionals only affects some people. In his experiments he found that while for half of the participants the believability of the conditional correlated with the believability of the conclusion, for other reasoners this was not the case. Instead, the believability of the conditional did not affect conclusions. Also Stevenson and Over (1995) found similar effects. They created conditionals such as "If John goes fishing, he will have a fish supper" and presented these together with additional information such as "If John catches a fish, he will have fish supper". The interesting manipulation was, however, that they varied the believability of defeaters by telling, for example, that John always is lucky or never is lucky when he goes fishing. Stevenson and Over found that when defeaters were presented as uncertain, their defeasible power was reduced.

Studies on the amount and strength of defeaters have all the underlying assumption that people take into account each possible defeater during reasoning. The more strong defeaters people can think of, the more likely a conclusion is withdrawn. However, an alternative explanation for the consideration of defeaters is that not the amount of possible

defeaters matters, but the general frequency of *p* and ¬*q* cases (i.e., the *frequency of exceptions*) or the frequency of ¬*p* and *q* cases (i.e., the *frequency of alternatives*) – irrespective of how many different disabling or alternative conditions are behind these frequencies. Usually, the amount of defeaters and the frequency of exceptions or alternatives are correlated. For instance, in the sunscreen example there are many disabling conditions and also the frequency of *p*¬*q* instances is high (i.e., it happens often that the sun is shining and people do not use sunscreen). But what about this conditional: "If the traffic lights are red, then the car will stop"? If someone is asked to generate disabling conditions for this conditional, it is possible to say that the driver had an emergency, that the car is an ambulance, that the driver did not care, or did not see the street light, etc. But, although there are many disabling conditions, the overall amount of *p*¬*q* cases is still low because, in general, there are not so many instances where the traffic lights are red and the car does not stop. Along the same lines it is possible to generate many alternative causes for the conditional "If the brake is depressed, then the car slows down" (e.g., out of gas, collision, driving uphill, etc., see De Neys et al., 2002). But the perceived frequency of cases where a car slows down without the brake being pressed is probably low. An important study on the relationship between the amount of defeaters and frequencies was done by Geiger and Oberauer (2007). They conducted four experiments on the relationship between disabling conditions and frequency of exceptions (i.e., *p*¬*q* cases). In the first three experiments they created conditionals with fictional content (e.g., If an animal belongs to the family of grocks, then it has six legs) and provided reasoners with information about frequencies and possible disabling conditions. In the fourth experiment they used everyday conditionals (e.g., If you open the fridge, then the light inside goes on), which differed orthogonally in their amount of disabling conditions and their overall frequency of exceptions (measured within a preliminary study). They found that it was not the amount, but the frequency of exceptions that best predicted MP and MT inferences.

The relationship between frequency of exceptions and the associative strength of defeaters is not clear. It is thinkable that what makes a defeater strong or salient in memory is related to how often it actually occurs. One attempt in combining both factors can be found in Fernbach and Erb (2013). Fernbach and Erb say that all disabling conditions have some base rate and also some disabling strength. When both factors are considered together, one has a measure of the disabling probability of one disabling condition. Accordingly, when the disabling probabilities of all possible disabling conditions are combined, one has a measure for the overall probability that the antecedent will not cause the consequent. It is possible to

imagine that this overall disabling probability is related to the perceived frequency of exceptions (i.e., instances of *p* but ¬*q*).

1.1.1.2. The Context of Conditionals

The amount, strength and frequency of defeaters affect defeasible reasoning through the content of conditionals. Context factors, instead, affect defeasible reasoning through the way in which conditionals are presented or framed. For instance, in a series of experiments, Stevenson and Over (2001) embedded conditionals describing health issues (e.g., If Bill has typhoid he will make a good recovery) in conversational contexts and varied the level of expertise of the speakers (first-year student vs. professor of medicine). They found that when the conditional is uttered by an expert in the field, people believed much more in the MP conclusion than when it was uttered by a novice. Similarly, when additional information denying the minor premise was added to the task (i.e., "Bill has typhoid", "No, Bill has cholera"), this additional information was considered more when it was uttered by an expert compared to a novice (see also Dieussaert, Schaeken, & d'Ydewalle, 2002; Evans & Twyman-Musgrove, 1998; Wolf, Rieger, & Knauff, 2012). Although denying the minor premise of inference tasks is not exactly the same as presenting defeaters, we can relate Stevenson and Over's findings to the acceptance of defeaters. If the conditional "If the sun is shining, then Jack puts sunscreen on his face" is uttered by a person who knows Jack very well, then people will presumably think less about defeaters than when the conditional is uttered by someone who does not know Jack.

In addition to the trustworthiness of the source, defeasible reasoning is also affected by other context factors. For instance, experimental manipulations such as instructions, response modality, and the phrasing of conditionals can also affect the degree to which people consider defeaters. These context factors are described in more detail in the following two sections.

1.1.1.2.1. Instructions and response modality

Instructions can influence the degree to which people reason strictly deductively or defeasible. For instance, Vadeboncoeur and Markovits (1999) found that the acceptance of MP and MT is quite low when participants are simply instructed to suppose the premises to be true, without further clarifications. However, when the instructions highlight the logical nature of the task,

telling participants that the statements have to be considered as true even though this may not be the case in everyday life, endorsements of MP and MT inferences increase. Similar effects were found by Singmann and Klauer (2011): when instructed to base conclusions on the logical form of the problem assuming the rule to be true, participants made much more MP inferences than when instructed to simply judge the probability of the conclusion given the rule and the fact (see also Evans, Handley, Neilens, & Over, 2010).

Also the response modality has an effect on reasoning. In the last section I presented a study of Geiger and Oberauer (2007) who claimed that not the amount, but the overall frequency of exceptions predicts inferences. However, Markovits, Forgues, and Brunet (2010) showed that Geiger and Oberauer's claims are only true when the response modality is scaled. When the response modality is categorical (i.e., dichotomous) the mere presence of defeaters was enough to withdraw from valid inferences, without any linear relationship to the frequency of exceptions. In fact, Markovits et al. noticed that some inconsistent findings between studies can be solved by looking at the response modality used.

1.1.1.2.2. Phrasing of the conditional

Most of the literature on defeasible reasoning has investigated the effect of defeaters by varying the content of conditionals, for example by presenting conditionals with either many or few defeaters. Little attention, however, has been given to how the conditionals are phrased. For instance, researchers only seldom give attention to whether their conditionals describe general rules (e.g., "If the sun is shining, then human beings use sunscreen") or specific rules (e.g., "If the sun is shining, then Jack uses sunscreen"). Little attention is also given to the tense in the premises. For instance, the categorical statement of an inference task can be presented in the present (e.g., the sun is shining) or in the past (e.g., the sun was shining).

Of particular importance for defeasible reasoning is the modal which is used to phrase conditionals. Let's go back to our initial example of Jack and the sunscreen. The relationship between the sun shining and Jack putting sunscreen on his face can be described in different ways. It can be described as an indicative conditional, describing what is factually the case "If the sun is shining, then Jack *will* put sunscreen on his face". On the other hand, the conditional can be phrased such that it suggests uncertainty, such as "If the sun is shining, then Jack *might* put sunscreen on his face". Further, conditionals can also be phrased as

describing a deontic relationship such as "If the sun is shining, then Jack *should* put sunscreen on his face" (e.g., Beller, 2010; Perham & Oaksford, 2005). It becomes clear that defeaters are weighted differently depending on how the conditional is phrased. If the sun is shining and one has to infer whether Jack *will* put sunscreen on his face (i.e., a factual conditional), then calculating the amount of defeaters – or the frequencies of exceptions – makes perfect sense. However, when one has to infer whether Jack *should* put sunscreen on his face (i.e., a deontic conditional), then counting defeaters or estimating "*p* and not *q*" instances is not very fruitful. Instead, weighing the advantages and disadvantages of a conclusion given the premises is more useful (e.g., Given that the sun is shining, is it really necessary for Jack to put sunscreen on his face?). These assumptions are strengthened by a study of Over, Manktelow and Hadjichristidis (2004). They found that when conditionals are not factual, but deontic, the acceptance of conditional rules does not depend on the perceived probability of *q* given *p* (see Section 1.1.2.3), but instead on the preference of the different outcomes such a rule can have. The more a reasoner prefers the outcome '*p* and *q*' (in our case: the sun shining and putting sunscreen on his face) over '*p* and ¬*q*' (in our case: the sun shining and Jack not putting sunscreen on his face), the more a deontic rule will be accepted (see also Oaksford & Chater, 2007; 2009).

1.1.2. Theories on Conditional Reasoning

So far I have described defeasible reasoning and explained how the content and context of conditionals influence people's willingness to defeat conclusions. Before moving on to the importance of defeasibility in law, it is necessary to give an overview on how the existing reasoning theories describe defeasible reasoning. Such a theoretical overview allows for a better understanding of the before mentioned effects and therefore provides a theoretical ground to base this thesis' hypotheses on. Most of the theories presented in this chapter were initially proposed as descriptions for monotonic deductive reasoning. Consequently, it is unavoidable to first describe their general assumptions, to then specify how each theory was adapted to account for defeasible reasoning.

1.1.2.1. Mental Rules Theories

Mental rules theories propose that human beings have some kind of abstract general reasoning rules that they apply to draw inferences. Inferences are thought to be made by first uncovering the logical form of the premises, then applying abstract reasoning rules to arrive at a conclusion, and finally, translating this abstract conclusion to the content of the premises (Evans et al., 1993; see also Braine, 1978; Braine & O'Brien, 1991). Mental rules theories are therefore syntactic theories, which only depend on the form of the arguments and not on their meaning (Knauff, 2006). The best known theories based on mental rules are those from Rips (1994) and Braine (Braine, 1978; Braine & O'Brien, 1991; 1998; Braine, Reiser, & Rumain, 1984). Braine explains deductive reasoning in terms of three components. The first component contains the inference schemas. The second is a reasoning program in charge of deciding which inference schemas should be applied. The third component is the comprehension component, which is based on pragmatic principles that are needed to obtain a semantic representation of the premises on which to apply the inferences schemas later on. The inference schemas are thought to be 13 (though in some papers 16 are reported; Knauff, 2006), e.g., MP, double negation, *and*-elimination, and *and*-introduction. Not all inference schemas are also found in logical formal systems. Rather, they are thought to represent some kind of natural logic, which also takes into account linguistic constraints (Knauff, 2006). In the case of conditional reasoning, the relevant inference schemas are *MP* and the *Schema for Conditional Proof* (CP), both contained in the *lexical entry* of the logical operator "if" in semantic memory. The MP schema states: "Given *if p then q* and *p*, one can infer *q*" and can be used to immediately reach conclusions from premises containing *p* and *if*-statements. CP states: "To derive or evaluate *If p then* …, first suppose *p*; for any proposition *q* that follows from the supposition of *p* taken together with other information assumed, one may assert *if p then q*", and can be thus used to derive *if*-statements. Conclusions are reached according to Braine and O'Brien (1991) by successively matching inference schemas against the premises. Each conclusion reached by the application of one inference schema is added to the premise set. A conclusion is considered true if there is a match between the conclusion provided in the inference task and the conclusion reached by the application of inference schemas. If there is no match, the conclusion was false. Along these lines, people's high endorsement of MP inferences is explained by the existence of the MP inference schema: MP inferences follow thus directly from the lexical entry "if" (Braine & O'Brien, 1991). MT inferences are more

difficult because more rules, e.g. reduction ad absurdum, have to be applied to reach a conclusion. AC and DA fallacies are explained by assuming that reasoners are invited by conversational implicatures to interpret the conditional as a biconditional (see Politzer, 2007; Evans et al., 1993). According to Braine's mental rules theory errors can therefore happen due to (1) comprehension errors (i.e., wrong interpretation of the premises), (2) strategy errors (i.e., difficulties in accessing some inference schemas), or (3) process errors (i.e., working memory constraints do not allow a correct application of the inference schemata; see Evans et al., 1993; Knauff, 2006). Similar to Braine's account, Rips (1994) also proposed that people apply mental rules during reasoning. Rips assumes that conclusions are drawn by the application of rules, each of them necessary to reach subgoals towards the final conclusions. However, Rips' inference schemas are more attached to the rules of formal logic than Braine's rules (Knauff, 2006). Another difference is that instead of assuming only MP and CP (here: if-introduction) as inference schemas for conditionals, he also proposes a MP backwards rule which is needed to reach subgoals.

Mental rule theories have received criticism. The main problem is that they do not specify how the structure of natural language propositions is transformed into semantic representations (Politzer, 2007). Consequently, mental rules theories cannot fully explain how the content of natural language conditionals affects reasoning (Evans et al., 1993). One might also criticize the fact that mental rules theories cannot account for defeasible reasoning because the rules are underlain by classical logic. However, there have been some attempts to adapt mental rules theories to defeasible reasoning. For instance, Politzer proposed that the MP and CP inference schemas can explain how uncertain conclusions can be reached by uncertain premises (Politzer & Bourmaud, 2002). He argues that conditionals are usually understood as "If {N}& *p*, then *q*", where N indicates implicit complementary necessary conditions (CNC) which are conditions that are necessary for *p* leading to *q*. When reasoners are confronted with defeaters, they are not sure anymore whether the CNC are actually fulfilled. This uncertainty is then transmitted to the conclusion by interplay between the MP and the CP inference schemas (Politzer & Bourmaud, 2002). However, Politzer's approach is not described in much detail and he declared years later that this approach has not been tested yet (Politzer, 2007). Stenning and van Lambalgen's (2004; 2005; 2008) non-monotonic approach to conditional reasoning can be also interpreted as an attempt to explain defeasible reasoning by the application of mental rules. However, one has to be cautious when relating Stenning and van Lambalgen's account to mental rules theories, because the kinds of rules pertain to different logics. Mental rules theories have the connotation of being monotonic

rules from classical logic, like it was in the approaches of Braine (1978) and Rips (1994). Instead, the rules proposed by Stenning and van Lambalgen belong to non-monotonic logics. Stenning and van Lambalgen's (2004; 2005; 2008) starting point is that before being able to reason with conditionals people must assign a *logical form* to the sentence. The assignment of logical form is tantamount to interpretation. Stenning and van Lambalgen call this process "reasoning *for* an interpretation", while further inferences with the conditional are "reasoning *from* that interpretation". Depending on how a conditional is interpreted, it may fall under the jurisdiction of different logics and hence correspondingly different inferences can be drawn. Stenning and van Lambalgen (2010) argue that conditionals are generally understood as default rules. For example, a conditional "if *p*, then *q*" is understood as "If *p* and nothing abnormal is the case, then *q*" (i.e., *if p & ¬ab, then q*, where *ab* stands for abnormality and represents defeaters). The decision about abnormalities is guided by *close world reasoning* (CWR). CWR states that a proposition is false, if we do not have reasons to believe that it is true. That is, normally reasoners assume that nothing abnormal is the case, i.e., no defeaters are available in the given situation. However, upon explicit evidence of abnormalities via, e.g., memory retrieval or explicit presentation in inference tasks (see Section 1.1.1.1), they incorporate them in the interpretation. For instance, in the sunscreen example reasoners usually assume the default situation of Jack putting sunscreen on his face when the sun is shining. Only if they become aware of abnormalities, i.e., defeaters such as the sunscreen bottle being empty, then they withdraw the conclusion that Jack puts sunscreen on his face. People therefore engage in defeasible reasoning whenever they are aware that abnormalities exist. Stenning and van Lambalgen's (2004; 2005; 2008) account can be seen as a mental rules theory in the following sense. The CW assumption, or the initial interpretation of the conditional as a default, 'absolutely normal' rule, can be seen as some kind of implicit and automatic mental rule used to understand conditionals and draw further inferences. However, different from the classical mental rules theories of Braine (Braine & O'Brien, 1991) and Rips (1994), Stenning and van Lambalgen place the emphasis on the interpretative processes necessary to reason with conditionals. In fact, most of the rules they propose are thought to be used to construct a certain interpretation of the conditional, e.g., the CW assumption, which in turn selects a specific set of rules to be applied in further reasoning about that conditional.

1.1.2.2. Mental Models Theory

Different from Braine and O'Brien's (1991; 1998) mental rules approach, in mental models theory (Johnson-Laird & Byrne, 1991) reasoning is understood as a semantic process, where the meaning and the interpretation of premises are considered (Knauff, 2006). Instead of simply applying abstract rules to the premises, mental models theory assumes that the information in the premises is first represented as mental models. These models are then inspected, evaluated, and conclusions are drawn. According to mental models theory, reasoning from conditionals has three steps (Johnson-Laird & Byrne, 1991). The first is *comprehension*, when reasoners use their general and linguistic knowledge to construct mental models of the state of affairs described in the premises. In the second *description phase*, these mental models are inspected in order to arrive at a putative conclusion. This conclusion should contain new information not explicitly given in the premises. If reasoners cannot arrive at such a conclusion, they conclude that nothing follows. But if a putative conclusion is found, they proceed with the third phase, the *validation phase*, and search for counterexamples to their putative conclusion. That is, they search models in which the premises are true, but their putative conclusion is false. When there are no counterexamples, the putative conclusion is valid. Finding counterexamples deems the conclusion false, and reasoners have to return to the second phase in search for an alternative conclusion for which no counterexamples exist.

Mental models follow the principle of truth, meaning that only what is true is represented. As explained in Section 1.1, according to material implication a conditional "if *p*, then *q"* is consistent with three true alternative situations: *p* and *q*, ¬*p* and *q*, and ¬*p* and ¬*q*. Accordingly, one could construct three mental models from a conditional, one for each true possibility (Byrne et al., 1999). However, because of working memory limitations, reasoners first construct just one mental model, the *explicit mental model p* and *q*, which represents the information explicitly provided in the conditional. All other possible mental models (¬*p* and *q*, and ¬*p* and ¬*q*) are only represented implicitly. For instance, for the conditional "If the sun is shining, then Jack puts sunscreen on his face" reasoners construct the following models:

Sun sunscreen

…

The three dots represent the implicit models, indicating that there might be other models for this conditional. Only if required by the task, these implicit mental models are fleshed out into

fully explicit mental models. This happens, for example, when reasoners are confronted with an MT inference. The second premise of a MT inference states ¬*q*, which is not represented in these initial mental models of the conditional. Therefore fully explicit mental models are necessary to infer ¬*p* follows from ¬*q*:

Sun sunscreen
¬sun sunscreen
¬sun ¬sunscreen

Mental models theory explains errors in reasoning by difficulties in searching for counterexamples, or by problems in fully fleshing out mental models (Johnson-Laird & Byrne, 1991; Johnson-Laird, Khelmani, & Goodwin, 2015). Along these lines, mental models theory also explains why some inferences seem harder than others: the more mental models are necessary to arrive at a conclusion, the more difficult the inference is (Johnson-Laird & Byrne, 1991). MP inferences are easy because reasoners can infer that *q* follows from *p* without the need to construct fully explicit mental models. Instead, the valid conclusion can be read off directly from the initial explicit mental model. MT inferences are more difficult, because as explained in the previous passage, reasoners have to construct fully explicit mental models in order to conclude that ¬*p* follows from ¬*q*. The difficulty of AC and DA inferences is explained on the one hand similarly. Mental models theory also posits an alternative explanation, namely that reasoners interpret the conditional as a biconditional of the form "if and only if" constructing the following explicit model:

[sun] [sunscreen]

The brackets indicate that [sun] and [sunscreen] are exhaustively represented, meaning that there are no other models that include sun or sunscreen (Johnson-Laird & Byrne, 1991). Therefore reasoners erroneously conclude that *p* follows from *q*, or that ¬*q* follows from ¬*p*.

In most of the conditional reasoning literature, mental models theory is used to explain deductive monotonic reasoning. However, mental models theory has been expanded to cover non-monotonic, defeasible reasoning as well (Byrne et al., 1999; Johnson-Laird & Byrne, 2002). The consideration of defeaters (more specifically disabling conditions) is explained by mental models theory by assuming that reasoners introduce other necessary conditions for *q* to happen as additional antecedents. Imagine, for instance, that a reasoner is confronted with the conditional "If the sun is shining, then Jack puts sunscreen on his face" and considers the

possibility of a defeater, for example that the bottle of sunscreen is empty. This results in the following mental model:

Sun empty bottle sunscreen

…

Because people know that it is not possible to use sunscreen from an empty sunscreen bottle, they will additionally construct the following models based on general knowledge:

empty bottle ¬sunscreen
¬empty bottle sunscreen
¬empty bottle ¬sunscreen

Thus, when people have to decide whether *q* follows from *p*, they will combine the models from background knowledge with the models of the premises, resulting in something like:

Sun empty bottle sunscreen ¬sunscreen

Johnson-Laird and Byrne (2002) state that general knowledge has priority here, allowing reasoners to conclude that, in case the sun is shining and the sunscreen bottle is empty, Jack will not put sunscreen on his face. This is called the principle of *pragmatic modulation* (Johnson-Laird & Byrne, 2002).

Mental models theory therefore describes the consideration of defeaters in terms of counterexample search in general knowledge. However, the principle of pragmatic modulation appears somewhat arbitrary – it is not clear why the model constructed from general knowledge about defeaters has priority over the mental model constructed from the information in the conditional. Politzer (2007) points out that in fact the relationship described in the conditional is part of general knowledge, too. So, why is the defeater weighted more? Further, some researchers argue that by relying on material implication, mental models theory is not as different from mental rules theories as proposed (e.g., Evans, 2008; Oaksford & Chater, 1995a). Another problem of mental models theory is that it has difficulties in explaining why the frequency of exceptions sometimes overweighs the mere amount of defeaters, and, in general, how different degrees in beliefs in the conditionals and conclusions are represented. Stevenson and Over (1995) made an attempt to explain frequencies and probabilities within mental models theory. They argue that within a mental model, one can represent proportions of representative cases, so that the belief in a conclusion depends on the proportions of models where this conclusion holds (see also Johnson-Laird, 2001). For

instance, the conditional "If the sun is shining, then Jack puts sunscreen on his face" may result in a model where every time the sun is shining, Jack puts sunscreen on his face:

Sun Sunscreen
Sun Sunscreen
Sun Sunscreen
Sunscreen

However, when people are aware of defeaters (i.e., that *p* sometimes happens without *q*) the exceptional cases are added to the model:

Sun Sunscreen
Sun Sunscreen
Sun Sunscreen
Sunscreen
Sun
Sun

Depending on how many "sun sunscreen" models are in comparison to "sun" models, a conclusion may be believed to different degrees. Johnson-Laird, Legrenzi, Girotto, Legrenzi, and Caverni (1999) also proposed that different degrees of belief in conclusions can be explained by the proportion of models in which this conclusion holds. Additionally, they claim that this proportion is calculated by assigning numerical values to the models. The numerical values represent frequencies. For instance, if there are 20 observations, and *pq* happens ten times, ¬*pq* four times and ¬*p*¬*q* two times, then each model has a notation of these frequencies and one can calculate the probability of the conditional (cf. Geiger & Oberauer, 2010). Although the incorporation of frequencies into mental models seems plausible, it is not clear to what extent such extended mental models are feasible for working memory. Johnson-Laird and Byrne (1991) already claimed that constructing fully fleshed out models is demanding for working memory. So, how could these models be additionally expanded to represent the proportion of cases in which certain models hold? This is especially problematic if we bear in mind that, on the one hand, reasoners do consider defeaters during reasoning, but, on the other hand, according to mental models' principle of truth instances of *p*¬*q* should not be explicitly represented (Geiger & Oberauer, 2010). That is why Geiger and Oberauer (2010) proposed a different approach to incorporate probabilities in mental models theory. They argue that reasoners start by constructing a mental model of the minor premise

(e.g., *p* in the case of MP), which is then expanded by the model of the other component of the conditional (e.g., *q*). Afterwards, reasoners search for information in memory regarding the probability of the situations in each model (i.e., *p* and *pq*). The subjective probability of the conditional corresponds to the ratio of P(*pq*) to P(*p*). However, to my knowledge Geiger and Oberauer's approach has not been tested empirically yet.

1.1.2.3. Probabilistic Theories

Many recent theories on conditional reasoning assume that people draw inferences by considering the probabilities of the premises. According to probabilistic theories people treat conditional statements probabilistically by understanding the probability of a conditional, P(if *p* then *q*), as the conditional probability of *q* given *p*, P(*q* | *p*) – a relationship known as the Equation (Edgington, 1995; Evans & Over, 2004). One way in which people calculate this probability is by performing the Ramsey test (1929/1990; see also Evans, Handley, Over, 2003). According to the Ramsey test, people calculate the conditional probability by first assuming that *p* holds. On the basis of this hypothetical belief people then calculate how probable it is that *q* follows by computing the ratio between the instances were *q* happens and those in which *q* does not happen. For instance, the conditional probability of "If the sun is shining, then Jack puts sunscreen on his face", reasoners first assume that the sun is shining. Then they start thinking about how often Jack actually puts sunscreen on his face when the sun is shining and how often he does not when the sun is shining. In other words, they assign probabilities to "The sun is shining and Jack puts sunscreen on his face" and "The sun is shining and Jack does not put sunscreen on his face". The higher the probability of P(*pq*) is relative to P(*p*¬*q*), the higher the conditional probability P(*q* | *p*) is perceived to be, and the higher also the perceived probability of P(if *p* then *q*) (Evans & Over, 2004).

The main advantage of probabilistic theories is that consideration of probabilities is well suited to human everyday reasoning, which is hardly ever in terms of "all or nothing". Probabilities capture the different degrees of certainty that exist in our daily lives. Contrary to rule based theories or mental model theory, probabilistic theories do not treat premises as certain. Instead, probabilistic approaches offer tools to model how uncertain premises can be combined to reach uncertain conclusions. Directly linked to the idea of capturing the uncertainty of everyday reasoning, probabilistic theories are also perfectly suitable to explain defeasible reasoning. This is because all information that influences the conditional

probability of *q* given *p* will also affect inferences. For example, additional information which lowers the perceived probability of *q* given *p* will also lower the degree of certainty on the conclusion, and, as a result, conclusions drawn without considering this new information are defeated. The more disabling conditions exist for a given conditional, the higher is the probability of "*p* but ¬*q*" instances, and the lower the conditional probability and the probability of the conditional P(if *p* then *q*) gets. In fact, several studies have already shown that probabilistic based theories can well predict human reasoning performances, sometimes even better than accounts based on material implication like mental models theory. For instance, Liu, Lo, and Wu (1996) varied the perceived conditional probability of *q* given *p* and found that the higher the perceived probability of the conditionals, the more logically correct responses people endorsed. Accordingly, Evans et al. (2003) showed that the perceived probability of a conditional is best predicted by conditional probability and not by material implication. The authors asked participants to imagine a pack of cards, whose cards are either yellow or red and have either a circle or a diamond on them. In addition, Evans and colleagues gave participants an abstract conditional "If the card is yellow, then it has a circle printed on it" and the corresponding frequencies of *pq*, *p*¬*q*, ¬*pq*, ¬*p*¬*q*. The participants' task was to indicate how likely they think it is that the conditional is true for a randomly selected card from the pack. If participants understand the probability of the conditional as the conditional probability, then their estimated likelihood of the conditional should correspond to $P(q \mid p) = P(pq)/[P(pq)+P(p\neg q)]$. If they instead understand the conditional as material implication, then the probability of the conditional should only depend on the true cases of the truth table: $P(MC) = P(pq) + P(\neg pq) + P(\neg p\neg q)$. Clear evidence against material implication was found. Instead, conditional probability was the best predictors for the probability of the conditional. Also the conjunctive probability P(*pq*) was well supported by the data, but the conjunctive probability cannot explain why the probability of the conditional raised almost linearly with increased *pq* frequencies, and decreased with increased *p*¬*q* frequencies. This can only be explained by the conditional probability. Evans' et al. findings were replicated by Oberauer and Wilhelm (2003) and Over, Hadjichristidis, Evans, Handley, and Sloman (2007). Especially interesting in this regard is the work of Over et al., because contrary to Evans et al. (2003) and Oberauer and Wilhelm (2003), they did not use abstract, but real world conditionals (e.g., "If the cost of petrol increases, then traffic congestion will improve"). This means it was not necessary to provide participants with frequency or probability information. Instead, Over and colleagues could ask for participants' subjective probabilities in order to see

to what extent these predict the probability of the conditional. Similar to previous findings, Over et al. found that conditional probability was the best predictor.

Because of the suitability of probabilistic approaches to everyday reasoning, many researchers have adopted the idea to incorporate probabilities in their theories. In fact, much of the work in the "new paradigm" of cognitive psychology actually assumes a probabilistic approach to conditionals (Elqayam & Evans, 2011; Evans, 2012; Oaksford & Chater, 2013; Pfeifer, 2013). On the one hand, there is the suppositional theory account of Evans and Over (2004). Evans also argues that the probability of the conditional equals to the conditional probability. MP inferences are therefore explained by arguing that the belief in the conclusion is a function of the probability of the conditional and thus a function of conditional probability (i.e., if *p* is given, how probable is it that *q* follows? see Geiger & Oberauer, 2010). MT inferences are also explained by suppositional inference. It is argued that reasoners first assume *p* and derive through MP that *q* follows. However, given that the categorical statement of MT inferences states ¬*q*, reasoners are confronted with a contradiction. This contradiction leads the reasoners to conclude that the previously made supposition is false, and so *p* must also be false (Evans & Over, 2004). AC and DA inferences are explained by assuming that reasoners add to the premises some new interpretation of the conditional. For AC inferences the converse of the conditional (if *q* then *p*) is added to premises, while for DA inferences its inverse (if ¬*p* then ¬*q*) (Evans and Over, 2004; see also Geiger & Oberauer, 2010).

On the other hand, there is also Oaksford and Chater's (2007; 2009) probabilistic theory, which is one of the first psychological theories of reasoning based on probabilities (Geiger & Oberauer, 2010). Oaksford and Chater were also among the first psychologists to emphasize the necessity to incorporate non-monotonic and defeasible reasoning in attempt to understand human reasoning (Oaksford & Chater, 1991; 1995a). Oaksford and Chater (2003b; Oaksford, Chater, & Larkin, 2000) propose that a person's degree of belief that a MP conclusion can be drawn depends on the belief in the conditional and thus on the conditional probability:

$$P(MP) = \mathrm{P}(\mathrm{q} \mid \mathrm{p}) = 1 - \varepsilon$$

Hereby ε is called the exception parameter, which corresponds to the probability of ¬*q* given *p*. ε depends on the availability of disabling conditions: the more disabling conditions there are for a given conditional, the higher the ε is, and consequently the less MP inferences are drawn. The exception parameter thus prevents the conditional probability from being 1. Similarly, the probability that a person draws a MT inference is described as:

$$P(MT) = P(\neg \mathrm{p}|\neg \mathrm{q}) = \frac{1 - P(q) - P(p)\varepsilon}{1 - P(q)}$$

Also here, the higher the exceptions parameter is, the smaller the enumerator gets and the less a MT inference is drawn. The probability of inferring MT is thus the probability of ¬*p* given ¬*q*.

Oaksford and Chater explain people's endorsement of AC and DA inferences in a similar way:

$$P(AC) = P(p|\mathrm{q}) = \frac{P(p)(1 - \varepsilon)}{P(q)}$$

$$P(DA) = P(\neg \mathrm{q}|\neg \mathrm{p}) = \frac{1 - P(q) - P(p)\varepsilon}{1 - P(p)}$$

Given that probabilistic theories do not assume mental representations of concrete *p*, *q*, ¬*p*, or ¬*q* instances, they can explain why in addition to the amount of defeaters, also the overall frequency of exceptions affects inferences. Contrary to mental models theory, probabilistic approaches do not require people to construct mental models from general knowledge to be aware that *p* can happen without *q*. Instead, the effect of defeaters is operationalized by the probability of ¬*q* but *p*, which is directly related to the frequency of exceptions. In other words, theories based on probabilities can predict the inferences people draw by computing the conditional probability. Along these lines the effect of trustworthiness can also be explained: conditionals uttered by highly trustworthy speakers are considered to have a higher conditional probability than those uttered by low trustworthy sources (cf. Stevenson & Over, 2001).

Probabilistic theories also have disadvantages. On the one hand, we seldom have concrete knowledge about frequencies or probabilities in our daily life in order to be able to compute probabilities for MP, MT, AC or DA inferences. One solution for this problem may be to assume that these are not factual frequencies, but subjective frequencies based on our experiences and semantic knowledge (see e.g. Over et al, 2007). However, the main disadvantage of probabilistic theories is that they do not describe what actually happens in the head of participants – they are no process models. Instead, they only describe on a computational level which conclusions will be drawn or withdrawn (Oaksford & Chater, 1995a; 2003a). So, there is still the open question of what actually happens online in the reasoners' minds. It is difficult to assume that people actually make these computations in

their minds. Several studies have shown that people are bad at dealing with probabilities (Kahneman, Slovic, & Tversky, 1982; Tversky & Kahneman, 1974). People reason better with frequencies (e.g., Kahneman et al., 1982), but even so it is still implausible than people run computations as complex as needed to reach or reject MT, AC or DA inferences. One solution to this problem could be that people only have rough categories of how probable something is, as for example: very probable, somehow probable, neither/ nor, somehow improbable or very improbable. It is certainly necessary to further investigate how probabilistic theories could go beyond computational level theories, and describe what happens in the minds of reasoners during inference tasks.

1.1.2.4. Dual Process Theories

In an attempt to explain why under some circumstances people are capable of reasoning deductively according to classical logic but in others not, many researchers have proposed dual process theories (e.g., Evans & Over, 2004; Sloman, 1996; Stanovich & West, 2000; Verschueren, Schaeken, & d'Ydewalle, 2005). Dual process theories state that depending on the task, time constraints, cognitive capacity, motivation, etc., people may use two kinds of thinking or modes of processing (Evans & Stanovich, 2013; Markovits, Brunet, Thompson, & Brisson, 2013). One kind is usually described as being heuristically driven, fast, automatic, unconscious, and effortless; it is often labeled Type 1 (e.g., Evans, 2012; Evans & Stanovich, 2013) or System 1 (e.g., Kahneman & Frederick, 2002). The other – labeled Type 2 or System 2 – is described as being analytic, slow, controlled, dependent on working memory, and deliberative. System 1 is therefore often seen as the default system, which can be either overridden or supported by System 2 (Evans, 2006; 2008; Singmann, Klauer, & Over, 2014). Besides cognitive psychology, dual process theories have also been used in other areas of psychology (Evans, 2008), for example, to explain attitudes change (Petty & Cacioppo, 1986) or moral reasoning (see Section 1.2.2).

Evans and Over (2004) argue that System 1 is responsible for implicit pragmatic inferences (e.g., "If the sun is shining, I use sunscreen"), which are thought to be the default way of reasoning. Only when the task requires and people are willing to make additional efforts System 2 comes into play and people facilitates engagement in deductive reasoning. The implicit inferences made by System 1 can sometimes look like deductive inferences; but actually they are reached on the basis of beliefs and uncertain premises, for example when

people infer that *q* follows from *p* because it is very likely that *q* follows from *p*. Along these lines, Evans and Over (2004) propose that most MP inferences are made by System 1, thus explaining why it is so easy to defeat MP conclusions. MT inferences are thought to require System 2, because reasoners' usual approach of supposing *p* and then deriving from this *q*, is contradicted by the categorical statement of the categorical premise ¬*q*. That is to say, automatic processes cannot deal with MT inferences, which call for additional deliberative processing. As for AC and DA inferences, they are thought of as a result of System 1. However, just like for MP inferences, Evans and Over argue that under some circumstances AC and DA inferences might involve explicit processing effort from the part of System 2.

There is a wide variety of evidence in favor of dual process theories in cognitive psychology. For instance, Verschueren et al. (2005) found that fast and slow responses in inference tasks are best explained by the postulation of two types of reasoning. Fast responses were best explained by the likelihood of *q* given *p* (for MP and MT inferences) or the likelihood of *p* given *q* (for AC and DA inferences). Slow responses were instead best described by the availability of disabling conditions or alternative causes. Verschueren and colleagues argue that Type 1 reasoning is based on probabilistic reasoning, whereas Type 2 is based on a counterexample search like the one proposed by mental models theory. Similar effects of time constraints on the usage of Type 1 or Type 2 reasoning were found by Evans & Curtis-Holmes (2005) and Markovits et al. (2013). Moreover, De Neys (2006a) found that when working memory load is high, people make more errors (according to classical logic) because they start reasoning according to their beliefs instead of considering the logical validity of the argument. Also evidence that intelligence correlates with correct deductive reasoning, but not with belief-based reasoning, supports the idea of two different ways of thinking (Evans, 2008; Evans et al., 2010; Stanovich & West, 2008). Further evidence in favor of dual process accounts comes from Weidenfeld, Oberauer, and Hörnig (2005). In a path model they showed that disabling conditions predicted inferences in two ways: via a direct pathway affecting endorsement of valid inferences, and via an indirect pathway, where disabling conditions first affected conditional probabilities which then affected the endorsement of inferences.

The main advantage of dual process theories is that they can explain a vast range of findings. The assumption of two distinct systems can also explain why people sometimes reason deductively and sometimes do not. Deductive reasoning performance and consideration of defeaters as counterexamples are usually explained by System 2 (see Verschueren et al., 2005). Considering uncertainties, different degrees of believability, and

reasoning under time constraint, are explained by System 1. However, one main problem of dual process theories is that it is not clear whether the two systems are qualitatively different or rather two extremes of a continuum (Kruglanski, 2013; see also De Neys 2006a; Osman, 2013). In the latter case it would still be difficult to fixate a turning point along the continuum (Varga & Hamburger, 2014). It also can be criticized that researchers talking about dual process theories often evoke the impression that there is only one dual process theory on reasoning. But dual process theories differ: whereas some researchers propose that the two systems work in parallel, others argue that the relationship is sequential, yet others do not make any assumptions about the underlying cognitive processes (Evans, 2008). It is therefore not clear what exactly distinguishes the two systems nor how they are cognitively implemented. Also the attributes assigned to each system differ across dual process theories, and some of these "typical" characteristics are questionable. For instance, some propose that System 1 is content dependent and System 2 abstract. But according to Evans (2008) this is not adequate, since System 2 is also relevant for reasoning with content rich material, as shown by Verschueren et al. (2005). The same happens for the automatic vs. non-automatic dimension. Osman (2013) pointed out that reasoning performances thought to be carried out by System 1 do not differ significantly in time from those carried out by System 2. Under some circumstances these "automatic" processes can take up to 26 or 47 seconds, throwing doubt onto the degree to which these processes are indeed automatic. Therefore, according to Evans (2008) the only reliable finding is that System 2 seems to depend on working memory, whereas System 1 does not. But also Osman (2013) highlights here that even reasoning performance thought to be carried out by System 1 can be impeded by working memory load (see De Neys, 2006b).

An alternative approach to explain how different conclusions are reached is proposed by Klauer, Beller, and Hütter (2010). In their dual source theory they propose that inferences depend on both, people's background knowledge about the content of the premises *and* the logical form of the inference. The background component is influenced by the conditional probability of *q* given *p*. It represents therefore the degree to which people endorse a conclusion given their background knowledge about the premises. The form component instead depends on the subjective probability of the logical inference presented in the task (Klauer et al., 2010; Singmann et al., 2014). The form component does therefore not refer to the actual logical status but to the "belief in the logicality of logical forms" (Singmann et al., 2014, p. 4), i.e., the perceived logical correctness of an inference. Based on studies employing abstract material and strict deductive instructions it has been shown that this perceived logical

correctness is highest for MP inferences, followed by MT inferences and AC and DA inferences. When confronted with an inference task, participants thus consider both, the content and the logical form of the inference. This explains why participants, when they are asked to estimate how probable it is that *q* follows from *p*, give higher ratings when they are first confronted with the corresponding "if-then" conditional rule than when they are asked (directly) without having been presented with a rule before (Klauer et al., 2010; see also Liu, 2003). The dual-source model thus differs from dual process theories, to the extent that it does not propose that people sometimes only consider the content while sometimes do abstract reasoning, but that people integrate both sources of information and weigh them according to the emphasis placed by task instructions either on content or on context variables (Klauer et al., 2010). Accordingly, Klauer et al. (2010) say that they abstain from ascribing the typical attributes of dual-process systems (e.g., automaticity, efficiency, etc.) to one of their dual sources.

In sum, although under some circumstances people seem to answer faster, less reflective and more belief-based, this does not necessarily mean that two different cognitive systems exist. On the one hand multiple cognitive or neural systems may account for the different types of processing (Evans & Stanovich, 2013). On the other hand, instead of qualitative, also quantitative differences in reasoning are plausible: many seemingly analytically or automatically reached conclusions can also be explained by assuming a single system that evokes different conclusions depending on the task difficulty (Osman, 2013).

1.1.2.5. Theories on the Utilities of Conditionals

Theories linking utilities to reasoning are not as wide spread as the other theories mentioned in this chapter. Nonetheless, utility based reasoning theories have gained attention in the last years. The main assumption is that the conclusions people draw or withdraw also depend on the utility of this conclusion for some agent (Bonnefon, 2009). Depending on the antecedent and/ or the consequent, some conclusions have more or less utility for reaching an agent's goals.

The role of utilities in conditional reasoning is often investigated with *utility conditionals*, which describe actions and their consequences (Bonnefon, 2009; 2012; Bonnefon, Girotto, & Legrenzi, 2012; see also Bonnefon & Hilton, 2004). The actions are usually described in the antecedent of the conditional, and the consequences – which can be

positive, negative or neutral – in the consequent, for example "If you turn the radio on one more time, then I will hit you". The participants' task is to decide whether an agent would perform the action described in the antecedent. Bonnefon showed that people conclude more often that an action will be taken if it is of high utility for the agent. That is, people attribute goals and motivations to the agents and only conclude that the action will be taken if the benefits of this action overweight the costs of the consequences (see also Evans, Neilens, Handley, & Over, 2008; Ohm & Thompson, 2006). This happens according to Bonnefon by folk axioms (Bonnefon, 2009). Bonnefon assumes that people have naïve understandings about how other people make decisions. The most important folk axioms in this respect are the folk axioms of *Self-Interested Behavior* and the one of *Self-Interested Attitude*. Self-interested behavior states that reasoners believe that agents take actions that increase their own personal utility. In the example above, reasoners would thus conclude that the agent will not turn the radio on. Self-interested attitude states that reasoners think that also other agents will perform actions which increase their personal utility. For example, given the conditional "If you testify against me, then you will have an accident" reasoners assume that the agent will not perform the action described in the antecedent (Bonnefon et al., 2012).

Utilities can also be used to explain defeasible reasoning. Bonnefon and Hilton (2004) argue that defeaters sometimes provide information about utilities which can be used to decide whether an inference should be drawn. In their experiment, for instance, they showed that when a conditional such as "If Marie's TV is broken, she will have it fixed" is followed by an utility conditional implicating defeaters like "If Marie has her TV fixed, she will not be able to pay the electricity bill", people refuse to endorse the MP inference that Marie will have the TV fixed given that her TV is broken. That is, they consider that the costs of not paying the bill are higher than not having a TV working and therefore defeat the MP conclusion. Similarly, Elqayam, Thompson, Wilkinson, Evans, and Over (2015) showed that when participants are confronted with conditionals like "If Robert wears the Rieti suit, then he will be offered a better job than he would have been otherwise", and with additional information such as "The Rieti suit has been made in a factory exploiting child labor. If Robert wears the Rieti suit, then he will be supporting child slavery", participants withdraw from the conclusion that Robert should wear the Rieti suit.

Besides utility conditionals, the role of utilities in reasoning has been also investigated with deontic conditionals. Deontic conditionals (roughly introduced in Section 1.1.1.2.2) describe what *should*, *may* or *must* happen, such as "When the sun is shining, then Jack *should* put sunscreen on his face". As already explained in Section 1.1.1.2.2, reasoning from

deontic conditionals can be best explained with utilities. For instance, if participants are confronted in an inference task with the conditional "When the sun is shining, then Jack *should* put sunscreen on his face" and are asked whether *q* should happen given *p*, then their answer depends on how much more they prefer the outcome *pq* over *p*¬*q* (Over et al., 2004). Perspective effects in reasoning can also be explained by utilities. Imagine you are confronted with a conditional rule uttered by a mother to her son: "If you tidy your room, then you may go out to play", and are asked to decide when this rule is violated. Manktelow and Over (1991) showed that people's perception about rule violation depends on the utilities of each outcome, which in turn depend on the perspective taken by the reasoner. For instance, if participants are asked to imagine that they are the mother uttering this conditional rule, then they say that the rule is violated if the son goes out to play without having tidied his room (¬*pq*). However, if participants are asked to imagine they are the son, then they say the rule is violated if the son tidies the room but is not allowed to go outside to play (*p*¬*q*). Manktelow and Over (1991) argue that this different perception of when the rule is violated results from different subjective utilities attributed to persons uttering rules and to persons obeying rules.

The idea that utilities can influence reasoning conflicts with many traditional approaches on reasoning. In the past, there has been a clear division between research on reasoning and on decision making. The reasoning community usually investigated deductive reasoning competence, and the decision making community investigated how utilities affect decision making (Bonnefon, 2009; Evans, 2012). However, the idea of breaking this division and considering utilities in investigations of reasoning too, is promising. In fact, taking into account utilities in reasoning is considered one of the main characteristics of the "new psychology of reasoning" introduced in Section 1.1. In everyday situations it is difficult to disentangle reasoning from decision making: people usually reason before deciding. Imagine for instance that someone gives you lots of candies so that you have to decide if you eat them or not. In order to make this decision, you might start thinking about your experiences with eating candies and perhaps remember that: *If a person eats lots of candies, then the person will get cavities*. How would you decide? If we allow utilities to affect reasoning processes, then we might conclude that people's decision about eating the candies reflects a reasoning process based on the utilities attributed to the antecedent and the consequent of the conditional. For instance, if we really love candies and the candies we got are our favorite ones, then the benefits of eating them might overweigh the costs of getting cavities. As a result, one might conclude that it ok to eat the candies.

Of course it is not possible to explain all sorts of conditional reasoning by utilities. There are for sure conditionals were no utilities are applicable, for example in cases were conditionals do not describe any actions nor present any agents, such as "If it rains, then the streets are wet". However, there are still many conditionals were utilities can help to understand which conclusions are drawn. In fact, the main advantage of opening the boundaries between reasoning and decision making is that it permits the investigation of many other types of reasoning which may be on a first sight more related to practical reasoning and decision making. One of these is the domain of *legal reasoning*, the focus of this thesis, which is described in the next section.

1.2. Legal Reasoning

So far, I have described how conditional reasoning is investigated in cognitive psychology. I have explained how psychologists' research interest has moved to the investigation of defeasible reasoning and how the current reasoning theories explain the withdrawal of conclusions. The overview on the psychological literature was necessary to get a general idea of what defeasible reasoning is and to understand when people defeat conclusions. However, to investigate legal conditional reasoning, it is now necessary to take a look on the literature on legal reasoning. The goal of this section is to describe legal reasoning from a law theoretical and a social psychological point of view, and to present some empirical findings on people's legal reasoning. By doing so, in Chapter 2, I will be able to combine the findings from conditional reasoning and legal reasoning to develop hypotheses on how human beings reason with legal conditionals.

1.2.1. Legal Reasoning and Conditionals in Legal Theory

Legal reasoning is the process of reaching conclusions and decisions in legal matters. Legal reasoning has thus two aspects: reasoning about evidence, and reasoning about rules or norms as reasons for action (MacCormick, 1998). This thesis focuses on the latter, especially

because there are several similarities between how legal theory describes reasoning about rules and norms and how cognitive psychology describes conditional reasoning.

When judges decide whether a legal rule should be applied or not, they usually have to make two sets of decisions. First, they must decide what the facts are (German: *Sachverhalt*) and whether there exists any legal rule applicable to this case. This step is called *external justification* (in German: *externe Rechtfertigung*; Alexy, 1983). For instance, if the case is about manslaughter, then the judge must justify that 1) a person indeed killed someone, and 2) that manslaughter is the right rule to be applied. The judge has to look at the concrete rule for manslaughter (§212 StGB: *Whoever kills a human being, without being a murderer, is punished for manslaughter with imprisonment for not less than five years* [Wer einen Menschen tötet, ohne Mörder zu sein, wird als Totschläger mit Freiheitsstrafe nicht unter fünf Jahren bestraft]) and interpret its components (e.g., what does "killing" mean? Is this what happened in the current case?). Only like this can a judge find out if the matters of facts agree with the prescribed elements of the crime (in German: *Tatbestand*). The external justification in legal reasoning is thus the most difficult step in legal reasoning, because it also depends on a good police investigation, and on weighting the evidence and arguments presented by lawyers and attorneys. Once the judge has decided which are the facts and which rule is appropriate, the next step is the *internal justification* (in German: *interne Rechtfertigung*; Alexy, 1983). Internal justification describes how a rule and the matter of facts are combined to a reach a verdict. According to Alexy (1983), the verdict ought to be reached deductively, more precisely, it must follow the structure of a MP inference: If there is a rule that punishes crime A, and somebody commits A, then the person should be punished for A. To better combine the legal rule and the matter of facts in a deductive way, many law theorists argue that legal rules should be understood as conditionals (cf. Bäcker, 2009, 2010; Koch & Rüßmann, 1982; MacCormick, 1998; Wang, 2004). I will call this conditional formulation of legal rules as *legal conditionals*. For example, if the evidence shows that somebody killed another human and the judge decided that the rule against manslaughter is appropriate, then the internal justification should follow the following schema:

If a person kills another human, then the person should be punished for manslaughter (Rule)
Bert killed another human (Facts)

Bert should be punished for manslaughter (Legal consequence)

Given these two premises, one can follow the MP inference and conclude that if a person kills another human (*p*), then the person should be punished for manslaughter (*q*). Law theorists call this the *judicial* or *legal syllogism* (Alexy, 1983; Bäcker, 2009; 2010; Koch & Rüßmann, 1982; Wróblewski, 1992). This term came about because the internal justification was originally introduced in the form of a syllogism (e.g., *Anyone who has brought about the death of another person is to be punished. Someone has brought about the death of another person. Therefore, the person is to be punished*; Bäcker, 2010). The aim of internal justification is to make the judge's argumentation irrefutable, and this is why it is best served by deductive reasoning: a verdict is internally justified if it follows logically from its rule and facts premises (Alexy, 1983). However, as discussed in the introduction of this thesis, reaching a certain conclusion is not always that easy. Law considers circumstances that prevent people from being punished even though punishable acts were committed – in other words: legal rules are defeasible (Bäcker, 2010; Dewitz, Ryu, & Lee, 1994; Prakken, 1997; Prakken & Sartor, 2004; Sartor, 2009). Consider the following example:

If a person kills another human, then the person should be punished for manslaughter (Rule)
Bert killed another human (Facts)
Because of a psychological disorder, Bert was not able to control his actions (Exculpatory circumstance)

Bert should not be punished for manslaughter (Legal consequence)

According to penal code, the offender should not be punished anymore, because absence of criminal responsibility due to psychological disorders is exculpatory evidence (or: exculpatory circumstance) according to law and a defeater in psychological terms.

Besides lack of criminal liability due to psychological disorders, there are several other exculpatory circumstances (i.e., defeaters) regulated in penal code, such as self-defense, necessity, mistake of law, or age (lack of criminal liability of children). In the German law most of the exculpatory circumstances are written in the General Part of the Penal Code (German: *Allgemeiner Teil*) and apply therefore to (almost) all specific offences (such as manslaughter, bodily injury, theft, etc.) which are written in the Special Part (German: *Besonderer Teil*). As all defeaters, exculpatory circumstances prevent the consequent of a legal conditional to happen although the antecedent is true. In other words, exculpatory circumstances void sentence completely, although the prescribed elements of the crime are fulfilled. In German law, exculpatory circumstances can void punishment in two ways: either

they justify the offence, or they excuse it. When an offence is justified, a person is not punished because the act is not seen as unlawful. When an offence is excused, the act is still considered unlawful, but the offender is not punished because the guilt is considered so low, that punishment is not judged necessary anymore. In this thesis, I will not make the distinction about exculpatory circumstances justifying or excusing offences. But I want to notice that exculpatory circumstances are not the same as mitigating circumstances. While exculpatory circumstances void sentence completely (i.e., no punishment), mitigating circumstances only lessen a sentence (e.g., 5 instead of 8 years of imprisonment).

The characteristic of the penal code of having separate rules for specific offences and for exculpatory circumstances, respectively, illustrates the defeasibility of legal rules. However, some legal theorists argue that this is no real defeasibility (Bäcker, 2012). The exculpatory circumstances presented in the last passage are actually all written down in the penal code. In this way, one could argue that such exculpatory evidence could simply be added to the antecedent as additional requirements for applying the rule (e.g., If somebody kills another human, and the person has no relevant psychological disorders, then the person should be punished for manslaughter), allowing still deductive conclusions. Some researchers therefore call this kind of defeasibility emerged from the structure of the penal code *improper defeasibility* (in German: unechte defeasibility; see Bäcker, 2012). However, this "improper defeasibility" labeling does not mean that legal rules are not defeasible. Legal theorists still acknowledge the defeasible character of legal rules; they argue that legal reasoning is defeasible because even though many exculpatory circumstances are prescribed in the penal code and could be added to the antecedent, it is not possible to know beforehand which defeaters will be relevant in future cases. That is to say, exculpatory evidence cannot be enumerated exhaustively (Bäcker, 2010; Wang, 2004). There will always be new cases and one cannot foresee all circumstances that will be relevant for these cases. This impossibility to enumerate all possible exculpatory circumstances beforehand as part of the antecedent is seen by legal theorists as *proper defeasibility* (in German: echte defeasibility; cf. Wang, 2004; see also Dewitz et al., 1994). Interestingly, the division between proper and improper defeasibility is not mentioned in the psychological literature on human reasoning. This is probably so because it only makes sense from a theoretical perspective. The conditionals with which we reason in our daily lives are not written in books which prescribe what to conclude from them. It is therefore difficult to claim that all possible defeaters can be included as additional antecedents. Moreover, from a cognitive point of view, the limited capacities of human working memory, and the ease with which people make MP inferences, also make this an

implausible assumption. This incapability of enumerating all relevant defeaters of an action with respect to a given consequence is related to the *frame problem* in philosophy and cognitive science (see Dennett 1987, Fodor 1987). To put it broadly, the problem is when to stop thinking about the things that may be relevant for a particular action (e.g., Fodor, 1987)[2]. Along these lines, because this thesis aims to investigate laypeople's legal reasoning (and not the theoretical defeasibility of law), I will not use any further the distinction between proper and improper defeasibility. The thesis thus follows the psychological and philosophical definition of defeasibility, as the withdrawal of conclusions in light of additional information, irrespective of whether this additional information is exhaustively regulated in written form.

Finally, legal defeasibility is not restricted to the application of rules the penal code. Some researchers argue that the whole legal system is based on defeasibility, in the sense that prosecutors and attorneys have to defeat each other's arguments during trials (Godden & Walton, 2008; Prakken, 2001; see also Prakken & Sartor, 2004). Another illustration of defeasible reasoning in law is weighing the evidence or the relevant matter of facts. For example, imagine it is known that people do not normally kill the people they love. So, if Anna's beloved husband Fred is found dead, then it will be concluded that Anna did not kill Fred. However, if new evidence shows that Anna's clothes were soaked in Fred's blood, then the previous conclusion that Anna did not kill Fred is defeated. But if further evidence shows that Linda put the blood on Anna's clothes, then the conclusion is once again defeated and it is concluded anew that Anna did not kill Fred (example taken from Sartor, 2009). Pollock (1987) calls the defeaters which bring reasoners to not believe anymore that *q* follows from *p* *rebutters* (e.g., such as discovering that there is blood on Anna's clothes). Defeaters that question the believability of *p* are called by Pollock *undercutters* (e.g., such as discovering that Linda put the blood on Anna's clothes).

[2] The frame problem originates in artificial intelligence (see McCarthy & Hayes, 1969) as a more narrow and technical issue concerning the challenge of representing the non-effects of actions (i.e., what remains unchanged by particular actions), but was expanded by philosophers (e.g., Fodor, 1987) to the problem of relevance (Murray, 2016).

1.2.2. Legal Rules and Justice in Social Psychology

In the previous section, legal reasoning has been explained from the theoretical perspective of law. But how do people reason about legal rules in real life? Lawyers will probably follow the rules from the penal code. But the question remains how people reason without knowledge of these norms and regulations. To understand how laypeople reason with legal rules, I start by delineating the literature in social psychology and social justice.

Research on social justice has shown that when laypeople are asked to decide about an offender's sentence, they base their decisions on their *own sense of justice*, which is often described as being retributive, following "just deserts" principles (Carlsmith & Darley, 2008; Carlsmith, Darley, & Robinson, 2002; Darley, Carlsmith, & Robinson, 2000; Keller, Oswald, Stucki, & Gollwitzer, 2010). This means that people feel that offenders should be punished according to the severity of an offence. But what does severity mean? When laypeople are confronted with offences, they often react with intuitive feelings of moral outrage and a "strong and immediate" desire to punish the offender (Darley, 2009, p. 2; Darley & Pittman, 2003; Carlsmith & Darley, 2008; Tetlock, Kristel, Elson, Green, & Lerner, 2000; see also Fehr & Gächter, 2002). Laypeople's severity perception thus depends primarily on the moral wrongfulness of an offence. The higher the feelings of moral outrage, the higher the desire to punish the offender, and also the higher the desired sentence for the offence (Alter, Kernochan, & Darley, 2007; Carlsmith et al., 2002; Darley, 2009; Darley et al., 2000; Gromet & Darley, 2009; see also Buckholtz et al., 2008; Young, Cushman, Hauser, & Saxe, 2007). Accordingly, also Haidt (2001; 2007) argued that moral judgments are highly intuitive and driven by emotions. Aspects such as utilitarian harmfulness (i.e., how severe the consequences of an offence are) are secondary (Darley, 2009). For instance, in one study Alter and colleagues (2007) created offences where moral wrongfulness and harmfulness were disentangled, such as an offender mistakenly spending 1000$ from another person's credit card compared with an offender knowingly trying to spend 1000$ from another person's credit card but being stopped on time. Alter et al. asked participants to indicate how long the sentence should be. They found that it was moral wrongfulness, and not harmfulness, that predicted the desired sentence severity. Principles of deterrence (i.e., punishing in order to prevent future offences) are also secondary. Carlsmith and colleagues (2002) showed that although people often report to think that deterrence is a good reason for punishing, their sentencing behavior is actually best described by just deserts principles.

Moral outrage is often described as "wide-spread, if not universal" (Tyler & Boeckmann, 1997, p. 237). Yet, there are situations when people's primary reaction towards offences is less emotionally driven. Some researchers talk in this respect about dual process theories of moral reasoning. For instance, Greene, Sommerville, Nystrom, Darley, and Cohen (2001) showed that there are situations when people make moral decisions based on more deliberative and utilitarian principles. In their experiments Greene and colleagues presented their participants with moral dilemmas: situations where people have to decide if it is acceptable to sacrifice the life of one person in order to save the life of more. They found that when people are confronted with problems that must be resolved by directly inflicting harm to another person (e.g., throwing a person from a bridge in order to prevent a train killing five persons working on the rails), participants seem to base their moral evaluation primarily on autonomous emotional responses rather than on deliberate decision-making processes. Consequently, people usually take longer in making judgments that go against their intuitive emotions. In contrast, in situations of impersonal dilemmas that do not evoke such strong emotional responses, people primarily make decisions based on more rational utilitarian principles (e.g., pulling a lever to change the direction of a train, so that only one, instead of five persons working on the rails are killed; Greene et al., 2001). Similarly, Darley also argues that if an offence is considered not severe or morally acceptable, then moral outrage and the desire for punishment are low or nonexistent (Robinson & Kurzban, 2007). This is why it is often proposed that moral outrage also has a cognitive component besides its emotional one (Darley & Pittman, 2003).

If reasoning about offences can be explained by dual process theories (see Darley, 2009), it should be possible to specify when people reason analytically and when they reason emotionally. One factor that may influence whether people reason about offences analytically or emotionally is legal training. Unlike laypeople, lawyers have learned how to decide about offences and should thus be less driven by emotions. Consequently, they can decide about offences in a more elaborate and analytical way. High need for cognition has also been related to people's ability to make moral judgments rationally, while controlling emotions (Haidt, 2001). Another factor might be the rules of behavior imposed by one's immediate social surrounding, e.g., religion. Religiosity is often related to forgiveness, altruism and prosocial behavior (e.g., McCullough & Worthington, 1999; Saroglou, Pichon, Trompette, Verschueren, & Dernelle, 2005; Hansen, Vandenberg, & Patterson, 1995; see also McCullough, 2001). Through forgiveness, religiosity could thus lessen feelings of moral outrage and increase deliberative reasoning; although religiosity is sometimes also related to

higher dogmatism (e.g. Juan & Haley, 1970; Kilpatrick, Sutker, & Sutker, 1980; Swindell & L'Abate, 1970; for a review see Ross, Francis, & Craig, 2005).

The cultural context may also influence the extent to which people follow feelings of moral outrage. For instance, if one considers the implications of moral outrage, then people experiencing feelings of moral outrage towards an offender could actually agree with taking justice into their own hands (e.g. Cook, 2006; Haas, de Keijser, and Bruinsma, 2012). After all, vigilantes often only follow the feelings of moral outrage they felt after hearing about some preceding offence. However, given that in our culture vigilantism is not accepted and we have a legal system in charge of punishing offenders and guarantying security, people usually do not take justice in their own hands – although in experiments people often sympathize with vigilantes (Cook, 2006; Haas et al., 2012), and even rate a murderer's death as fair, regardless of whether it was achieved by vigilantism or by a legal trial ending in the death penalty (Skitka & Houston, 2001).

1.2.3. Empirical Findings on Legal Reasoning

As far as I know, there are almost no psychological studies investigating how people reason with conditionals describing legal rules. That is, the judicial syllogism presented in Section 1.2.1 has only been debated theoretically in legal theory, but has not received empirical support yet. Similarly, there are only a few studies investigating empirically the defeasibility of legal rules. In this section I present the few studies that can be related to legal conditional reasoning. Then, in Chapter 2, I develop my hypotheses on how people reason with legal conditional rules.

The study more closely related to the investigation of legal reasoning was conducted by Manktelow, Fairley, Kilpatrick, and Over (2000). Manktelow and colleagues created conditionals describing two types of road traffic violations: drink-driving or speeding (e.g., If a car driver travels above 30 mph in a built-up area and is stopped by the police, then she or he is liable to a fine). These conditionals were paired with circumstance information which could be mitigating (e.g., doctor on call) or aggravating (e.g., late for a party). The participants task was to decide whether the driver must/ should/ ought to/ may /may not/ ought not/ should not/ or must not be fined. If participants thought that none of these answer options is correct, they could write down in their own words what follows (e.g., more than a

fine such as imprisonment). Manktelow and colleagues found that mitigating circumstances influenced conclusions much more strongly for speeding than for drink-driving. In fact, drink-driving was always perceived to deserve some punishment, whereas speeding was sometimes completely condoned in light of mitigating circumstances. In addition, Manktelow and colleagues found that the effect of mitigating circumstances depended on how strongly the rule was violated. For instance, mitigating circumstances had less effect when the driver drove 60mph than when he drove 35mph.

Bonnefon, Haigh, and Steward (2013) investigated how people react to undesired actions. Bonnefon et al. showed that when the antecedent of a conditional describes somebody doing something bad (e.g., insulting or hurting someone) people expect that the consequent will describe something negative happening to this person (e.g., "If Brian insults Mandy, then he will get told off"). Similarly, if the antecedent describes somebody doing something good, then people expect that the consequent will describe something positive happening to this person (e.g., "If John helps Sarah, then he will be rewarded"). Bonnefon et al. talk in this respect about a *justice template* guiding people's interpretation of conditionals describing people doing good or bad things. These justice templates even help to assign utilities to conditionals with nonsense words. For instance, when people are confronted with a conditional like "If Lisa murbs Ian, then she will be tymped" and are told that Lisa does not like to being *tymed*, then they conclude that *murbing* is something bad.

Manktelow's et al. (2000) and Bonnefon's et al. (2013) studies show that people favor conclusions which punish undesired behavior, as for example violating traffic norms or acting disrespectful with other people. This high endorsement of conditionals describing punished or rewarded behaviors can also be found in the literature on inducements and advice, which is usually more concerned with how people reason deductively with the rules of classical logic. They are nonetheless related to legal reasoning because legal rules are seen as some kind of inducement, aiming to regularize the behaviors of a society. Newstead, Ellis, Evans, and Dennis (1997) constructed conditionals describing inducements like threats (e.g., "If you lose your next game then I shall take you out of the team) or promises (e.g., "If you wash the car then I'll let you borrow it tonight"), as well as advices such as tips (e.g., "If you stand by the pillar you'll be served immediately") and warnings (e.g., "If you pull the dog's tail then he'll bite you"). They found that people endorse more MP, MT, AC, and DA inferences when the conditional describes inducements than advices. Follow up studies showed that this preference for endorsing inducements more than advices is primarily because speakers of inducements are perceived to have more control than the ones of advices (Evans & Twyman-

Musgrove, 1998) and because the antecedent of inducements is perceived more sufficient and necessary for the consequent (Ohm & Thompson, 2004). Given the findings on inducements and advice, legal rules should also be perceived as being uttered by a high control source and consequently, their antecedents should be perceived of being sufficient and necessary for their consequent.

In the literature on conditional reasoning one can also find studies on deontic reasoning where people must decide in the so called Wason's (1968) selection task if a rule (e.g., If a person is drinking beer, then that person must be over 18 years of age) has been violated (e.g., Cosmides, 1989; Cox & Griggs, 1982; Holyoak & Cheng, 1995a; 1995b). At first sight it is tempting to relate those studies to my current aim of investigating defeasible reasoning with legal conditionals. However, such comparisons are difficult. First, the interpretation of deontic rules in those studies differs from the meaning of legal rules in law theory. Second, the experimental paradigm used in those studies differs in very important aspects from the conditional inference task, and thus also from the judicial syllogism presented in Section 1.2.1. I discuss further in the General Discussion why it may be problematic to relate the classical literature on deontic reasoning to legal reasoning (Section 7.1.4).

Chapter 2: Combining Conditional and Legal Reasoning - Aims and Hypotheses

In Chapter 1, conditional and legal reasoning have been dealt with separately. It has been shown that conditional reasoning is defeasible, and that people introduce their background knowledge and preferences into reasoning tasks. Relatedly, I emphasized that although people usually wish offenders to be punished, the penal code itself is defeasible: when exculpatory evidence becomes available, people should not be punished despite having committed an offence. But are legal rules defeasible for laypeople, too? The empirical evidence is still scarce. To be able to answer this question it is necessary to combine findings from conditional and legal reasoning. This is the goal of this Chapter.

Predicting how lawyers or judges reason with legal rules is rather easy. As explained in Section 1.2.1, reasoning from legal rules can be illustrated by drawing MP inferences. Therefore, if lawyers are confronted with a case where somebody killed another human, they should answer the following question with "yes":

If a person kills another human, then the person should be punished for manslaughter.
A person killed another human.

Should the person be punished for manslaughter?

Even when exculpatory evidence is presented as an additional premise, it is rather clear what lawyers should conclude. Given that lawyers know the regulations of the penal code, they know which circumstances count as exculpatory evidence and which do not. Consequently, if exculpatory evidence is introduced as an additional premise, lawyers can retrieve information from their knowledge of the penal code and conclude that the offender should not be punished. Lawyers should therefore answer the following problem with "no":

If a person kills another human, then the person should be punished for manslaughter.
A person killed another human.
Because of a psychological disorder it was impossible for this person to control and see the wrongfulness of his/her acts.

Should the person be punished for manslaughter?

In this way, lawyers' defeasible reasoning depends only on the availability of one defeater. As soon as they realize that the premises include exculpatory evidence, they can withdraw their previously drawn conclusion and conclude that the offender should not be punished. But how do people without legal background knowledge reason from legal conditionals?

Chan and Chua (1994) argued that one can only consider and weigh defeaters correctly if one has prior knowledge about the domain in question. Accordingly, several studies found differences in inferences depending on whether a participant has domain knowledge or not (see e.g., Cummins, 1995; Markovits, 1986). For instance, Cummins (1995) compared how people reason with conditionals with everyday content (e.g., "If it rains, then the streets will be wet"), and conditionals with fictitious content (e.g., "If it thardrons, then the streets will be sticky"). She found that participants rejected valid inferences more often for everyday than for fictitious conditionals, presumably because participants had background knowledge about defeaters for the former but not for the latter. Also Markovits (1986) found that familiarity with the content of conditionals increases the consideration of information not directly presented in the inference task. He found that the amount of correctly rejected DA and AC inferences and therefore the consideration of alternative causes were positively correlated with the domain familiarity. Moreover, Chan and Chua (1994) tested how undergraduates and policemen weight potential defeaters to everyday conditionals (e.g., "If Steven is invited, then he would attend the dance party") or conditionals describing police scenarios (e.g., "If he wants to steal car-radios, then the thief will break into the car"). They created different kinds of potential defeaters which differed in their relevance for the situations described in the conditionals ("Steven knows the host well" or "Parking attendants are on duty" vs. "Completes the report tonight" or "Carpark is dimly lit"). They found that while undergraduates and policemen did not differ in the weights they attached to the potential defeaters for everyday conditionals, they did so for conditionals describing police scenarios. Only policemen weighted the defeaters as they are actually weighed in real police situations.

On the basis of these studies, one can expect that laypeople and lawyers differ in legal reasoning. Cognitive psychologists, however, also want to understand *how* laypeople differ from lawyers. More precisely, cognitive psychologists are interested in how laypeople actually cognitively process potential exculpatory evidence. One option would be to simply ignore potential exculpatory evidence: because laypeople do not know that given information is exculpatory, they simply do not use it and reason monotonically. Alternatively, it could be possible that people without domain knowledge use other knowledge bases to compensate, consciously or unconsciously, their lack of knowledge. Chan and Chua (1994) proposed that

while experts in a field have elaborated schemas about the domain in question, people without domain knowledge only have "simple and ill-defined" schemas, with loosely related elements (p. 234). Yet, the authors did not specify the meaning of "ill-defined schemas" into detail.

My thesis expands Chan and Chua's assumption by assuming that the ill-defined schemas laypeople use are actually imprinted by criteria from different, but related domains. Which criteria these are, certainly varies from one domain to another. For instance, if a conditional is about police scenarios, as in Chan and Chua (1994), people might use their knowledge about police investigations from TV shows (see Schweitzer & Saks, 2007). If a conditional is about machines or engineering they might use their knowledge about lay physics or their experiences in fixing household appliance. Or, if a conditional is about psychological constructs, they might apply folk psychology concepts or guide their answers on their own personality traits (Bonnefon, 2010). But which is the most closely related domain laypeople have access to when reasoning with legal conditionals? According to the literature on social justice, it is people's own sense of justice.

As explained in Section 1.2.2, the research on social justice shows that laypeople's own sense of justice follows a just deserts principle, guided by the moral wrongfulness of offences instead of their harmfulness. Laypeople's severity perception of offences is thus guided by feelings of moral outrage, which are difficult to control and evoke a desire to punish the offender. Accordingly, I hypothesize that when laypeople reason about legal conditional rules, they should have difficulties in accepting possible exculpatory circumstances. More precisely, their consideration of exculpatory circumstances (and thus their willingness to defeat conclusions drawn from legal rules) should depend on how morally outraging the offence is. This hypothesis is supported by several studies showing how people's own preferences and subjective utilities influence their reasoning (see Section 1.1.2.5). As an illustration consider the following two problems:

If a person kills another human, then the person should be punished for manslaughter.
A person killed another human.
Because of a psychological disorder it was impossible for this person to control and see the wrongfulness of his/her acts.

Should the person be punished for manslaughter?

Vs.

If a person participates in an illegal game of chance, then the person should be punished for illegal gambling.
A person participated in an illegal game of chance.
Because of a psychological disorder it was impossible for this person to control and see the wrongfulness of his/her acts.

Should the person be punished for illegal gambling?

Although both problems include the very same exculpatory circumstance (which are according to law defeaters and thus reasons for voiding punishment), the offence described in the first legal conditional is clearly more morally outraging than the second. As a consequence, I expect laypeople to feel more reluctant to accept the exculpatory circumstance as a reason for voiding punishment. Laypeople should thus answer "no" and decide that the offender should not be punished more often in the second than in the first example. In other words, the more morally outrageous the initial offence is, the harder it should be for laypeople to accept circumstance information as exculpatory and the more they will deny this circumstance as being sufficiently strong to refrain from punishing the offender. This higher difficulty should be reflected in less no-punishment conclusions and longer decision times whenever they decide contrary to their feelings of moral outrage. These hypotheses are investigated in Chapter 3, in Experiments 1-3.

The assumption that people use their own sense of justice to decide about the defeasibility of legal rules is supported by the empirical findings on legal reasoning presented in Section 1.2.3. If one assumes that people use their own sense of justice when reasoning about offences, then it is understandable why in Bonnefon's et al. (2013) study reasoners concluded that antecedents describing someone doing something undesirable should be followed by consequents describing something negative happening to this person. Along the same lines, people's own sense of justice and feelings of moral outrage also explain why in Manktelow's et al. (2000) study reasoners were less influenced by mitigating circumstances in cases of drink-driving than in speeding. Although the authors did not relate their findings to moral outrage, drink-driving is probably more morally outraging than speeding, making reasoners more reluctant to forgive drink-driving. In this way one could say that the reasoners' own preferences about what should happen to offenders will influence their conclusions in a legal conditional reasoning task. Also the findings on inducements and

advices strengthen the assumption that people should have difficulties to consider defeaters in legal reasoning. Given that legal rules are inducements, the antecedents of such conditionals are perceived to have a high necessity to its consequent (see Ohm & Thompson, 2004).

One direct consequence of using one's own sense of justice for legal reasoning is that differences in one's sense of justice reflect in differences in legal reasoning. In Section 1.2.2 I presented the literature showing that people's reactions towards offenders sometimes differ. Sometimes the desire for punishment results in acts of vigilantism, and sometimes one's cultural norms of behavior, such as religion, dictate how we should deal with offenders. This thesis therefore also aims to investigate how such behavioral norms affect laypeople's defeasible reasoning with legal conditionals. The effect of behavioral norms on legal reasoning is investigated in Chapter 4. For instance, Experiment 4 investigates how religiosity affects legal reasoning. There is a wide range of studies relating religiosity to either forgiveness or dogmatism. Regardless of which is true, religiosity should affect people's defeasible reasoning and their consideration of exculpatory circumstances as reasons for voiding punishment. Similarly, Experiment 5 investigates cross-cultural differences in legal reasoning. It is investigated under which circumstances additional information describing offenders as vigilantes is considered exculpatory and thus a reason to withdraw conclusions.

Besides investigating how one's own sense of justice affects legal reasoning, another aim of this thesis is to explore how to moderate this effect. One way to do so is to vary the participants' instructions. By asking participants to reason according to their own sense of justice or to try to reason like a real judge, it should be possible to affect the degree to which one's own sense of justice affects legal reasoning. This is tested in Experiment 2. However, another more subtle influence on the degree to which people's conclusions are based on their own sense of justice is the phrasing of the legal conditional. The effect of phrasing in legal reasoning is investigated in Chapter 5. According to legal theory, legal conditionals are deontic, describing what *should* happen to an offender. Already in Sections 1.1.1.2.2 and 1.1.2.5 it has been shown how deontic conditionals often trigger people's enhanced reliance on their own preferences and subjective utilities. Consequently, one way to moderate the effect of one's sense of justice on legal reasoning is to phrase legal conditionals non-deontically. They can be phrased as factual conditionals by changing the modal auxiliary and describing what *will* happen if an offence is committed (e.g., If a person kills another human, then the person will be punished for manslaughter). As shown in Section 1.1.1.2.2, the modal auxiliary used in conditionals affects the criteria used to draw conclusions. In this way, by being factual, participants' tendency of guiding their conclusions on their own preferences

should diminish and their consideration of factual information, such as exculpatory evidence, should increase. The role of the modal auxiliary (should vs. will) is investigated in Experiments 6 and 7. Another way in which phrasing can influence the effect of one's own sense of justice on legal reasoning is by keeping the modal auxiliary deontic, but changing how general the rule is phrased. If a conditional is phrased in such a way that it already suggests the existence of exculpatory circumstances, for example by introducing uncertainty or weakening the relationship between antecedent (the offence) and the consequent (the punishment), then people should be less prone to conclude that an offender should be punished even though an offence was committed. This can be done by modifying the amount of cases in which *q* follows from *p*, for instance by stating that *all* people who commit an offence are punished or that only *some* people are punished. This is tested in Experiment 8.

Finally, one last aim of this thesis is to test whether people's reluctance to withdraw from negatively-charged conclusions may also occur in everyday situations. More precisely, this thesis' main assumption – that laypeople have problems in accepting exculpatory evidence for morally outrageous offences – can be interpreted as a difficulty in withdrawing from negative emotions. In other words, if something bad happens, it is difficult to let go of – or defeat – the negative emotions associated with this event. In Chapter 6, I go beyond the domain of legal reasoning to test if also in everyday situation people have difficulties in defeating conclusions resulting from negative emotions. In Experiments 9 and 10 I will therefore construct conditionals containing negative events (e.g., If my mother dies, then I am sad) but also positive ones (e.g., If I get my dream job, then I am happy) and test in how far conclusions from those conditionals can be defeated by subsequent information.

To sum up, the main hypotheses of this thesis are:

1) Laypeople's decisions about exculpatory circumstances in legal conditional reasoning depend on their own sense of justice and thus on how morally outraging the offence is: High morally outraging offences should result in fewer acceptances of exculpatory circumstances, and thus less defeasible reasoning.

2) Differences in people's preferences about how to deal with offenders and offences should affect the conclusions laypeople draw from legal conditionals.

3) Linguistic factors that minimize the importance of one's own sense of justice or that highlight the existence of exculpatory circumstances should reduce the effect of one's own sense of justice on legal conditional reasoning.

4) People's reluctance to withdraw conclusions from conditionals describing negatively laden events – such as defeating the conclusion that an offender should be punished – also occurs in everyday situations.

Chapter 3: Legal Rules, Moral Outrage and Domain Knowledge

The aim of the first three experiments was to test the first main hypothesis of this thesis. It is tested whether laypeople's consideration of exculpatory circumstances depends on moral outrage by comparing their conclusions with those of lawyers, who know the regulations of the penal code and should therefore answer accordingly.

In all three experiments, laypeople were students from disciplines outside law, with no specific law knowledge. The group of lawyers consisted of graduate lawyers (who completed at least the first German state examination) and advanced law students. In the latter group, it was ensured that all law students were already familiar with the rules of the penal code used in the experiments; this is usually the case after the first three to four semesters of law studies in Germany. Strictly speaking these law students are not yet fully-trained lawyers; nonetheless they are already familiar with the relevant legal rules and certainly have more legal experience than laypeople. Hence, for simplicity reasons, henceforth the group of people with law studies are referred to as 'lawyers'.

3.1. Experiment 1: The Legal Conditional Reasoning Paradigm

In Experiment 1, I developed the experimental paradigm of this thesis. I created legal conditionals by selecting legal rules from the German penal code and by putting those in conditional form. In Experiment 1, only severe offences were selected. These legal conditionals were combined with circumstance information that could be exculpatory, neutral or aggravating. It was tested whether a) people defeat logically valid conclusions in light of such circumstances, and b) this task is appropriate to measure differences between lawyers and laypeople. The hypothesis is that when faced with exculpatory circumstances, lawyers will withdraw the conclusion to punish the offender. Laypeople may also do so to some extent, but since only severe offences were presented – and severe offences are related to high moral outrage (Darley et al., 2000) – withdrawing the conclusion to punish the offender

should be more difficult and thus less pronounced for laypeople. As a consequence, laypeople should take longer in concluding not to punish the offender than to punish him. I also added aggravating circumstances to the experimental design to check if, besides defeating conclusions, circumstance information also enhances punishment conclusions by strengthening the association between the antecedent (the offence) and the consequent (the punishment) of the legal conditional (cf. Manktelow & Fairley, 2000; Stevenson & Over, 1995).

Hypothesis 1: While lawyers withdraw the conclusion to punish the offender in light of exculpatory circumstances, laypeople should have difficulties in doing so and thus withdraw the conclusion to punish the offender less often than lawyers.

Hypothesis 2: Laypeople's difficulty in considering exculpatory circumstances should be reflected in decision times: no-punishment conclusions should take longer than punishment conclusions.

Hypothesis 3: Circumstance information that strengthens the link between antecedent (offence) and consequent (punishment) should lead to more punishment conclusions compared to instances where this link is not strengthened.

3.1.1. Methods

3.1.1.1. Participants

Participants were 22 lawyers (16 female) and 26 laypeople (14 female). The mean age of lawyers was 26.5 (*SD* = 6.7); the mean age of laypeople was 23.3 years (*SD* = 2.3). Within the lawyers' group, 8 already graduated from law school, the rest were still at university but already had knowledge about the offences presented in the experiment[3]. Law students had studied for 4.6 semesters on average.

[3] One participant reported no knowledge of the legal text on arson. However, as this participant did report knowledge about exculpatory circumstances, the participant's answers were not excluded.

3.1.1.2. Materials and Design

The problems followed the structure of a defeasible MP inference, but were adapted to the legal context. They consisted of (1) a legal conditional "if *p* then *q*", where *p* refers to an offence (*manslaughter*, *arson*, *bodily injury*, or *theft*) and *q* to a punishment, (2) the fact *p* stating that someone committed the offence, (3) additional information about circumstances, and (4) the conclusion phrased as a question about *q*, that is, whether the offender should be punished for the offence or not (yes vs. no). For each problem, participants were also asked to indicate on a 7-point Likert scale (5) how certain they were about their conclusion (1 = *very unsure*, 7 = *very sure*) and (6) how severe they perceived the offender's action to be (1 = *not severe at all*, 7 = *very severe*). The latter question was meant to show whether participants incorporated the information about circumstances into their mental representation of the offence. An example problem is as follows:

Legal conditional rule:	If a person kills another human, then the person should be punished for manslaughter.
Fact:	Bob killed another human.
Circumstances:	Bob is schizophrenic and had a delusion of an attack against him.
Conclusion:	Should Bob be punished for manslaughter?
Certainty:	How certain are you?
Severity:	How severe do you perceive Bob's action to be?

In total, there were 48 conditional reasoning problems. The experimental manipulation was that all problems were presented either with situations which were potentially exculpatory, aggravating, or neutral (i.e., crime-irrelevant) for the given legal rule. Among the exculpatory circumstances, half were legally relevant (i.e., potentially exculpatory for the offence according to penal code, or at least permissible as such at a judge's discretion), and half were legally irrelevant (i.e., probably exculpatory, or at least mitigating, according to some personal standards, but not according to law). The same distinction of legally relevant or irrelevant information was made for the aggravating circumstances. Neutral circumstances were, per definition, always legally and morally irrelevant. Thus, the problems with neutral circumstances represented the base acceptance rate of the conditional legal rule. Examples of the circumstances are presented in Table 4. All exculpatory, aggravating, and neutral

circumstances were selected from a larger pool of problems (N = 192) that were tested in a pilot study. In this pilot study, participants (n = 16 for theft and manslaughter; n = 17 for bodily injury and arson) rated how mitigating or aggravating they perceived a particular circumstance for a given offence. Besides exculpatory circumstances, mitigating circumstances were also used in this pilot study. For the main experiment, I selected those combinations of offences and circumstances that received the highest mean "aggravating" and "mitigating" ratings in the pilot study ("neutral" circumstances were those which obtained the mean value in the scale). In the actual experiment all circumstance descriptions were of similar length (61 ± 2 characters including spaces). Moreover, the name and the gender of the offender was varied between subjects to avoid possible effects of attitudes or preferences (Sporer & Goodman-Delahunty, 2009), but not as another independent variable. Overall, the experiment followed a 2 (group: laypeople vs. lawyers) × 2 (relevance: legally relevant vs. legally irrelevant) × 3 (circumstances: exculpatory vs. aggravating vs. neutral) mixed design. The factor expertise was a between-subjects factor; all other factors were within-subjects factors. I did not differentiate between the different offences in the 48 problems (manslaughter, arson, bodily injury or theft), so they were not treated as an additional factor.

In addition, I also measured participants' legal attitude using a German translation of the revised version of the Legal Attitudes Questionnaire (RLAQ; Kravitz, Cutler, & Brock, 1993) administered immediately after the experiment. Because RLAQ scores did not differ among participants, these results are not reported.

Table 4

Examples for legally relevant and legally irrelevant exculpatory, neutral, and aggravating circumstances used in Experiment 1 (original material was in German language)

	Exculpatory	Neutral	Aggravating
Legally relevant	The victim was the first to stab Bob during the fight.	Bob loves traveling by train while listening to music	Bob wanted to get the money from the victim's live insurance.
Legally not relevant	The victim raped Bob's wife several times when she was a child.	Bob drank a glass of water a few minutes ago	The victim was a single parent of a child.

3.1.1.3. Procedure

The experiment was programmed in Cedrus® SuperLab® 4.X and conducted on a computer. Participants were tested individually. The experiment was introduced as an experiment about reasoning in law. Participants were told they will be confronted with legal cases in which a person committed an offence, and their task is to decide as a judge whether the person should be punished for the offence. The legal conditional was introduced as a general legal rule. Problem components (i.e., rule, fact, circumstance, and conclusion) were each presented on separate screens. Participants could switch from one screen to the next by pressing the space bar and gave their conclusions by pressing a "y" (yes) or "n" (no) key. The number pad was used to provide ratings for the last two questions on certainty and severity. All statements were presented in black font except the conclusion question, which was presented in red. Participants had the opportunity to take a break between problems. At the beginning of the experiment, participants completed six practice problems. For both practice and experimental trials, the order of problems was randomized. The experiment took about 45 minutes. All participants received monetary compensation for their participation.

3.1.2. Results

General note: In this and in all the following experiments in this thesis, data were analyzed with analyses of variances (ANOVAs), which will be described in more detail in the corresponding sections. In cases where Maulchy's sphericity test was significant Greenhouse Geisser corrected values were used. Significant effects in the ANOVAs were scrutinized with follow-up ANOVAs or t-tests where appropriate. P values in these pairwise comparisons were tested against Bonferroni adjusted alpha levels. In all experiments, standardized mean differences (*d*) were computed as described by Borenstein (2009). For decision times I always computed the time between presentation of the potential exculpatory evidence and participants' conclusions, excluding times resulting from mistyped/invalid answers. Besides for Experiment 5 (see Section 4.2), all problems were presented in German language.

3.1.2.1. Perceived Severity Ratings (Manipulation Check)

Severity ratings were analyzed with a 2 (group: laypeople vs. lawyers) × 2 (relevance: legally relevant vs. legally irrelevant) × 3 (circumstances: exculpatory vs. aggravating vs. neutral) mixed ANOVA. Descriptive data can be found in Table 5. The ANOVA revealed a main effect of circumstances, $F(1.36, 62.77) = 152.85$, $p < .001$, $\eta_p^2 = .769$, a main effect of relevance, $F(1, 46) = 49.77$, $p < .001$, $\eta_p^2 = .520$, an interaction between circumstances and relevance, $F(2, 92) = 118.55$, $p < .001$, $\eta_p^2 = .720$, and a three way interaction between all factors, $F(2, 92) = 4.43$, $p = .015$, $\eta_p^2 = .088$. While the effect of circumstances did not differ between lawyers and laypeople for problems with legally relevant circumstances, $F(1.41, 64.79) = 0.71$, $p = .447$, $\eta_p^2 = 0.015$, it did for problems with legally irrelevant circumstances, $F(1.47, 67.43) = 4.82$, $p = .011$, $\eta_p^2 = .108$. Laypeople were descriptively more influenced by irrelevant exculpatory and irrelevant aggravating circumstances than lawyers, although pairwise *t*-tests did not reach significance (*t*s < 1.52, *p*s > .137). Overall, problems with exculpatory circumstances ($M = 4.17$; $SD = 0.86$) were perceived as less severe than problems with neutral circumstances ($M = 5.43$; $SD = 0.83$), $t(47) = 9.81$, $p < .001$, $d = 1.49$, and those as less severe than problems with aggravating circumstances ($M = 6.05$; $SD = 0.65$), $t(47) = 10.16$, $p < .001$, $d = 0.75$. All other effects were not significant (*F*s < 2.48, *p*s > .110). Because different circumstances influenced the perceived severity of offences in the expected direction, it can be concluded that the manipulation was effective.

Table 5

Mean severity ratings (and standard deviations) for problems with legally relevant and legally irrelevant circumstances in Experiment 1

	Circumstance information		
	exculpatory	neutral	aggravating
Legally relevant circumstances			
Laypeople	3.56 (0.99)	5.47 (0.91)	6.14 (0.65)
Lawyers	3.74 (0.89)	5.36 (0.84)	6.22 (0.63)
Legally irrelevant circumstances			
Laypeople	4.53 (0.99)	5.55 (0.85)	6.04 (0.76)
Lawyers	4.90 (0.64)	5.33 (0.82)	5.76 (0.71)

Note. Severity ratings range from 1 (not severe at all) to 7 (very severe).

3.1.2.2. Defeated Conclusions (no-punishment)

For the analysis of the amount of defeated conclusions, I calculated the percentage of no-punishment conclusions for each of the problem categories (see Table 6). As no-punishment decisions were scarce for neutral and aggravating circumstances, only a 2 (group: laypeople vs. lawyers) × 2 (relevance: legally relevant vs. legally irrelevant) ANOVA on no-punishment conclusions for exculpatory circumstances was conducted. The ANOVA showed a main effect of relevance, $F(1, 46) = 224.05$, $p < .001$, $\eta_p^2 = .83$, and an interaction between group and relevance, $F(1, 46) = 11.59$, $p = .001$, $\eta_p^2 = .201$. In cases of legally irrelevant exculpatory circumstances, lawyers and laypeople showed no difference in percentage of no-punishment conclusions, $t(32.10) = 0.47$, $p = .644$, $d = 0.13$, with participants in both groups almost never considering legally irrelevant exculpatory circumstances as valid defeaters. In cases of legally relevant circumstances, however, lawyers and laypeople differed: lawyers accepted legally relevant exculpatory circumstances as valid defeaters much more often than laypeople (51.14% vs. 35.58%), $t(46) = 2.39$, $p = .021$, $d = 0.69$. No main effect of group was found, $F(1, 46) = 2.14$, $p = .151$, $\eta_p^2 = .044$.

Table 6

Percentages (and standard deviations) for the no-punishment conclusions in Experiment 1

	Circumstance information		
	exculpatory	neutral	aggravating
Legally relevant circumstances			
Laypeople	35.58 (25.17)	1.44 (4.07)	0.00 (0.00)
Lawyers	51.14 (18.86)	0.57 (2.67)	3.41 (8.78)
Legally irrelevant circumstances			
Laypeople	6.25 (17.41)	0.96 (3.40)	0.48 (2.45)
Lawyers	4.55 (6.15)	1.70 (5.84)	1.14 (3.68)

3.1.2.3. Decision Times and Certainty Ratings

Decision times and certainty ratings were analyzed separately for punishment and no-punishment conclusions (Figure 1). As the majority of no-punishment conclusions were made

in light of legally relevant exculpatory circumstances, only these problems were analyzed. For both analyses, a 2 (decision: punishment vs. no-punishment) × 2 (group: laypeople vs. lawyers) mixed ANOVA was conducted.

For the analyses of decision times the ANOVA revealed a significant interaction between group and punishment decision, $F(1, 42) = 4.36$, $p = .043$, $\eta_p^2 = .094$. Whereas lawyers showed no differences in their decision times for punishment and no-punishment conclusions, $t(21) = 0.94$, $p = .358$, $d = 0.21$, laypeople required significantly more time to select no-punishment than to select punishment conclusion, $t(21) = 2.80$, $p = .011$, $d = 0.62$. However, a main effect of group also indicated that, in general, lawyers (M = 10778ms; SD = 3496) had longer decision times than laypeople (M = 8236ms; SD = 3496), $F(1, 42) = 5.82$, $p = .02$, $\eta_p^2 = .122$. No main effect of decision was found, $F(1, 42) = 0.16$, $p = .689$, $\eta_p^2 = .004$.

The analyses of certainty ratings showed a similar pattern. Although the interaction between group and punishment decision failed to reach significance, $F(1, 41) = 2.70$, $p = .108$, $\eta_p^2 = .062$, descriptively only laypeople were less certain about no-punishment than about punishment decisions. Additional main effects revealed that lawyers were more certain than laypeople, $F(1, 41) = 5.67$, $p = .022$, $\eta_p^2 = .122$, and all participants were more certain about punishment than no-punishment conclusions, $F(1, 41) = 6.29$, $p = .016$, $\eta_p^2 = .133$.

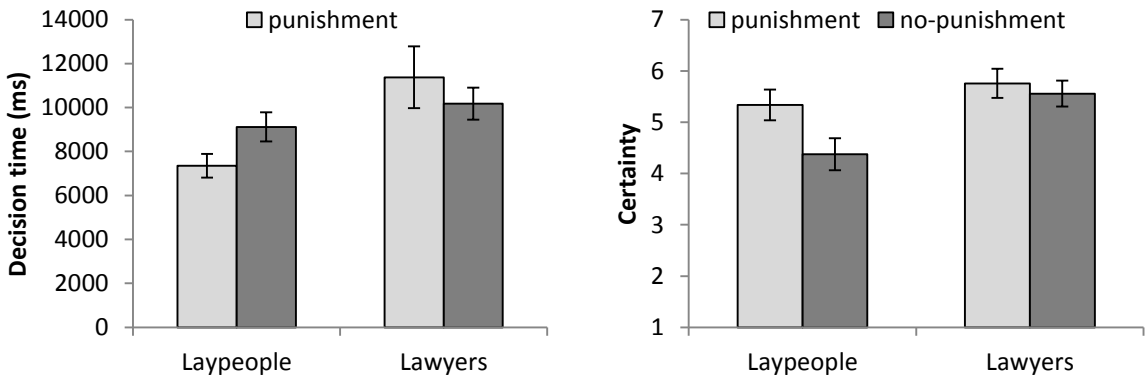

Figure 1. Decision times and certainty ratings for punishment and no-punishment conclusions in Experiment 1. Error bars show standard errors.

3.1.3. Discussion

The data shows that people indeed use additional information about the circumstances of an offence when reasoning with legal conditionals. However, in accordance with Hypothesis 1, lawyers decided not to punish offenders more often than laypeople. Since severe offences were used, it can be assumed that the moral outrage laypeople felt influenced their consideration of exculpatory circumstances. Accordingly, lawyers were quite certain of all their decisions, but laypeople felt more insecure and needed more time when deciding contrary to what moral outrage predicts (Hypothesis 2). Lawyers needed more time to make a decision than laypeople, without showing differences in the time needed to arrive at a punishment or no-punishment conclusion. This finding suggests that the underlying cognitive processes of lawyers were more deliberate (cf. Evans, 2008). Yet, lawyers did not consider all the legally relevant exculpatory circumstances that were presented, probably because for some of those presented in the experiment it lays within the judges' discretion whether they refrain from punishing or whether they only consider them as mitigating circumstances.

In addition to exculpatory circumstances I also added aggravating circumstances into the experimental paradigm to test whether such information enhances the logically valid answer of punishing the offender (Hypothesis 3; cf. Manktelow & Fairley, 2000; Stevenson & Over, 1995). However, this was not the case, probably because of a ceiling effect. Laypeople's severity perception of offences with neutral circumstances was already pretty high (M = 5.51; SD = 0.86). Accordingly, laypeople's rate of no-punishment conclusions for offences with neutral circumstances was already low. So, it is possible that further aggravating information did not have an additional effect on participants' preference of punishment conclusions.

In summary, the results of this experiment support the hypothesis that laypeople but not lawyers have difficulties in accepting exculpatory circumstances as reasons for defeating the MP conclusion of punishing the offender. The results support the appropriateness of this paradigm to further investigate legal reasoning and to test whether laypeople's punishment conclusions depend on moral outrage.

3.2. Experiment 2: Legal Conditional Reasoning and Moral Outrage

In Experiment 1, it was shown that lawyers and laypeople differ in their acceptance of exculpatory circumstances as defeaters to conditional legal rules. However, I did not test systematically whether this difference depends on moral outrage. It could be possible that lawyers and laypeople differed only because laypeople always reject violations of norms per se, irrespective of how morally outrageous an offence is. To test the hypothesis of laypeople's punishment conclusions depending on moral outrage, it is necessary to pair offences of differing degrees of moral outrage with the same kinds of circumstance information. If laypeople's consideration of defeaters depends on how morally outrageous the offence is, then the difference between lawyers and laypeople in no-punishment conclusions found in Experiment 1 should diminish for low moral outrage offences, but remain for high moral outrage offences. Lower moral outrage towards an offence should make laypeople more willing to accept evidence as exculpatory, as they do not feel the strong desire to punish the offender. Overall, lawyers' decisions should not vary with the degree of moral outrage an offence might evoke, but only with what is prescribed by the penal code.

Also participants' perspective was varied by phrasing two different instructions: one condition asked them to act according to their own sense of justice, while the other asked to act as they think a real judge would do. If decisions about exculpatory circumstances are influenced by moral outrage evoked by the offences, its effect should be higher for participants in the former group. Also, laypeople should be more certain of their decisions when instructed to decide based on their own sense of justice.

Hypothesis 1: The more morally outraging an offence is, the higher a layperson's reluctance in accepting a given exculpatory circumstance as a reason for voiding punishment. Lawyers' conclusions should not depend on moral outrage.

Hypothesis 2: The more morally outraging an offence is, the more laypeople's conclusions will differ from lawyers'.

Hypothesis 3: The level of moral outrage evoked by an offence will only impact laypeople's decision times, but not lawyers'.

Hypothesis 4: Instructing laypeople to decide as a real judge should diminish the effect of moral outrage in legal conditional reasoning and thus make laypeople more uncertain about their decisions.

3.2.1. Methods

3.2.1.1. Participants

Participants were 24 lawyers (15 female) and 40 laypeople (20 female). Three participants in the lawyers' group were excluded from analysis because they failed to fulfill the inclusion criteria of having studied law for at least 4 semesters or having passed their intermediate law exam. Thus, the final sample of lawyers consisted of 21 participants (12 female). The mean age of lawyers was 26.48 years (*SD* = 4.06); the mean age of laypeople was 24.15 years (*SD* = 5.31). Six participants from the lawyers' group had already graduated, the rest were still at university, having studied for 9.6 semesters on average.

3.2.1.2. Materials and Design

The material consisted of 36 conditional problems that followed the structure of those in Experiment 1, but used legal conditionals that differed in the level of moral outrage evoked by the offences. *Maltreatment of wards* and *child sexual abuse* were considered high moral outrage offences, *handling stolen goods* and *breach of domestic peace* were considered medium moral outrage offences, and *illegal gambling* and *obtaining benefits by devious means* were considered low moral outrage offences. These different offences were selected from a large and representative (*N* = 448; 315 female) preliminary study in which participants rated on a scale from 1 to 7 the level of moral outrage felt in response to *N* = 36 offences from the German penal code. High moral outrage offences received a mean rating of *M* = 6.83 (*SD* = 0.48), medium moral outrage offences a mean rating of *M* = 3.91 (*SD* = 1.42), and low moral outrage offences a mean rating of *M* = 2.34 (*SD* = 1.24).

As additional information relevant exculpatory and irrelevant control circumstances were used. Relevant exculpatory circumstances were taken from the General Part of the German penal code and described scenarios of (1) *absence of criminal responsibility due to psychological disorders*, (2) *mistakes of law*, or (3) *necessity brought about by coercion*. Irrelevant control circumstances also pertained to psychological disorders, mistakes of law, and situations of coercion, but were completely irrelevant to the offence (e.g., psychological disorders with no legal connection to the crime, like having crime irrelevant memory problems in a case of maltreatment of a ward). These control circumstances were selected from a larger pool from three online studies (N = 21, N = 20, and N = 27) and were used to ensure that participants attended to the task and read all of the circumstance information to make a decision. The crucial manipulation of Experiment 2 was that each circumstance was paired with each legal conditional. This allowed assessing whether the same circumstances were weighted differently depending on the degree of moral outrage of the offence with which it is presented. Offenders described in the problems were always male. Examples of the circumstances can be found in Table 7.

There were two different sets of instructions specifying the perspective that participants should take during evaluation of the conclusion. All lawyers and half of the laypeople were instructed to imagine that they were a judge who always relies on prescriptions of the legal system ("legal system" instruction). The other half of the laypeople were instructed to imagine that they were a judge who makes decisions based on his or her own sense of justice irrespective of regulations of the legal system ("own sense of justice" instruction). The experiment used a 3 (moral outrage: high vs. medium vs. low) × 3 (group: laypeople - own sense of justice vs. laypeople - legal system vs. lawyers - legal system) × 2 (circumstances: exculpatory vs. control) mixed design. The subject condition was a between-subjects factor, and the other factors were within-subject factors.

Table 7

Example of problems used in Experiment 2 translated into English language. This problem describes relevant exculpatory vs. irrelevant control circumstances in a case concerning maltreatment of wards ("If a person maltreats a minor in their charge, then the person should be punished for maltreatment of wards").

	Moral outrage evoked by the offence		
	High	Medium	Low
Relevant exculpatory circumstances	Relevant case of: Psychological disorder *Because of a psychological disorder, it was impossible for Bob to control his action and realize that his acts were harmful.* Mistake of law *It was impossible for Bob to know that this was an illegal educational method.* Situation of necessity brought about by coercion *Under threats to his life and physical integrity, Bob was coerced by a third person into harming the child.*		
Irrelevant control circumstances	Irrelevant case of: Psychological disorder *Because of a psychological disorder, Bob could not remember what the minor likes or the gifts the minor gave him.* Mistake of law *It was impossible for Bob to know that children also have a sense of humor.* Situation of necessity brought about by coercion *Under threats to his life and physical integrity, Bob was coerced by a third person into not seeing any violent movies.*		

Note. Offences of high, medium and low moral outrage were always paired with all six kinds of circumstance information.

3.2.1.3. Procedure

The procedure was the same as in Experiment 1. Similarly, the experiment was introduced to participants as an experiment about reasoning in legal contexts. Participants were told that they would be confronted with legal cases, and their task would be to decide whether a legal conditional rule should be followed. However, in contrast to the previous experiment, the instruction focused more explicitly on the aspect of applying rules and mentioned that the application of a rule would lead to punishment of the offender. Different components of the problems were presented on sequential screens; participants could switch from one statement to the next by pressing the space bar. Decisions about applying the conditional rule and punishing the offender were indicated by pressing a "y" (yes) or "n" (no) key. Ratings about certainty and severity were provided using the number pad. The perspective to be taken by participants was given during the instructions and was highlighted in blue. After reading the instructions, it was made sure that participants understood the perspective to be taken by asking them to rephrase the instructions. Before the experiment, participants solved six practice problems. All problems were presented randomly. The experiment took about 30 minutes. All participants received either course points or a monetary compensation.

3.2.2. Results

3.2.2.1. Perceived Severity Ratings (Manipulation Check)

Perceived severity ratings (upper part of Table 8) were analyzed using a 3 (moral outrage: high vs. medium vs. low) × 3 (group: laypeople - own sense of justice vs. laypeople - legal system vs. lawyers - legal system) × 2 (circumstances: exculpatory vs. control) mixed ANOVA. The ANOVA revealed a main effect of moral outrage, $F(1.37, 79.57) = 323.64$, $p < .001$, $\eta_p^2 = .848$. High moral outrage offences ($M = 6.17$, $SD = 0.87$) were perceived as more severe than medium moral outrage offences ($M = 3.58$, $SD = 1.25$), $t(60) = 16.36$, $p < .001$, $d = 2.37$, and those as more severe than low moral outrage offences ($M = 2.86$, $SD = 1.26$), $t(60) = 8.39$, $p < .001$, $d = .0.57$. Also an interaction between moral outrage and group was found, $F(2.74, 79.57) = 5.34$, $p = .003$, $\eta_p^2 = .156$, however, pairwise *t*-tests did not reach the Bonferroni adjusted alpha of 0.0167. Additionally, also a main effect of circumstances, $F(1,$

58) = 72.30, $p < .001$, $\eta_p^2 = .555$, and an interaction between circumstances and moral outrage was found, $F(2, 116) = 3.99$, $p = .021$, $\eta_p^2 = .064$. Problems with exculpatory circumstances were perceived as less severe than problems with control circumstances, yet this was especially the case for medium moral outrage offences, $t(60) = 7.82$, $p < .001$, $d = 0.72$ (for high moral outrage: $t(60) = 6.82$, $p < .001$, $d = 0.70$; for low moral outrage: $t(60) = 5.77$, $p < .001$, $d = 0.51$). All other effects were not significant (Fs < 1.66, ps > .200).

3.2.2.2. Certainty Ratings

Certainty ratings (lower part of Table 8) were analyzed using a 3 (moral outrage: high vs. medium vs. low) × 3 (group: laypeople - own sense of justice vs. laypeople - legal system vs. lawyers - legal system) × 2 (circumstances: exculpatory vs. control) mixed ANOVA. The ANOVA showed main effects of group, $F(2, 58) = 25.08$, $p < .001$, $\eta_p^2 = .464$, of circumstances, $F(1, 58) = 69.80$, $p < .001$, $\eta_p^2 = .546$, and of moral outrage, $F(2, 116) = 8.36$, $p < .001$, $\eta_p^2 = .126$, and interactions between circumstances and moral outrage, $F(1.84, 106.41) = 16.30$, $p < .001$, $\eta_p^2 = .219$, moral outrage and group, $F(4, 116) = 2.94$, $p = .023$, $\eta_p^2 = .092$, and group and circumstances, $F(2, 58) = 4.99$, $p = .010$, $\eta_p^2 = .147$. Participants were more certain about their decisions in cases of control ($M = 6.20$, $SD = 0.69$) than in cases of exculpatory circumstances ($M = 5.38$, $SD = 0.89$), primarily in cases of high moral outrage, $t(60) = 10.60$, $p < .001$, $d = 1.53$, followed by medium, $t(60) = 4.64$, $p < .001$, $d = .0.67$, and low moral outrage, $t(60) = 3.53$, $p = .001$, $d = .51$. In cases of high moral outrage, certainty ratings did not differ between laypeople and lawyers, $F(2, 58) = 1.78$, $p = .117$, $\eta_p^2 = 0.058$. However, they did in cases of medium, $F(2, 58) = 8.24$, $p = .001$, $\eta_p^2 = 0.221$, and low moral outrage, $F(2, 58) = 11.596$, $p < .001$, $\eta_p^2 = .286$. Laypeople in the own sense of justice and in the legal system group were less certain than lawyers in cases of medium ($t(39) = 2.47$, $p = .018$, $d = 0.77$; $t(37.99) = 4.03$, $p < .001$, $d = 1.25$; respectively), and in cases of low moral outrage ($t(29.04) = 4.20$, $p < .001$, $d = 1.29$; $t(39) = 3.75$, $p = .001$, $d = 1.17$, respectively). In cases of exculpatory circumstances, laypeople in the own sense of justice group ($M = 5.31$, $SD = 0.59$) were more certain about their decisions than laypeople in the legal system group ($M = 4.65$, $SD = 0.8$), $t(34.94) = 2.95$, $p = .006$, $d = 0.93$ (for irrelevant circumstances $p > .370$). The three way interaction was not significant, $F(3.670, 106.41) = 0.61$, $p = .643$.

Table 8

Mean severity ratings and certainty ratings (and standard deviations) for problems with irrelevant control and relevant exculpatory circumstances in Experiment 2

	Moral outrage of the conditional		
	High	Medium	Low
Severity ratings			
Irrelevant control circumstances			
Laypeople (own sense of justice)	6.73 (0.49)	3.95 (1.40)	2.64 (1.06)
Laypeople (legal system)	6.58 (0.51)	3.98 (1.18)	3.26 (1.30)
Lawyers (legal system)	6.22 (1.33)	4.26 (1.71)	3.71 (1.76)
Relevant exculpatory circumstances			
Laypeople (own sense of justice)	6.18 (0.90)	3.06 (1.30)	2.22 (0.97)
Laypeople (legal system)	6.06 (0.64)	3.15 (0.84)	2.50 (1.26)
Lawyers (legal system)	5.30 (1.18)	3.06 (1.53)	2.81 (1.36)
Certainty ratings			
Irrelevant control circumstances			
Laypeople (own sense of justice)	6.60 (0.57)	6.01 (0.93)	5.67 (0.91)
Laypeople (legal system)	6.34 (0.74)	5.86 (1.08)	5.45 (1.05)
Lawyers (legal system)	6.77(0.29)	6.58 (0.60)	6.48 (0.43)
Relevant exculpatory circumstances			
Laypeople (own sense of justice)	5.20 (1.02)	5.41 (0.77)	5.32 (0.84)
Laypeople (legal system)	4.73 (0.92)	4.69 (0.82)	4.53 (1.01)
Lawyers (legal system)	5.84 (0.80)	6.39 (0.55)	6.20 (0.58)

Note. Severity and certainty ratings range from 1 (not severe at all/ not certain at all) to 7 (very severe/ very certain).

3.2.2.3. Defeated Conclusions (no-punishment)

Percentages of no-punishment conclusions are shown in Table 9. As participants hardly considered irrelevant control circumstances, only a 3 (moral outrage: high vs. medium vs. low) × 3 (group: laypeople - own sense of justice vs. laypeople - legal system vs. lawyers - legal system) mixed ANOVA for problems with relevant exculpatory circumstances was

conducted. The ANOVA revealed a main effect of group, $F(2, 58) = 12.70$, $p < .001$, $\eta_p^2 = .305$, of moral outrage, $F(1.80, 104.31) = 45.11$, $p < .001$, $\eta_p^2 = .437$, and an interaction between both factors, $F(3.60, 104.31) = 3.75$, $p = .009$, $\eta_p^2 = .114$. For low moral outraging offences there were no differences between lawyers and both groups of laypeople, $F(2, 58) = 2.63$, $p = .081$, $\eta_p^2 = .083$. However, there were differences in cases of medium, $F(2, 58) = 6.84$, $p = .002$, $\eta_p^2 = .191$, and especially in cases of high moral outrage, $F(2, 58) = 18.91$, $p < .001$, $\eta_p^2 = .395$. Laypeople in the own sense of justice and in the legal system group made less no-punishment conclusions than lawyers in cases of medium moral outrage ($t(34.31) = 2.83$, $p = .008$, $d = .89$; and $t(39) = 3.77$, $p = .001$, $d = 1.18$; respectively) and even less no-punishment conclusions in cases of high moral outrage ($t(39) = 5.55$, $p < .001$, $d = 1.73$; and $t(39) = 4.92$, $p < .001$, $d = 1.54$; respectively). Both groups of laypeople did not differ from each other neither in cases of high, nor in cases of medium moral outrage (ts < 0.595, ps > .555). Note that according to this Chapter's Hypothesis 1, lawyers' no punishment decisions for high and low moral outrage offences did not differ significantly, $t(20) = 1.60$, $p = .126$, $d = 0.47$.

Table 9

Percentages (and standard deviations) for no-punishment conclusions in Experiment 2

	Moral outrage of the conditional		
	High	Medium	Low
Irrelevant control circumstances			
Laypeople (own sense of justice)	0.00 (0.00)	6.67 (11.34)	21.67 (17.19)
Laypeople (legal system)	3.33 (10.26)	9.17 (15.74)	12.50 (16.11)
Lawyers (legal system)	4.76 (15.04)	5.56 (12.17)	10.32 (12.33)
Relevant exculpatory circumstances			
Laypeople (own sense of justice)	24.17 (21.95)	60.00 (26.16)	64.17 (23.74)
Laypeople (legal system)	28.33 (22.36)	55.83 (22.48)	56.67 (27.78)
Lawyers (legal system)	63.50 (23.35)	80.16 (18.72)	73.81 (20.12)

3.2.2.4. Decision Times

Decision times for punishment and no-punishment conclusions for problems with relevant exculpatory circumstances were analyzed in two separate 2 (conclusion: punishment vs. no-punishment) × 3 (moral outrage: high vs. medium vs. low) within-subjects ANOVAs – one for laypeople and one for lawyers (Figure 2). Laypeople were analyzed as a single group because the two sets of instructions (own sense of justice and legal system) did not affect their punishment conclusions. Only participants from whom I had punishment and no-punishment conclusions in each moral outrage condition were considered in the analysis (25 laypeople and 12 lawyers). This was necessary to be able to make reliable within subject comparisons. Due to technical problems, decision times of one participant were not included in the analysis. To control for different sentence lengths, decision times were adjusted by computing the latency per character for each sentence and multiplying it by the mean sentence length.

For laypeople, the ANOVA revealed no main effects (Fs < 2.38, ps > .136), but a significant interaction between conclusion and moral outrage, $F(1.63, 39.13) = 5.39$, $p = .013$, $\eta_p^2 = .183$. As shown in Figure 2, whereas laypeople's decision times for punishment conclusions did not differ according to moral outrage, $F(1.35, 32.44) = 0.96$, $p = .392$, $\eta_p^2 = .038$, the decision times for their no-punishment conclusions did, $F(1.21, 28,93) = 5.81$, $p = .018$, $\eta_p^2 = .195$. Descriptively, in cases of no-punishment, decision times were longer for high than for medium moral outrage, $t(24) = 2.22$, $p = .036$, $d = 0.26$, and decision times of medium moral outrage were longer than those of low moral outrage , $t(24) = 2.08$, $p = .048$, $d = 0.43$. Even though the significance did not reach the Bonferroni adjusted alpha level of 0.025, the linear trend analysis was significant, $F(1, 24) = 6.58$, $p = .017$, $\eta_p^2 = .215$. This interaction between conclusion and moral outrage was not replicated for lawyers, $F(2, 22) = 0.53$, $p = .596$, $\eta_p^2 = .046$. Rather, there was only a main effect of decision, $F(1, 11) = 5.23$, $p = .043$, $\eta_p^2 = .322$, with lawyers taking overall more time to decide punishment (M = 11531ms; SD = 5235) than no-punishment (M = 8107ms; SD = 1439). Also the main effect of moral outrage was not significant, $F(1.35, 14.80) = 0.32$, $p = .647$, $\eta_p^2 = .028$.

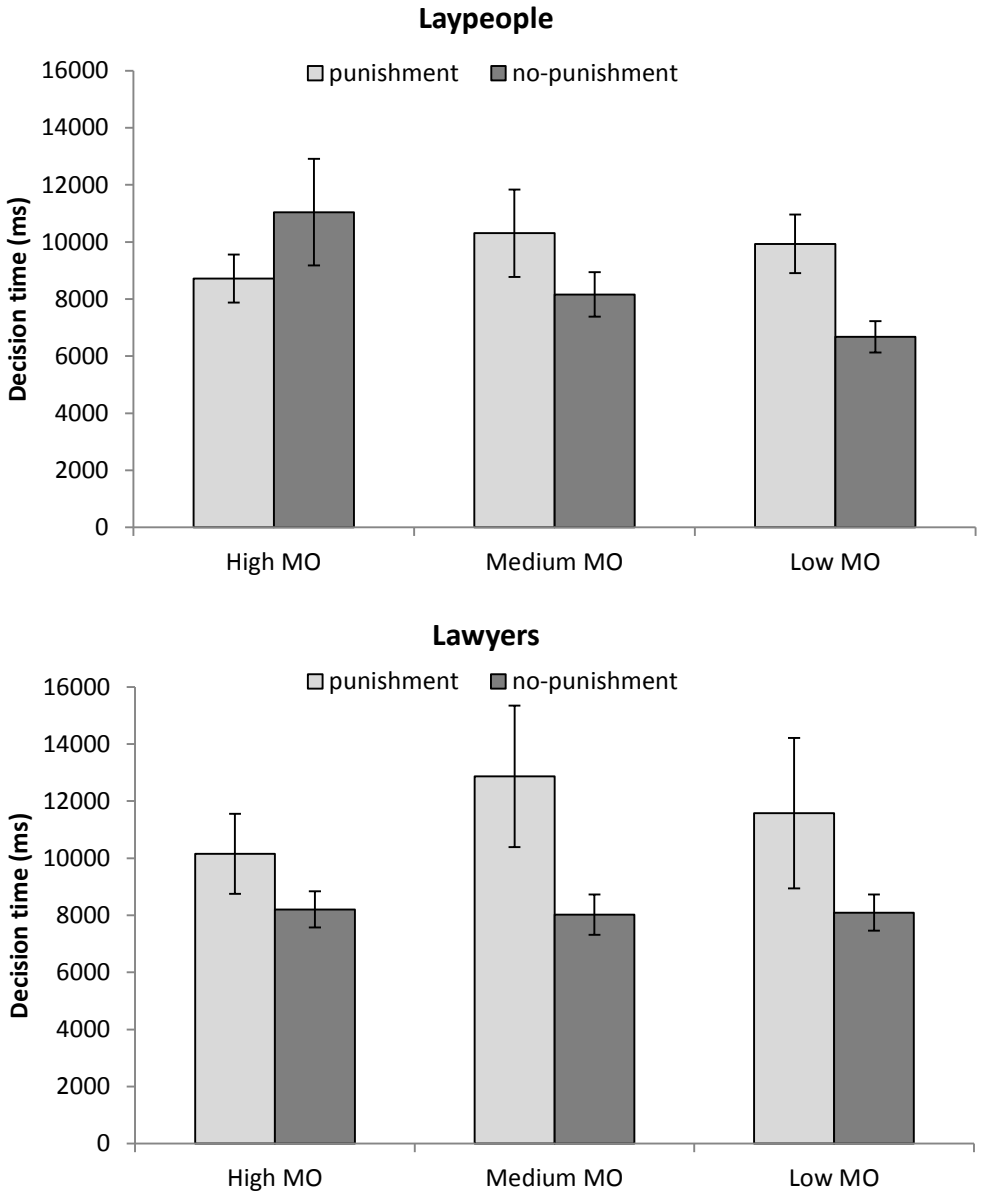

Figure 2. Decision times for punishment and no-punishment conclusions for laypeople and lawyers in Experiment 2, separated by the moral outrage (MO) evoked by the conditionals. Error bars represent standard errors.

3.2.3. Discussion

The results of Experiment 2 show that laypeople's decisions about exculpatory circumstances depend on how morally outrageous the offence in the legal conditional is. This supports Hypothesis 1. When the offence was of high moral outrage, laypeople only seldom decided not to punish offences in light of exculpatory circumstances. Yet, when the offence was of

low moral outrage, laypeople decided in the majority of the cases not to punish the offender. Consequently, laypeople's punishment conclusions did not differ from lawyers in cases of low moral outrage, but only in cases of high moral outrage. This supports Hypothesis 2 and suggests that laypeople do not reject exculpatory circumstances because of the violation of a norm per se, but because the moral outrage evoked by the violation affects the way they reason about exculpatory circumstances. Likewise, since the different offences were paired with the very same exculpatory circumstances (absence of criminal responsibility due to psychological disorders, mistakes of law, and situations of necessity brought about by coercion), the different punishment conclusions laypeople made cannot be attributed to not recognizing these circumstances as exculpatory. All in all, the fact that laypeople sometimes decided to punish in light of a given exculpatory circumstance and sometimes not indicates that their consideration of exculpatory circumstances as defeaters depended on the degree of moral outrage evoked by the offence.

The hypotheses are also supported by the decision times: the higher the moral outrage, the longer laypeople took to reach a no-punishment decision; this reflects the difficulty in deciding against moral outrage; as proposed by Hypothesis 3. As shown in Figure 2, when the offence was only of low moral outrage, decision times for not punishing the offences were faster than for punishing. Cases of illegal gambling or obtaining benefits by devious means are not offences with a high moral necessity of punishment, so deciding in favor of punishment is almost counterintuitive and may consequently take longer. Accordingly, laypeople in the own sense of justice group also showed a tendency to not to punish offenders for low moral outrage offences with irrelevant circumstances (see Table 9). However, also in cases of medium moral outrage no-punishment decisions were somewhat faster than punishment decisions. Though they were chosen to evoke some amount of moral outrage, the severity ratings showed that these offences were not considered very severe (around 3.5 on a 7 point scale). Therefore these offences, too, were likely judged not to deserve strict punishment.

Lawyers decided about exculpatory circumstances according to the penal code. They were somewhat stricter in cases of high moral outrage, but this was probably only because of the legal principle of proportionality, but not primarily because of moral outrage. This interpretation is supported by the decision times, where no significant differences depending on moral outrage were found. In fact, lawyers were always faster in selecting a no-punishment conclusion, indicating that most of the exculpatory circumstances were recognized quickly and without bias. The high decision times for punishment conclusions indicate that when

lawyers decided to incorrectly reject an exculpatory circumstance, this was a hard decision for them. However, because of the small sample size of people selecting punishment as well as no-punishment conclusions for all conditions, these results should be interpreted with caution.

The two different instructions (own sense of justice vs. legal system) given to laypeople did not affect their punishment conclusions. The results did thus not support Hypothesis 4. I expected a higher moral outrage effect for laypeople in the own sense of justice condition than for laypeople in the legal system condition. Yet, the effect of moral outrage was found in both conditions. One possible explanation is that participants did not follow the instructions to decide on the basis of the regulations of the legal system. However, I do not think this was the case: laypeople given legal system instructions seemed to understand the perspective they were to take. On the one hand, they were less certain than laypeople in the own sense of justice condition in deciding about problems with relevant exculpatory circumstances. On the other hand, laypeople assigned to the legal system condition reported in an open-ended questionnaire at the end of the experiment that they followed the instructions and tried to reason like a real judge. Nevertheless, 65% of them also said that this was a difficult task due to conflicts with their own sense of justice or that they were aware that their opinions and sense of morality still influenced their decisions. This indicates how deeply our morality and sense of justice is engrained in our beliefs about exceptions to legal rules and how this affects people's willingness to withdraw the conclusion from a legal conditional rule.

3.3. Experiment 3: Generating Exculpatory Circumstances

In the previous two experiments exculpatory circumstances were always presented explicitly together with the legal conditional and the categorical statement; participants were not instructed to think of exculpatory circumstances themselves. But how well can people themselves retrieve from memory exculpatory circumstances to a legal rule? And is the availability of exculpatory circumstances affected by the level of moral outrage evoked by an offence? In Section 1.1.1.1 I discussed the importance of memory in accepting conditional rules (Chan & Chua, 1994; De Neys et al., 2003a; Markovits & Barrouillet, 2002; see also Markovits & Quinn, 2002). These studies showed that when people make a conditional

inference they search their memory for domain relevant information, e.g., potential defeaters to the rule. The discovery of defeaters in memory increases the probability of not accepting the conditional rule and triggers the withdrawal of MP inferences (De Neys et al., 2003a). Hence, if the search for defeaters in memory is essential to the application of conditional rules, then the previous experiments might indicate that the ability to recall exculpatory circumstances for legal rules varies between lawyers and laypeople. To test this, I changed the experimental paradigm and asked participants to generate exculpatory circumstances in a paper-and-pencil task. My assumptions are that (1) lawyers know exculpatory circumstances from their law studies and should therefore be able to recall them independently of moral outrage, whereas (2) laypeople's capacity to retrieve exculpatory circumstances depends on the moral outrage evoked by the offence: the higher the feelings of moral outrage, the more difficult it should be to retrieve an exculpatory circumstance. As the number of exculpatory circumstances in memory may be confounded with the familiarity of the domain, I also asked participants to generate aggravating circumstances and compared those with the number of exculpatory circumstances. It is predicted that it should be more difficult for laypeople to think of exculpatory circumstances than aggravating circumstances, and this difficulty should vary with the moral outrage evoked by the offence. In contrast, lawyers' amount of retrieved exculpatory circumstances should not depend on moral outrage.

Hypothesis 1: Laypeople generate less exculpatory circumstances than lawyers. However, laypeople do not differ from lawyers in the generation of aggravating circumstances because aggravating circumstances do not conflict with laypeople's desire to punish offenders.

Hypothesis 2: The amount of exculpatory circumstances laypeople generate depends on the degree of moral outrage evoked by the offence: The more morally outraging the offence is, the less exculpatory circumstances compared to aggravating circumstances laypeople generate. Lawyers are not affected by moral outrage.

3.3.1. Methods

3.3.1.1. Participants

Participants were 20 lawyers (9 female) and 20 laypeople (13 female). One additional layperson also took part but was unable to complete the experiment and was therefore removed from the data file. The mean age of lawyers was 25.4 years (*SD* = 1.96); the mean age of laypeople was 23 years (*SD* = 1.41; 5 missing values). Two participants from the lawyers' group finished their law studies. The rest were still at university and studied for 9.2 semesters on average.

3.3.1.2. Material and Design

For the experiment six offences from the German penal code were selected: *theft, coercion, bodily injury, abortion, manslaughter* and *incest.* These offences differ in their penalty range and were selected on the basis of the amount of exculpatory and aggravating circumstances in the German penal code. An online study (*N* = 312; 224 female) was conducted to measure levels of moral outrage evoked by these offences. Participants rated their level of moral outrage on a 7-point Likert scale (1 = *no moral outrage*, 7 = *great moral outrage*). This online study showed that the offences evoke different levels of moral outrage: manslaughter (*M* = 6.54; *SD* = 0.83), bodily injury (*M* = 5.71; *SD* = 1.12), coercion (*M* = 5.15; *SD* = 1.24), theft (*M* = 4.33; *SD* = 1.30), incest (*M* = 4.31; *SD* = 1.89), and abortion (*M* = 2.55; *SD* = 1.72).

In the main study, offences were presented in a paper booklet consisting of two parts. One part asked for exculpatory and mitigating circumstances, and the other part asked for aggravating circumstances. The order of these parts was counterbalanced across participants. On each page there were two offences. The sequence of pairs of offences over all problems was randomized. Participants were also asked for mitigating circumstances to guarantee that exculpatory circumstances were actually considered exculpatory and not just mitigating. The experiment followed thus a 3 (category: exculpatory vs. mitigating vs. aggravating) × 2 (group: laypeople vs. lawyers) design. However, as mitigating circumstances were only used to ensure the clarity of the distinction between exculpatory and aggravating circumstances, these were not included in the analysis.

3.3.1.3. Procedure

The experiment was a paper and pencil experiment and participants were tested either in groups or individually. The instructions explained the meaning of exculpatory, mitigating, and aggravating circumstances. Participants were instructed to write down all thinkable situations they would consider exculpatory, mitigating, or aggravating circumstances for a given offence. Exculpatory circumstances were described as circumstances which prevent punishment entirely, mitigating circumstances as those that lower a sentence, and aggravating circumstances as those that elevate a sentence. Participants were told that it was irrelevant whether the situations were regulated in the penal code. One sample problem was given to illustrate the tasks. There were no time restrictions. The experiment took about 45 minutes. All participants received monetary compensation for their participation.

3.3.2. Results

Two raters independently counted the number of situations generated for the different offences (Kendall's tau =.967 for exculpatory circumstances; Kendall's tau=.949 for aggravating circumstances). The mean number of these situations (i.e., defeaters) was analyzed using a 2 (circumstances: exculpatory vs. aggravating) × 2 (group: laypeople vs. lawyers) mixed ANOVA. The ANOVA revealed main effects of group, $F(1, 38) = 14.03$, $p = .001$ $\eta_p^2 = .270$, of circumstances, $F(1, 38) = 6.43$, $p = .015$, $\eta_p^2 = .145$, and an interaction between group and circumstances, $F(1,38) = 9.28$, $p = .004$, $\eta_p^2 = .196$. Laypeople ($M = 3.56$; $SD = 1.42$) and lawyers ($M = 4.38$; $SD = 1.56$) did not differ in the amount of aggravating circumstances generated, $t(38) = 1.73$, $p = .092$, $d = 0.55$, but laypeople generated significantly fewer exculpatory circumstances ($M = 1.60$; $SD = 0.75$) than lawyers ($M = 4.56$; $SD = 3.17$), $t(21.133) = 4.06$, $p = .001$, $d = 1.28$. Moreover, lawyers did not generate different amounts of exculpatory and aggravating circumstances, $t(19) = 0.29$, $p = .778$, $d = 0.06$, whereas laypeople listed twice as many aggravating than exculpatory circumstances, $t(19) = 6.15$, $p < .001$, $d = 1.67$.

To test whether the difference in number of exculpatory and aggravating circumstances was related to moral outrage, the difference between the amount of aggravating and amount of exculpatory circumstances was computed for each offence. As expected, I did

find such an effect: the higher the moral outrage evoked by an offence, the fewer exculpatory (compared to aggravating) circumstances laypeople generated (Figure 3), with the following trend: manslaughter > bodily injury > coercion > theft > incest > abortion. This rank order was corroborated by Page's trend test, Page's $L = 1628$, $p < .01$, and resembles the moral outrage ratings from the online study for the different offences. Lawyers did not show this trend (although Page's trend was still significant, Page's $L = 1545.5$, $p < .05$, but as can be seen Figure 3, the pattern among offences was not clear for this group and did not resemble that of laypeople at all).

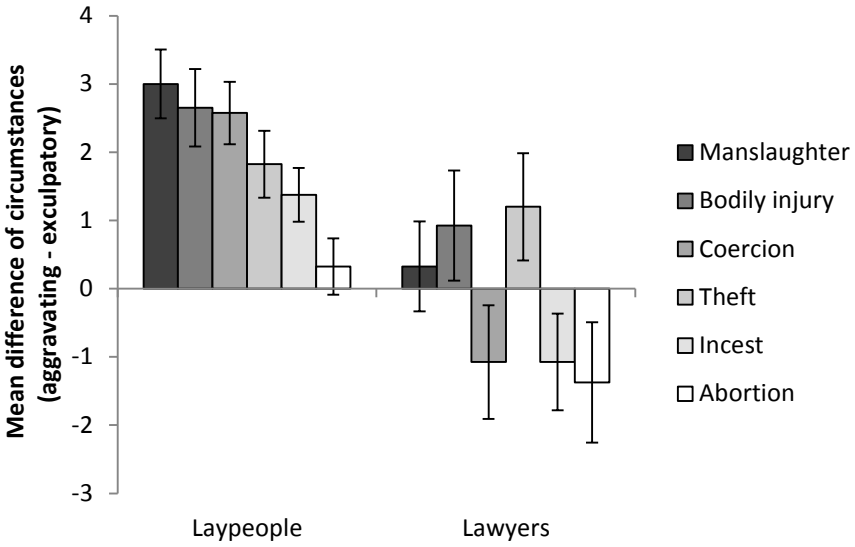

Figure 3. Mean differences between the amount of aggravating and exculpatory circumstances per offence in Experiment 3. Error bars represent standard errors.

3.3.3. Discussion

The results of this experiment suggest that lawyers and laypeople have different mental representations of exculpatory circumstances. Whereas lawyers easily generated exculpatory and aggravating circumstances, laypeople had difficulties in thinking of exculpatory circumstances, especially for offences of high moral outrage. This supports Hypotheses 1 and 2. These results show that the effect of moral outrage is not limited to inferences with legal conditionals; it also affects retrieval of exculpatory circumstances from memory, and thereby the interpretation of the conditional itself. However, one can still argue that the difficulty to retrieve exculpatory circumstances for specific offences does not indicate that they are not

stored in memory. Exculpatory circumstances may be stored in memory, but not retrieved because they are not in accordance with the person's moral values. This explanation is plausible and might also apply to other experiments where participants are asked to generate exceptions (e.g. Cummins, 1995; Cummins et al., 1991; De Neys et al., 2003b). However, when investigating how people reason with legal conditionals, only the defeaters that are actually retrieved and considered are important for the inference process. The results of Experiment 2 show that, when instructed to act like a real judge, laypeople still decide according to moral outrage, which in turn indicates that even if there are some defeaters available in memory, they are rarely considered and therefore have no observable effect on reasoning.

Despite these correspondences between the amount of exculpatory circumstances retrieved from memory and the moral outrage of an offence, the relationship is only correlational. One cannot know whether the difficulty in generating exculpatory circumstances is caused by the moral outrage evoked by the offence or whether the moral outrage evoked by an offence is caused by a small number of exculpatory circumstances stored in memory. Although this aspect cannot be clarified here, I believe that the influence is bidirectional: if not finding many exculpatory circumstances leads participants to classify an offence as highly morally outrageous, then this assessment will in turn hinder them when searching for other possible exculpatory circumstances. Further empirical evidence is needed to ground this.

3.4. Summary and Implications of Chapter 3

In this first block of experiments, I showed that lawyers and laypeople defeat conclusions from legal conditionals in light of exculpatory circumstances, but in a different way: lawyers seem to weigh circumstance information according to what is prescribed by the penal code, but laypeople seem to base their decisions on their own sense of justice, guided by feelings of moral outrage. Because of that, laypeople had difficulties in accepting exculpatory circumstances when the offence was of high moral relevance, adhering therefore more strongly to an initial conditional rule than lawyers. Consequently, compared to lawyers, laypeople had difficulties in withdrawing a logically (and perhaps morally) valid conclusion,

even when instructed to decide like an actual judge. Experiment 3 suggests that this difficulty seems to arise from an incapability to retrieve exculpatory circumstances for morally outrageous offences from memory.

The results of all three experiments show that accepting a given fact as a defeater is not a trivial task. The acceptance of defeaters depends on a person's domain knowledge and on a person's attachment to the initial conditional rule – which in the case of legal conditionals is a person's own sense of justice. In this way, the feelings of moral outrage evoked by the offences in the legal conditionals probably affected the perceived sufficiency of the antecedent (i.e., the offence) to its consequent (i.e., the punishment). The more morally outraged reasoners were by the offence described in the antecedent, the more they concluded that the offender should be punished and – consequently – the less affected they were by potential defeaters. Applying the terminology of Markovits and Potvin (2001) and De Neys and colleagues (2002; 2003b) one could argue that the associative strength between offence and punishment in cases of high moral outrage is so strong, that it is hard for defeaters to break it. That is, reasoners' conclusions are modulated by their own preferences and attitudes towards offences. In this respect it is also possible to relate the effect of moral outrage on laypeople's acceptance of exculpatory circumstances to the literature on the importance of utilities in reasoning. As explained in Section 1.2.2, feelings of moral outrage are related to a desire to punish offenders. It is therefore possible that the perceived utility of punishing somebody is greater than the subjective utility of acquitting somebody of an offence. This utility based explanation would account for why laypeople decided to rely on the legal conditional rule (Experiments 1 and 2) even when they can actually think of at least one exculpatory circumstance (Experiment 3). However, utilities cannot account for the lawyers' conclusions, because they actually know which information invalidates which conclusions without the need of computing utilities. The discussion on how the reasoning processes of laypeople and lawyers differ, and which reasoning theory best explains the effects observed in this chapter, is continued in the General Discussion in Chapter 7.

In sum, the results of this first experimental block support the hypothesis that laypeople's own sense of justice affects the conclusion they draw from legal conditionals. But what happens if people or groups of people differ in their feelings of justice? In the next Chapter, I will investigate how differences in a person's attitudes to offenders and offences affect legal conditional reasoning and the withdrawal of conclusions.

Chapter 4: Legal Conditional Reasoning, Religiosity, and Culture

Chapter 3 showed that laypeople's sense of justice influences their withdrawal of conclusions in legal reasoning. The more morally outrageous an offence is, the less laypeople accept exculpatory circumstances as reasons for voiding punishment, and the less they defeat the otherwise valid conclusion of punishing the offender. As explained in Section 1.2.2, these reactions of moral outrage and the desire to punish offenders are usually described as widespread (Tyler & Boeckmann, 1997; see also Rozin, Lowery, Imada, & Haidt, 1999). However, people's intuitive desire to punish offenders can be lessened. In the previous chapter I showed that lawyers' punishment decisions were not influenced by moral outrage. Similarly, also societal rules of behavior may shape people's perception of offences. For instance, if societal norms of behavior enhance forgiveness, then people might be more open to except exculpatory circumstances and thus engage more in defeasible reasoning. Similarly, if one's close social environment enhances the condemnation of offences, then people should be even more reluctant to withdraw punishment. The aim of this second block of experiments is to investigate the second main hypothesis of this thesis: whether differences in a layperson's attitudes about offenders and offences affect legal conditional reasoning. If cultural or ideological norms of behavior enhance or inhibit the excuse of offences, then it should have the same twofold effect on laypeople's defeasible reasoning with legal rules. I will test this hypothesis by focusing on two fields: religiosity and culture.

4.1. Experiment 4: Religiosity[4]

Religions often provide behavioral guidelines which are concerned with injustices and offences. A good example are Christianity's Ten Commandments, which prohibit killing, stealing, adultery, and giving false testimony. Experiment 4 therefore investigates how religiosity affects people's consideration of exculpatory circumstances in legal reasoning. I

[4] Data for this experiment was partially gained during the Bachelor Thesis of Christian Kirchner (Kirchner, 2013), which I supervised.

will concentrate on Christianity whose primary guidelines concerning offences are set out in the Ten Commandments. The existence of such clearly formulated and rather inflexible rules suggests that highly religious people are more dogmatic than less religious ones. In fact, several studies have found such correlations (e.g. Juan & Haley, 1970; Feather, 1964; Kilpatrick et al., 1980; Swindell & L'Abate, 1970; for a review see Ross et al., 2005). Moreover, studies also show that highly religious people are more compliant with rules (e.g., Grasmick, Kinsey, & Cochran, 1991). Consequently, by relating this high dogmatism to defeasible reasoning, we can expect highly religious people to be more reluctant to accept violations of legal rules and therefore show little consideration of exculpatory circumstances; even for low morally outraging offences. That is, their defeasible reasoning should be rather low.

However, the literature on religiosity is inconsistent. Religion is not only related to dogmatism; but also to forgiveness. For instance, Laurin, Shariff, Henrich, and Kay (2012) showed that people's belief in a powerful intervening god leads to diminishing choice of a stronger punishment. In fact, several studies show that greater religiosity is related to a more prosocial behavior (e.g., Hardy & Carlo, 2005; Mattis, et al., 2000; Saroglou et al., 2005; Smith, 1999; see also Batson & Gray, 1981; Hansen et al., 1995), and that forgiveness and prosociality are closely linked (see e.g. Karremans, Van Lange, & Holland, 2005; McCullough, 2001). As a result – contrary to what could be expected when linking religiosity to high dogmatism – this relationship between religiosity and forgiveness should lead highly religious people to be more open to accept exculpatory circumstances when compared to less-religious people; even when the offence causes strong moral outrage. That is, highly religious people's defeasible reasoning should be rather high.

The inconsistent literature on religiosity does not permit reliable predictions on how religiosity might influence people's defeasible reasoning with legal rules. That is, it is difficult to predict whether highly religious people will be more open to defeat punishment conclusions than less-religious ones. However, the fact that one's religion promotes both, dogmatism and forgiveness, should result in feelings of ambivalence and uncertainty. This ambivalence should lead highly religious people to feel conflict when presented with exculpatory circumstances. Such people cannot rely solely on their feelings of moral outrage to determine punishments; they must also consider their religion's doctrines of dogmatism and forgiveness which might point towards different verdicts. Although this conflict inhibits reliable predictions about concrete punishment decisions, it should result in lower certainty ratings and longer decision times. Conversely, less-religious people should not feel this

conflict between dogmatism and forgiveness, or at least not as strongly as highly religious people. Less-religious people can decide based on what they spontaneously feel, without having to compare their preferences to religious standards.

Hypothesis 1: Highly religious people need more time to decide about exculpatory circumstances than less-religious people.

Hypothesis 2: Highly religious people should be less certain than less religious people when deciding about exculpatory circumstances.

4.1.1. Methods

4.1.1.1. Participants

Participants were selected based on a preliminary study (N = 604) using a German translation of the religiosity scale used by Hardy and Carlo (2005). The preliminary study was conducted online via SoSci Survey (Leiner, 2014) and LimeSurvey™, as well as in person in the lab. The religiosity scale by Hardy and Carlo (2005) consists of seven items which measure religious salience or commitment, religious activity or involvement, and religious identity. It uses 5-point Likert scales in which higher ratings are related to greater religiosity. The items were: (a) How important is religion in your life? (b) How often do you go to church? (c) How often do you attend church-related activities like e.g. youth activities (other than worship services)? (d) I am a spiritual person. (e) I practice my religion. (f) My faith never deserts me during hard times; and (g) My faith makes me who I am. Hardy and Carlo (2005) report to have taken the last four items from the spiritually subscale of the Values in Action Inventory of Strengths for Youth (Peterson & Seligman, 2004).

In order to find highly and less-religious participants I computed the mean score over all items (cf. Hardy & Carlo, 2005). Participants with one point were classified as less religious and participants with four or more points as highly religious. For the experiment, 20 participants (10 females) were classified as highly religious and 20 participants (12 females) as less religious. The group of highly religious participants reached a mean score of 4.46 points (SD = 0.25) and the group of less-religious participants reached a mean score of 1.00

point (*SD* = 0.00). The difference in mean scores was highly significant, $t(19) = 62.21$, $p < .001$, $d = 19.67$. Highly religious participants had a mean age of 24 years (*SD* = 2.94) and less-religious participants had a mean age of 23.2 years (*SD* = 3.46). It was ensured that all participants in the high-religious group were Christians: 40% were Catholics, 55% were Protestants (4 of them from a free church), and one participant simply described himself as a Christian. In the less-religious group 40% reported to be Christians by baptism, but given their scores in the preliminary study it was concluded that religion was not important to them. In an open interview in the end of the experiment many of them reported that religion was unnecessary, manipulative, and only for tradition.

4.1.1.2. Materials, Procedure, and Design

Material and procedure were the same as in Experiment 2 (Chapter 3), with the only difference that now all participants were instructed to use their own sense of justice during the experiment. The experiment followed thus a 2 (religiosity: high vs. low) x 3 (moral outrage: high vs. medium vs. low) x 2 (circumstance: legally exculpatory vs. control) mixed design. Religiosity was varied between individuals; all other factors were varied within individuals.

4.1.2. Results

4.1.2.1. Perceived Severity

Severity ratings were analyzed with a 2 (religiosity: high vs. low) x 3 (moral outrage: high vs. medium vs. low) x 2 (circumstances: legally exculpatory vs. control) mixed ANOVA. Results can be found in the upper part of Table 10. The ANOVA showed a main effect of circumstances, $F(1, 38) = 54.23$, $p < .001$, $\eta_p^2 = .588$, a main effect of moral outrage, $F(1.55; 58.77) = 328.06$, $p < .001$, $\eta_p^2 = .896$, and an interaction between moral outrage and circumstances, $F(1.80; 68.47) = 10.25$, $p < .001$, $\eta_p^2 = .212$. Problems with control circumstances were perceived as more severe than problems with legally exculpatory circumstances (*M* = 4.71; *SD* = 0.75 and *M* = 3.8; *SD* = 0.84, respectively). High-moral outrage offences were always perceived as more severe than medium-moral outrage offences (*M* = 6.41; *SD* = 0.66 and *M* = 3.60; *SD* = 0.91, respectively), $t(39) = 19.86$, $p < .001$, $d =$

3.49, and medium-moral outrage offences as more severe than low-moral outrage offences (M = 2.74; SD = 1.03), $t(39) = 7.62$, $p <. 001$, $d = 0.89$. However, the interaction between circumstances and moral outrage shows that the mitigating effect of legally exculpatory circumstances was higher for problems with a medium level of moral outrage, $t(39) = 7.443$, $p < .001$, $d = 1.20$, compared to cases with high, $t(39) = 4.616$, $p < .001$, $d = 0.68$, or low-moral outrage, $t(39) = 5.722$, $p < .001$, $d = 0.76$. No main effect of religiosity or interactions concerning religiosity were found (all Fs < 0.60, $p > .450$). That means participants' religiosity had no effect on severity ratings.

Table 10

Severity and certainty ratings of highly and less-religious participants for legally exculpatory and control circumstances in Experiment 4

	Moral Outrage		
	High	Medium	Low
Severity ratings			
Highly religious			
Legally exculpatory	6.11 (0.84)	2.94 (1.00)	2.32 (1.02)
Control	6.70 (0.60)	4.40 (1.15)	3.28 (1.19)
Less-religious			
Legally exculpatory	6.14 (1.10)	2.99 (1.10)	2.28 (1.10)
Control	6.71 (0.40)	4.08 (1.03)	3.08 (1.30)
Certainty ratings			
Highly religious			
Legally exculpatory	4.87 (1.28)	4.87 (1.17)	4.83 (1.15)
Control	6.53 (0.52)	5.98 (0.94)	5.58 (1.00)
Less-religious			
Legally exculpatory	5.40 (1.24)	5.41 (1.20)	5.60 (1.08)
Control	6.55 (0.76)	5.90 (1.25)	5.78 (0.96)

Note. Severity and certainty ratings range-from 1 (*not severe at all* or *very unsure*, respectively) to 7 (*very severe* or *very sure*, respectively). The numbers in brackets indicate the corresponding standard deviations.

4.1.2.2. Defeated Conclusions (no-punishment)

The percentages of no-punishment conclusions were analyzed with a 2 (religiosity: high vs. low) x 3 (moral outrage: high vs. medium vs. low) x 2 (circumstances: legally exculpatory vs. control) mixed ANOVA. Results can be found in Figure 4. The ANOVA revealed a main effect of circumstances, $F(1, 38)=158.27$, $p < .001$, $\eta_p^2 = .806$, a main effect of moral outrage, $F(1.54, 58.67) = 38.34$, $p < .001$, $\eta_p^2 = .502$, and an interaction between moral outrage and circumstances, $F(1.65, 62.79) = 33.13$, $p < .001$, $\eta_p^2 = .466$. However, there was also a three-way interaction showing that participants' punishment decisions depended on the quality of the circumstance information, $F(1.65, 62.79) = 6.96$, $p = .003$, $\eta_p^2 = .155$. Therefore, I decided to further analyze the data by conducting two separate 2 (religiosity: high vs. low) x 3 (moral outrage: high vs. medium vs. low) ANOVAs, one for legally exculpatory circumstances and one for control circumstances. The ANOVA for legally exculpatory circumstances only showed the expected moral-outrage effect, $F(1.56, 59.39) = 51.281$, $p < .001$, $\eta_p^2 = .574$. Participants made fewer no-punishment decisions when offences were of medium-moral outrage compared to high-moral outrage ($M = 62.50\%$; $SD = 25.53$ and $M = 26.67\%$; $SD = 26.09$, respectively), $t(39) = 7.78$, $p < .001$, $d = 1.39$. Descriptively participants made also fewer no-punishment conclusions for offences of low-moral outrage ($M = 67.50$; $SD = 25.30$) compared to medium-moral outrage. However, this last comparison was not significant, $t(39)=1.64$, $p = .110$, $d = 0.20$. No effects of religiosity were found (all Fs < 0.3, $p > .600$).

The ANOVA for control circumstances, however, showed a different pattern. Although no-punishment decisions were generally scarce, the ANOVA revealed a main effect of religiosity, $F(1, 38) = 5.64$, $p = .023$, $\eta_p^2 = .129$, a main effect of moral outrage, $F(1.31, 49.78) = 13.70$, $p < .001$, $\eta_p^2 = 0.265$, and an interaction between both factors, $F(1.31, 49.78) = 7.09$, $p = .006$, $\eta_p^2 = .157$. As shown in Figure 4, highly religious and less-religious people differed in their percentages of no-punishment conclusions for offences with irrelevant defeaters. Even though the offences were paired with crime irrelevant circumstance information, less-religious people decided not to punish the offender for medium moral outrage offences in 9.2% of the cases, and for low moral outrage offences in 30.8% of the cases. Highly religious people did not show this effect and almost always decided to punish the offender. This preference of less-religious people to not punish offenders with control circumstances was highly significant for low-moral outrage offences, $t(23.43) = 2.68$, $p =$

.013, d = 0.85, but did not reach significance for medium-moral outrage offences, $t(22.58)$ = 1.93, p = .068, d = 0.61.

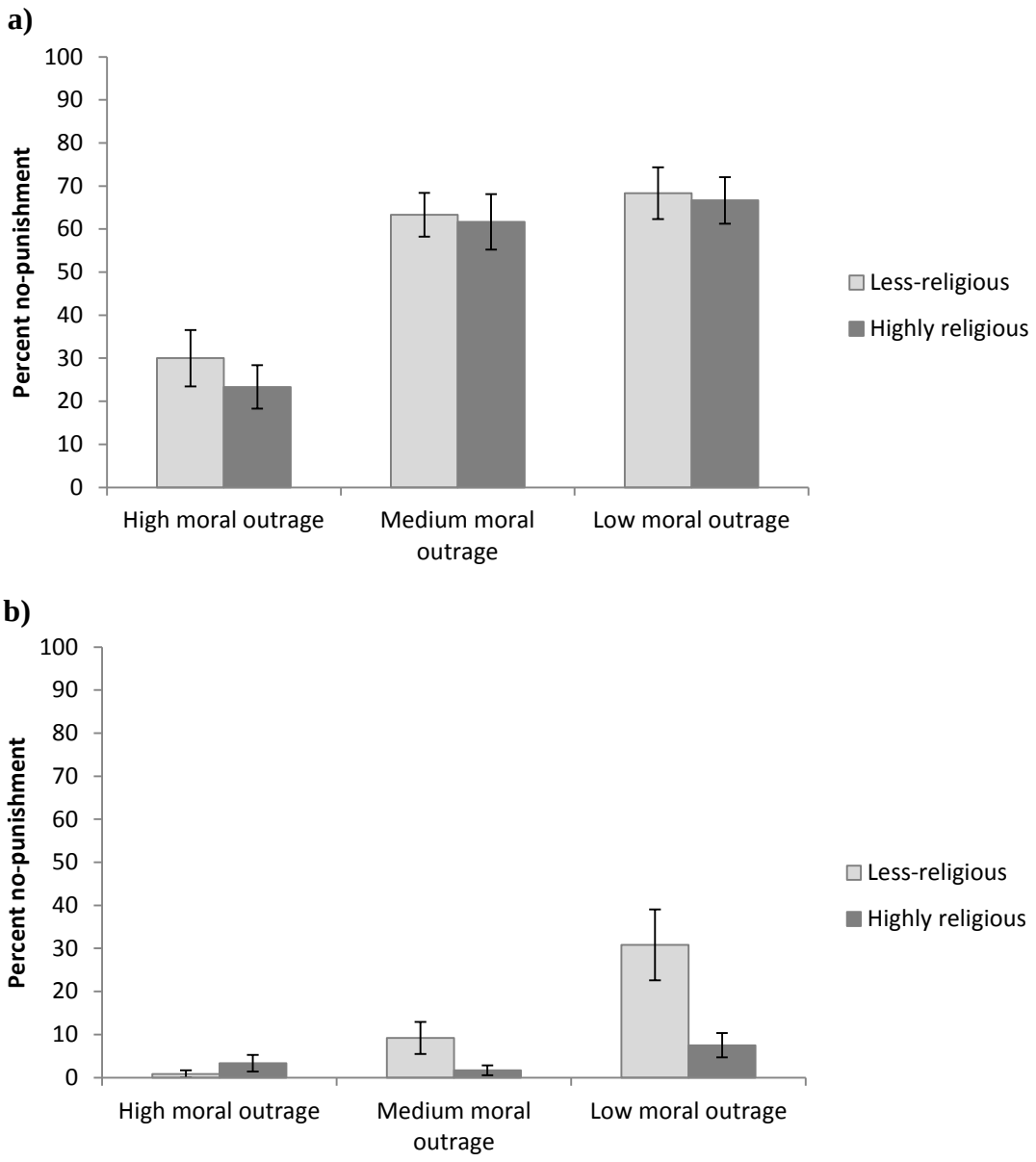

Figure 4. Percentages of no-punishment decisions made by less and highly religious participants for problems with a) relevant exculpatory and b) control circumstances in Experiment 4.

4.1.2.3. Decision Times

Decision times were only analyzed for the problems with legally exculpatory circumstances since they were the only ones where participants selected punishment and no-punishment conclusions for all levels of moral outrage in a representative manner (cf. Section 3.2.2). These decision times were adjusted for reading time as in Experiment 2 and are shown in Figure 5. First, decision times were analyzed with a 2 (religiosity: high vs. low) x 3 (moral outrage: high vs. medium vs. low) x 2 (decision: punishment vs. no-punishment) mixed ANOVA. This ANOVA revealed only an interaction between moral outrage and decision, $F(1.48, 26.54) = 4.68$, $p =. 027$, $\eta_p^2 = .206$, but no main effect of religiosity, $F(1, 18) = 2.45$, $p = .135$, $\eta_p^2 = .120$. However, the three way interaction was also close to reach significance, $F(1.48, 26.54) = 2.89$, $p = .087$, $\eta_p^2 = .138$. As can be seen in Figure 5, decision times of highly and less religious participants seem to follow a different pattern. It was therefore decided to analyze the decision times of both groups of participants separately with two 3 (moral outrage: high vs. medium vs. low) x 2 (decision: punishment vs. no-punishment) ANOVAs. For less-religious people no significant main effects or interactions were found (all Fs < 1.03, $ps \geq .379$). However, the ANOVA for highly religious people showed a significant interaction between decision and moral outrage, $F(2, 18) = 4.91$, $p = .020$, $\eta_p^2 = .353$. As shown in Figure 5, when the offence was of high-moral outrage, highly religious people showed longer decision times when deciding not to punish compared to deciding to punish. But when the offence was of low-moral outrage, highly religious people showed longer decision times when deciding to punish compared to deciding not to punish. These differences were tested with one-tailed post hoc *t*-tests. For these *t*-tests, only the participants that were considered in the ANOVA were considered (i.e., the participants that had selected punishment and no-punishment for all three conditions of moral outrage). Even though the results showed a trend, they did not reach the Bonferroni-adjusted alpha level of 0.016, probably because of the small sample size and large variance ($t(9) = 2.20$, $p = .028$, one-tailed, $d = 0.48$ for high moral outrage; $t(9) = 1.96$, $p = .041$, one-tailed, $d = 0.95$ for low moral outrage). Therefore, I decided to test the same differences with non-parametric one-tailed Wilcoxon tests. As expected, when the offence was of high-moral outrage highly religious people took longer to decide not to punish compared to deciding to punish, $Z = -2.29$, $p = .011$ (one-tailed), $r = .73$. When the offence was of low-moral outrage highly religious participants took longer to decide to punish compared to deciding not to punish, $Z = -2.50$, $p = .007$ (one tailed), $r = .79$.

Decision times for punishment and no-punishment decisions for offences with medium-moral outrage did not differ, $Z = -0.36$, $p = 0.361$ (one-tailed), $r = .11$. All other main effects of the ANOVA were not significant (all Fs < 0.31, $ps \geq .597$)

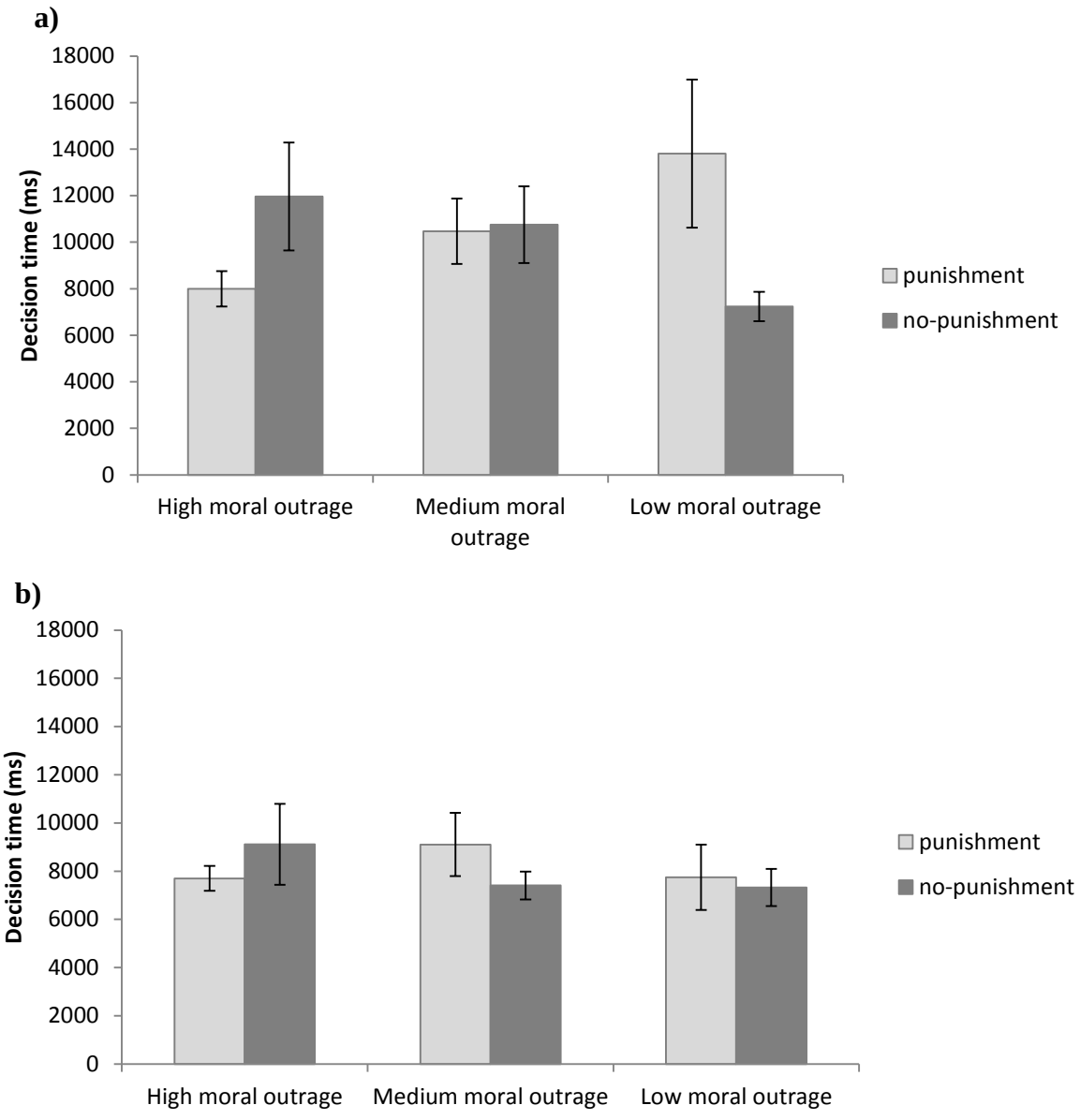

Figure 5. Decision times for punishment and no-punishment decisions for problems with relevant exculpatory circumstances in Experiment 4 for a) highly and b) less-religious participants.

4.1.2.4. Certainty

Certainty ratings were analyzed with a 2 (religiosity: high vs. low) x 3 (moral outrage: high vs. medium vs. low) x 2 (circumstances: legally exculpatory vs. control) mixed ANOVA. Results can be found in the lower part of Table 10. Again, the ANOVA revealed a main effect of circumstances, $F(1, 38) = 32.84$, $p < .001$, $\eta_p^2 = .464$, a main effect of moral outrage, $F(1.81, 68.87) = 7.82$, $p = .001$, $\eta_p^2 = .171$, and an interaction between moral outrage and relevance, $F(1.56, 59.43) = 16.59$, $p < .001$, $\eta_p^2 = .304$. Participants were more certain when problems contained control circumstances ($M = 6.05$, $SD = 0.84$) compared to legally exculpatory circumstances ($M = 5.16$, $SD = 1.10$). Participants were also more certain when the problem was of high-moral outrage ($M = 5.84$, $SD = 0.82$) compared to medium-moral outrage ($M = 5.54$, $SD = 1.00$), $t(39) = 2.58$, $p = .014$, $d = 0.32$, but not when the problem was of medium-moral outrage compared to low-moral outrage ($M = 5.45$, $SD = 0.92$), $t(39) = 1.03$, $p = .310$, $d = 0.10$. However, the interaction between moral outrage and circumstances showed that the higher certainty in control problems compared to experimental problems was highest for high-moral ($M = 6.54$, $SD = 0.64$; $M = 5.13$, $SD = 1.27$, respectively), $t(39) = 7.64$, $p < .001$, $d = 1.30$, followed by medium-moral outrage offences ($M = 5.94$, $SD = 1.15$; $M = 5.14$, $SD = 1.20$, respectively), $t(39) = 4.13$, $p < .001$, $d = 0.68$, and low-moral outrage offences ($M = 5.68$, $SD = 0.97$; $M = 5.22$, $SD = 1.17$, respectively), $t(39) = 2.58$, $p = .014$, $d = 0.42$. Interestingly, the ANOVA also revealed a trend towards an interaction between religiosity and circumstance, $F(1, 38) = 3.33$, $p = .076$, $\eta_p^2 = .08$: As can be seen in Table 10, highly religious people were somewhat less certain when deciding about legally exculpatory circumstances than less-religious people. No further effects were found (all Fs < 1.60, $ps \geq .214$).

4.1.3. Discussion

The results of Experiment 4 show that religiosity influenced participants' legal reasoning. It was expected that due to Christianity's ambivalence between dogmatism and forgiveness, highly religious people – in contrast to less-religious people – should have conflicting feelings when deciding if an offender should be punished or not. This ambivalence should result in longer decision times and less certainty ratings. As expected by Hypothesis 2, highly religious

people where descriptively somehow less certain than less religious people when deciding about legal conditionals with exculpatory circumstances. However, I did not find support for Hypothesis 1. Highly religious people were not overall slower than less-religious people. Instead, there were differences in their overall response pattern. While less-religious participants showed no preferences in choosing between punishment and no-punishment, the decision times of highly religious participants depended on how morally outrageous the offence was. When the offence was of high moral outrage, highly religious people needed longer to decide not to punish than to punish. And when the offence was of low moral outrage, they took longer in deciding to punish than not to punish. These differences in decision times can be explained by assuming that the conflict between dogmatism and forgiveness is particularly high when dogmatism and forgiveness also conflict with feelings of moral outrage, which, according to Experiments 1-3 (Chapter 3), are stronger in cases of high and low moral outrage. However, even though these differences in decision times can be related to conflicts between the dogmatism and the forgiveness triggered by Christianity, it is still not clear why highly religious participants are not overall slower than less-religious people. In fact, when comparing the decision times of highly religious people with those of the laypeople in Experiment 2 (see Figure 2), there are salient similarities. In both cases participants needed longer for no-punishment decisions in cases of high moral outrage, and for punishment decisions in cases of low moral outrage. The explanation that the participants in Experiment 2 were probably also religious is not plausible. The preliminary study of this experiment suggests that the Median value on the religiosity scale in a student's population lays at 2.35 (1 = *little religious*; 5 = *highly religious*). Another explanation could be instead that although highly religious people felt somewhat unsure about what to decide (as shown by the smaller certainty ratings), their ambivalence between dogmatism and forgiveness did not have any particular effect on decision times. Instead, it could be that less-religious people were particularly spontaneous when deciding to punish or not to punish offenders, without experiencing conflicts when deciding against feelings of moral outrage. This, however, conflicts with the results from Experiments 1 and 2 and requires further investigations and replications.

Also the punishment decisions on problems with control circumstances showed an interesting pattern of results. Although highly and less-religious participants did not differ in their severity ratings, they did differ in their punishment decisions. For low-moral outrage problems, less-religious people decided not to punish the offender although there was no real reason for not punishing. In contrast, highly religious people did not show this effect. This

difference can be explained by assuming that highly and less-religious people use different standards when making punishment decisions. When less-religious people perceive an offence as not severe they feel little or no moral outrage and decide not to punish the offender even though the offender has committed a legal offence. But when highly religious people perceive an offence as not severe and also feel little or no moral outrage, they nonetheless decide to punish the offender because an offence has been committed and must be punished. This explanation fits with the higher dogmatism attributed to highly religious people and is therefore relevant when trying to pin down the effect of moral outrage on defeasible reasoning. The fact that highly religious people decided to punish even though the severity ratings suggest that they were not morally outraged shows how norms of behavior imposed by one's religion can moderate the effect of moral outrage on punishment decisions. However, since this result was not replicated for problems with legally exculpatory circumstances, further studies are necessary.

4.2. Experiment 5: Culture

Another factor that may influence people's legal reasoning and withdrawal of conclusions is culture. Culture can shape people's legal reasoning by prescribing which behavior is desirable and which not. Most criminal offences such as manslaughter, theft, sexual abuse or bodily injury are prohibited by law and also condemned by people across cultures. However, sometimes there are behaviors which are prohibited by law, but not necessarily condemned equally across cultures. One of these is vigilantism. Vigilantism, or vigilante justice, refers usually to groups of people enforcing rules they believe will not be enforced by the legal system (Robinson & Darley, 2007). Yet, vigilantism can also be applied to refer to single individuals taking justice into their own hands (Haas et al., 2012). Taking justice into one's own hands is usually prohibited by law because the judiciary is the only entity allowed to decide if a behavior deserves punishment. Without the prohibition of vigilantism, the whole idea of a legal system would be obsolete. However, people's attitude towards vigilantism differs across cultures. Whereas in countries like the Netherlands support of vigilantism is low (see Haas et al., 2012), cases of vigilantism are more frequent in countries like Nigeria, Ghana, Indonesia, Bolivia, Venezuela and Peru (e.g., Benson, Fischer, & Thomas, 2008;

Harnischfreger, 2003; Huggins, 1991; Goldstein, 2003; Tankebe, 2009, Tyson, 2013; Onken, 2011). People from those countries often even sympathize with vigilantes and thank them for punishing criminals (e.g., Goldstein, 2003; Harnischfreger, 2003, Tankebe, 2009). Some researchers argue that these differences between countries in people's acceptance of vigilantism depend on how well-functioning people perceive their legal systems to be. The less people feel they can trust in their legal system and their police, the more they sympathize with vigilantes (e.g., Adinkrah, 2005; Benesh & Howell, 2001; Black, 1983; Cook, 2006; Goldstein, 2003; Haas, de Keijser, & Bruinsma, 2014; Harnischfeger, 2003; Tankebe, 2009). Consequently, when people perceive their legal system as legitimate, they show a greater compliance with the system's rules (e.g., Tyler, 2006; Tyler & Huo, 2002; see also Mishler & Rose, 2001).

Experiment 5 aims to investigate laypeople's defeasible reasoning in law by making use of the cross-cultural differences that exist in peoples acceptance of vigilantism. It is assumed that people from countries with a higher acceptance of vigilantism should evaluate vigilantism differently than people from countries with less acceptance of vigilantism. Imagine for instance you are confronted with a case of manslaughter. According to the findings of Chapter 3, people should conclude that the offender should be punished. But what happens if additional information explains that the offender decided to kill the victim because the "victim" sexually abused the offender's child? According to the literature on vigilantism people with positive attitudes towards taking justice into one's own hands should be more willing to excuse this offender. After all, by killing the "victim" the offender only fulfilled the desire for punishment that moral outrage evokes, and consequently reaffirmed the violated moral rule (see Cook, 2006; Haas et al., 2012; Skitka & Houston, 2001; Vidmar & Miller, 1980). That is, people who favor vigilantism should defeat the conclusion of punishing offenders if they become aware that an offence was committed in an attempt to do justice. People against vigilantism, instead, should not excuse the killing.

For this cross-cultural study I decided to test people from Germany and Peru, because, on the one hand, essential parts of the Peruvian penal code – specifically the ones related to the definitions and regulations of culpability – historically originate from German penal code (Código Penal, 1991/2014, section *Exposición de Motivos*). This makes the systems comparable. On the other hand, the two countries significantly differ in their criminality (Global Study on Homicide - United Nation Office on Drugs and Crime, 2013) and their citizens' trust in the courts and police. The World Values Study (2014), for instance, shows that whereas 71.3% of Germans declare to have a great deal or quite a lot of confidence in

courts, only 17.4% of Peruvians does so. Similarly, whereas 81.7% of Germans declare to have a great deal or quite a lot confidence in the police, only 27.9% of Peruvians do so (World Values Study Association, 2014, World Values Survey Wave 6: 2010-2014).

To test people's consideration of vigilantism as a defeater in legal reasoning I created legal conditional inference tasks which had either no information about circumstances (i.e., baseline), or additional information describing exculpatory circumstances (as hitherto) or circumstances that put the offender as a vigilante ("Oscar killed the victim because the victim had killed the offender's wife before"). Participants from Germany and Peru were asked to decide whether the offender should be punished. It was hypothesized that participants from both countries should not differ in their conclusions about legal conditionals without circumstance information. That is, according to the results of Experiments 1-3 (Chapter 3) they should favor punishment as long as the offence is morally outraging. Due to the similarities between the two penal codes, participants from Germany and Peru should also not differ significantly when deciding about legally relevant exculpatory circumstances. Their acceptance should depend on how morally outrageous the offence was (see Experiments 1-3). However, differences are expected when deciding about circumstances framing the offence as an act of vigilantism. Germans should be reluctant to excuse an offence if it happened as an act of vigilantism. Peruvians instead should be more open to conclude that an offender should not be punished if the offence was committed only to make justice. In addition, considering that moral outrage is related to peoples acceptance of vigilantism (see Cook, 2006; Haas et al., 2012; Skitka & Houston, 2001; Vidmar & Miller, 1980), Peruvians' acceptance of vigilantism as a defeater should be highest for high morally outraging offences.

Hypothesis 1: People from Peru should be more prone to accept vigilante circumstances as defeaters and reasons for voiding punishment than people from Germany. Especially, when the prior offence (to which the vigilante reacts) was morally outraging.

Hypothesis 2: In light of no additional information about circumstances, or when the circumstance information is exculpatory, people from Peru and Germany should not differ in their punishment decisions. Their decisions should only depend on how morally outraging the offence described in the legal conditional is.

4.2.1. Methods

4.2.1.1. Participants

The experiment was conducted online (see Section 4.1.1.2). In total, 191 participants completed the online experiment. However, because the target sample included only participants between 20 and 40 years old, without scholar legal knowledge, and living in their corresponding countries, all participants who did not fulfill these requirements were excluded from the analysis. The final sample consisted of 54 Peruvians (27 female) and 108 Germans (81 female). The mean age of Peruvians was 26.96 years old (*SD* = 4.23; Median = 26), the mean age of the Germans was 24.48 years old (*SD* = 3.53; Median = 24). Participants from both countries had similar high levels of education. All German participants finished high school, all but one studied at a university, and 39.9% indicated having some kind of university degree. All of the Peruvian participants also finished high school and 92.5% were studying at a university or an institute, or already had a university degree. Three participants gave no educational specification, but indicated they worked as a commercial pilot, flight dispatcher, and audiovisual communicator.

4.2.1.2. Material, Procedure, and Design

The experiment was programmed and conducted online on SoSci Survey (Leiner, 2014). The link was administered via a university database and social networks.

The experiment consisted of two parts. The first part was designed to measure people's general attitudes towards the offence used in this experiment. For this, participants were presented with three legal conditional rules without any information about circumstances. The problems consisted of (a) the legal conditional rule, (b) the fact that somebody committed an offence, and (c) the question about the conclusion. As offences I selected manslaughter ("If a person kills another human, then the person should be punished for manslaughter"), bodily injury ("If a person physically maltreats another human, then the person should be punished for bodily injury"), and defamation (in German "Üble Nachrede"– "If a person asserts denigrating facts about another person, then the person should be punished for defamation"). A online pilot study (*N* = 568) showed that these offences are perceived differently with respect to moral outrage, with manslaughter evoking the highest and

defamation the lowest degree of moral outrage. All information (rule, fact, and question about the conclusion) were presented at once, and the participants had to answer with a *yes* or *no* key whether the offender should be punished (an example is given in the upper part of Table 11). The presentation order was randomized. After having solved all problems, participants were again presented with each of the offences and had to rate their severity on a 1 (*not at all severe*) – 7 (*very severe*) Likert scale. The severity ratings were added to the procedure to measure how the participants actually perceived the offences and whether participants from both countries perceived the offences as equally wrong and deserving of punishment.

The second part of the experiment was designed to test people's perception of vigilantism as a defeater. For this, I created defeasible conditional inferences tasks like in Experiments 1, 2, and 4. Participants received the same three legal conditionals from the first part (manslaughter, bodily injury, and defamation), but this time offences were paired with additional information about circumstances. The circumstance either described 1) the "victim" as the offender in a preceding crime and the offender taking justice into their own hands (henceforth: vigilante circumstances), or it described 2) legally exculpatory circumstances (henceforth: exculpatory circumstances). Each of the three legal conditionals (manslaughter, bodily injury, and defamation) was combined with two vigilante circumstances and two exculpatory circumstances, making a total of twelve problems. The problems consisted thus of (a) the legal conditional rule, (b) the fact that somebody committed an offence, (c) the vigilante or exculpatory circumstance information, and (d) the question about the conclusion.

The vigilante circumstances described two different eye-for-an eye scenarios. In one scenario, the offender is doing to the "victim" the same as the "victim" did to his wife before (i.e., killing the victim after the victim had killed the offender's wife; physically maltreating the victim after the victim had physically maltreated the offender's wife; and defaming the victim after the victim had defamed the offender's wife). In the other scenario, the "offender" responded with a different but similarly severe offence (i.e., killing the victim after the victim had sexually abused the offender's child; physically maltreating the victim after the victim had threatened the offender's child with harm; and defaming the victim after the victim had pinched the offender's girlfriend). These scenarios were used to make sure that the severity and level of moral outrage in the vigilant act and the preceding offence were comparable. The legally exculpatory circumstances included mistakes of law or situations of necessity. They served as control problems to which participants from Peru and Germany were expected to respond similarly. Examples of problems with vigilante and exculpatory circumstances can be found in Table 11. Both offender and victim were always male. The names of offender and

victim exist in both countries (sometimes in slightly adapted forms, e.g. Georg vs. Jorge) and were varied randomly between the offences. All premises were presented at once and participants had to decide whether they accepted the conclusion or not by pressing either a yes or no key. The problems were presented randomly. After the conditional inference task participants were again asked to rate the severity of the offences on a 1 (*not at all severe*) – 7 (*very severe*) Likert scale, but this time the offences were presented in combination with the corresponding circumstance information (e.g., "Oscar physically maltreated Daniel because Daniel threatened Oscar's child with harm").

Table 11

Exemplary problems used in the second part of Experiment 5 illustrated with the example of bodily injury

	Example
No circumstance	If a person physically maltreats another human, then the person should be punished for bodily injury. A person physically maltreated another human. *Should the person be punished for bodily injury?*
Vigilante circumstances	
Eye-for-an-eye	If a person physically maltreats another human, then the person should be punished for bodily injury. Oscar physically maltreated Daniel. Oscar did this because his wife had been physically maltreated by Daniel. *Should Oscar be punished for bodily injury?*
Comparable severity	If a person physically maltreats another human, then the person should be punished for bodily injury. Oscar physically maltreated Daniel. Oscar did this because his child had been threatened with harm by Daniel. *Should Oscar be punished for bodily injury?*

Exculpatory circumstances	
Mistakes of law	If a person physically maltreats another human, then the person should be punished for bodily injury. Oscar physically maltreated Daniel. Oscar believed erroneously that Bert tried to attack him *Should Oscar be punished for bodily injury?*
Necessity brought about by coercion	If a person physically maltreats another human, then the person should be punished for bodily injury. Oscar physically maltreated Daniel. Under threats to his life and physical integrity, Oscar was coerced into committing the crime. *Should Oscar be punished for bodily injury?*

Participants had no time limits and were instructed to decide using their own sense of justice. Peruvians were tested in Spanish and Germans in German. After the experiment they had the opportunity to take part in a lottery for a gift card.

4.2.2. Results

4.2.2.1. No Circumstance Information

First the severity ratings and the mean number of no-punishment decisions (in percent) for problems without any circumstance information were analyzed. This served to determine if participants from both countries had in general similar attitudes to the offences used in the reasoning problems. This is an important prerequisite for this experiment. I also added gender as an additional factor to control for the different distribution of females and males in both samples. This results in two separate 2 (country: Peru vs. Germany) x 3 (offence: manslaughter vs. bodily injury vs. defamation) x 2 (gender: female vs. male) ANOVAs (one for the percentage of no-punishment decisions and one for severity ratings). The results are presented in Table 12. The ANOVA for the percentage of no-punishment decisions without any circumstance information did not show differences between participants from Peru and Germany. In fact, all participants agreed with the legal rule and almost never decided not to

punish the offender. Accordingly, no main effects of country, gender, and offence, and no interactions were found (all Fs ≤ 3.08, ps ≥ .053).

The ANOVA on severity ratings showed a main effect of offence, $F(1.52, 240.68) = 222.95$, $p < .001$, $\eta_p^2 = .585$, and an interaction between offence and country, $F(1.52, 240.68) = 4.35$, $p = .023$, $\eta_p^2 = .027$. Manslaughter was always perceived as more severe than bodily injury ($M = 6.90$; $SD = 0.40$ and $M = 6.30$; $SD = 0.88$, respectively), $t(161) = 9.18$, $p < .001$, $d = 0.82$, and bodily injury as more severe than defamation ($M = 4.67$; $SD = 1.37$), $t(161) = 20.13$, $p < .001$, $d = 2.17$. Yet, German participants perceived manslaughter and bodily injury as slightly more severe than Peruvians ($t(61.74) = 2.48$, $p = .016$, $d = 0.53$; $t(77.30) = 2.54$, $p = .013$, $d = 0.48$, respectively). However, these differences were not high (see Table 12). No main effect of country was found, $F(1, 158) = 0.63$, $p = .430$, $\eta_p^2 = .004$. A main effect of gender also showed that female participants ($M = 6.10$; $SD = 0.55$) perceived offences somewhat more severe than male participants ($M = 5.65$; $SD = 0.76$), $F(1, 158) = 19.43$, $p < .001$, $\eta_p^2 = .109$. All other effects were not significant (all Fs ≤ 3.69, ps ≥ .056).

Table 12

Percentage of no-punishment decisions and severity ratings of participants from Germany and Peru for problems without circumstance information

	Offence		
	Manslaughter	Bodily injury	Defamation
No-punishment decisions (%)			
German participants	5.56 (23.01)	1.85 (13.54)	7.41 (26.31)
Peruvian participants	11.11 (31.72)	5.56 (23.12)	14.81 (35.86)
Severity ratings			
Germany participants	6.96 (0.23)	6.44 (0.73)	4.62 (1.32)
Peruvian participants	6.76 (0.58)	6.02 (1.09)	4.78 (1.48)

Note. Severity ratings range from 1 (*not severe at all*) to 7 (*very severe*). The numbers in brackets indicate the corresponding standard deviations.

4.2.2.2. Vigilante and Exculpatory Circumstances

In the main part of the analyses, I compared the mean frequency of punishment decisions (in percent) and the severity ratings as a function of the additional vigilante and exculpatory circumstance information. This was done with two separate 2 (country: Peru vs. Germany) x 2 (circumstance: vigilante vs. exculpatory) x 3 (offence: manslaughter vs. bodily injury vs. defamation) x 2 (gender: female vs. male) mixed ANOVAs; one for punishment decisions (Figure 6) and one for severity ratings (Table 13).

The ANOVA for the percentage of no-punishment decisions revealed main effects of offence and of circumstance (both Fs $\geq$ 6.19, ps $\leq$ .003), two-way interactions between country and circumstance and between offence and circumstance (both Fs > 7.04, ps $\leq$ .007), and a three-way interaction between offence, country, and circumstance, $F(2, 316) = 3.59$, $p = .029$, $\eta_p^2 = 0.022$. Also a main effect of gender was found showing that the percentage of no-punishment decisions was higher for male participants ($M = 37.65\%$; $SD = 22.88$) than for female participants ($M = 28.40\%$; $SD = 17.39$), $F(1, 158) = 5.71$, $p = .018$, $\eta_p^2 = 0.035$. All other effects were not significant (all Fs $\leq$ 2.90, ps $\geq$ .059). Since the three way interaction showed that the effect of country and offences depended on which kind of circumstance information was presented, I continued the analyses by conducting two separate 2 (country: Peru vs. Germany) x 3 (offence: manslaughter vs. bodily injury vs. defamation) ANOVAs for problems with legally exculpatory circumstances and for problems with vigilante circumstances. In these analyses, the ANOVA for problems with legally exculpatory circumstances only revealed a main effect of offence, $F(2, 320) = 8.70$, $p < .001$, $\eta_p^2 = .052$. The percentage of no-punishment decisions was lower for manslaughter ($M = 33.64\%$; $SD = 33.36$) compared to bodily injury ($M = 46.30\%$; $SD = 37.20$), $t(161) = 4.39$, $p < .001$, $d = 0.36$, and compared to defamation ($M = 47.84\%$; $SD = 32.18$), $t(161) = 4.71$, $p < .001$, $d = 0.43$. No-punishment decisions for bodily injury and defamation did not differ, $t(161) = 0.50$, $p = 0.62$, $d = 0.04$. All other effects were not significant (all Fs $\leq$ 2.36, ps $\geq$ .096).

In contrast, the ANOVA for problems with vigilante circumstances showed differences between countries. This main effect of country shows that the percentage of no-punishment decisions for offences with vigilante circumstances was higher for participants from Peru than for participants from Germany ($M = 30.56\%$; $SD = 27.23$ and $M = 15.28\%$; $SD = 23.80$, respectively), $F(1, 160) = 13.45$, $p < .001$, $\eta_p^2 = .078$. Yet, an interaction between country and offence shows that this effect depended on how morally outrageous the prior

crime was, $F(1.89, 302.9) = 6.29$, $p = .003$, $\eta_p^2 = .038$. As can be seen in Figure 6, the percentage of no-punishment decisions for offences with vigilante circumstances was only higher for participants from Peru than from Germany for bodily injury, $t(86.96) = 4.41$, $p < .001$, $d = 0.79$, but not for defamation, $t(96.74) = 0.92$, $p = .360$, $d = 0.16$. Peruvian participants also condoned vigilantism more often than German participants for manslaughter, $t(83.84) = 2.21$, $p = .030$, $d = 0.40$, but the effect did not reach the Bonferroni adjusted alpha level of 0.016. Because of this interaction, the main effect of offence was significant, $F(1.89, 302.9) = 5.18$, $p = .007$, $\eta_p^2 = .031$.

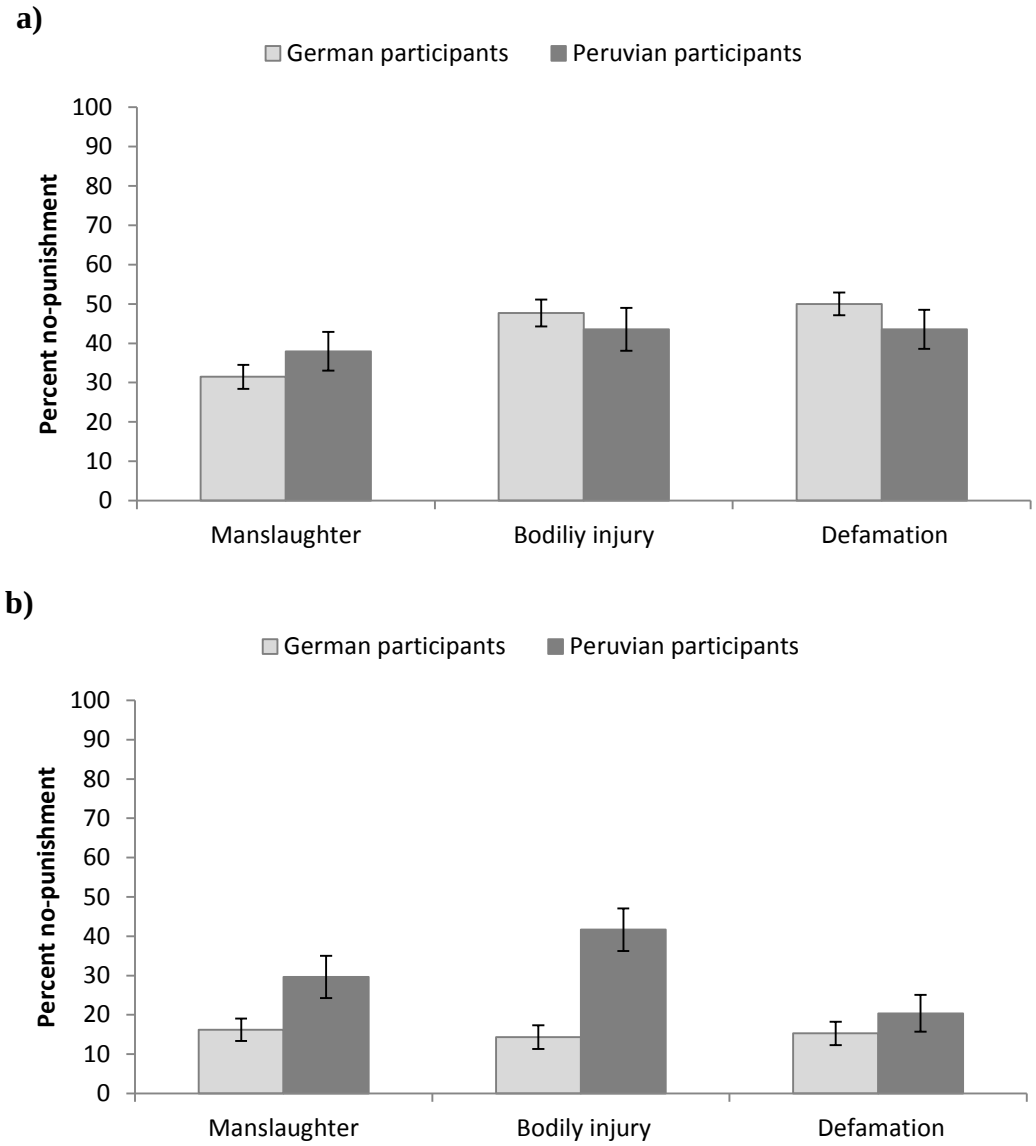

Figure 6. Percentages of no-punishment decisions made by participants from Germany and Peru for problems with a) legally exculpatory and b) vigilante circumstances.

Results for the severity ratings are presented in Table 13. The analysis of the severity ratings revealed a main effect of offence, $F(1.45, 229.53) = 69.19$, $p < .001$, $\eta_p^2 = 0.305$, two-way interactions between country and offence and between offence and circumstance (both Fs ≥ 5.08, ps $\leq .008$), and a three-way interaction between offence, country, and circumstance, $F(1.84, 291.21) = 4.12$, $p = .02$, $\eta_p^2 = .025$. Additionally, a main effect of gender showed that female ($M = 4.82$; $SD = 1.08$) participants perceived offences as more severe than male participants ($M = 4.31$; $SD = 1.25$), $F(1, 158) = 5.21$, $p = .024$, $\eta_p^2 = 0.024$. All other effects were not significant, including the interaction between country and offence (all Fs ≤ 3.79, ps $\geq .053$). As for the analysis of no-punishment decisions, the tree way interaction allowed to continue the analyses by conducting two separate 2 (country: Peru vs. Germany) x 3 (offence: manslaughter vs. bodily injury vs. defamation) ANOVAs; one for problems with legally exculpatory circumstances and one for problems with vigilante circumstances. Both ANOVAs – the one for exculpatory and the one for vigilante circumstances – revealed a main effect of offence ($F(1.75, 280.40) = 95.25$, $p < .001$, $\eta_p^2 = .373$; and $F(1.40, 224.03) = 32.02$, $p < .001$, $\eta_p^2 = .167$; respectively): manslaughter was perceived as more severe than bodily injury and bodily injury as more severe than defamation (all ts ≥ 7.21, ps ≤ 0.001). In both ANOVAs interactions between offence and country were found ($F(1.75, 280.40) = 5.32$, $p = .01$, $\eta_p^2 = .031$; and $F(1.40, 224.03) = 11.34$, $p < .001$, $\eta_p^2 = .066$; respectively). However, only for problems with vigilante circumstances the post-hoc t-tests reached the Bonferroni adjusted alpha of 0.016: when offences were paired with vigilante circumstances, participants from Germany perceived manslaughter and bodily injury as more severe than participants from Peru ($t(81.97) = 2.79$, $p = .007$, $d = 0.51$; and $t(87.31) = 3.80$, $p < .001$, $d = 0.68$; respectively). For defamation there were no differences, $t(160) = 0.80$, $p = .427$, $d = 0.13$. As a consequence, the main effect of country was significant for problems with vigilante circumstances, $F(1, 160) = 7.64$, $p = .006$, $\eta_p^2 = .046$, but not for problems with exculpatory circumstances, $F(1, 160) = 1.85$, $p = .176$, $\eta_p^2 = .011$.

Table 13

Severity ratings of participants from Germany and Peru for problems with legally exculpatory and vigilante circumstances

	Offence		
	Manslaughter	Bodily injury	Defamation
Legally exculpatory circumstances			
German participants	5.68 (1.23)	4.68 (1.43)	3.77 (1.45)
Peruvian participants	5.08 (1.76)	4.32 (1.52)	3.88 (1.58)
Vigilante circumstances			
German participants	5.62 (1.1.49)	5.06 (1.41)	3.94 (1.57)
Peruvian participants	4.75 (2.04)	4.01 (1.78)	4.15 (1.46)

Note. Severity ratings range from 1 (*not severe at all*) to 7 (*very severe*). The numbers in brackets indicate the corresponding standard deviations.

4.2.3. Discussion

Experiment 5 aimed to investigate cross-cultural differences in legal reasoning. It was expected that the consideration of vigilantism as a defeater in legal reasoning varies between people from Peru and Germany. The results show that all participants, regardless of their country of origin, agreed that offences like manslaughter, bodily injury, and defamation should be punished. This supports Hypothesis 2. In addition, they agreed that in light of legally exculpatory circumstances it is acceptable not to punish offenders (with slight differences depending on how moral outraging the offence is). However, Peruvians and Germans differed in their punishment decisions for problems with vigilante circumstances. As expected by Hypothesis 1, participants from Peru were more inclined not to punish an offender if the circumstances described the offence as an act of vigilantism. Germans did not show this effect. That is, only Peruvians withdrew the conclusion of punishing the offender in light of information framing the offence as an act of vigilantism. Nevertheless, even for Peruvians the prior offence to which the vigilante reacted had to be severe enough to engage in defeasible reasoning. If the vigilante's act was only committed to avenge low morally outraging offences like defamation, participants from both countries decided to punish the

offender. This finding suggests that moral outrage moderates the acceptance of vigilantism as a defeater.

An unexpected finding of the study, however, is that the highest acceptance of vigilantism was found for bodily injury and not for manslaughter. One explanation could be that the acceptance of an-eye-for-an-eye retribution is limited to less extreme cases of vigilantism. It is possible that paying for a life with another life is too extreme to be excused. In such cases the vigilante's act probably also evokes strong feelings of moral outrage so that condoning the act becomes unacceptable. Such conflicting effects of strong feelings of moral outrage against both the prior offender and the vigilante have been described before as possible reactions to vigilantism (see Cook, 2006; Haas et al., 2012). In this context, also the gender differences in the attitudes towards vigilantes are notorious. Interestingly, men perceived offences as less severe than women: Men decided more often not to punish the offender. This was the case for all men – both from Peru and Germany – and across all kinds of circumstance information – no circumstance, exculpatory circumstances, and vigilante circumstances. This finding agrees with previous studies that show men to be more supportive of vigilantism than women (e.g., Briceño-León, Camardiel, & Avila, 2006).

One open question that still remains is why exactly Peruvians and Germans differed in their attitudes towards vigilantism. According to the literature on vigilantism, it could be that Peruvians and Germans differ in their trust in the legal system. As shown by the World Values Study less than one fourth of Peruvians trust their legal system, whereas the majority of Germans do. However, in the present experiment I did not measure people's trust in the legal system directly. It may be possible that the academic population tested in this study differs from the general population of the World Values study. In an attempt to have a more reliable measure on people's trust in the legal system and the relationship to vigilantism, I conducted an additional online study via LimeSurvey™ with a sample from the same population of this experiment. In this study, participants (N = 39 Germans; N = 24 Peruvians) had to answer several questions about their experiences and attitudes on legal matters. I will report here only the results on trust and vigilantism. The single questions and descriptive data can be found in Table 14. As expected, Peruvians scored significantly lower on the trust-related questions than Germans, $t(61) = 11.78$, $p < .001$, $d = 3.06$, and had accordingly more positive attitudes towards vigilantism than Germans, $t(61) = 3.10$, $p = .001$, $d = 1.07$. In fact, measures of trust and vigilantism correlated significantly, $r = -.622$, $p < .001$. Especially this last correlation suggests that the differences found between Peruvians and Germans in the

consideration of vigilantism as a defeater are related to differences in their trust in the legal system.

Table 14

Results of the follow-up study on Peruvians' and Germans' trust in the legal system and attitude towards vigilantism (Experiment 5)

	***M* (*SD*)**			
	PE	**DE**	***p***	***d***
Trust towards the legal system				
(1 = *yes a lot*; 7 = *not at all*)				
- Are you satisfied with your LS?	6.2 (0.9)	2.8 (1.0)		
- Do you trust in your LS?	6.2 (0.7)	2.7 (1.0)		
- Does your LS persecute some offences too little?[a]	5.9 (1.5)	4.5 (1.6)		
Mean	6.1 (0.9)	3.3 (0.9)	<.001	3.06
Attitude towards vigilantism				
(1 = *yes, a lot*; 7 = *not at all*)				
- Do you think vigilantism is acceptable?	4.6 (2.1)	5.9 (1.4)		
- Do you think there are circumstances were vigilantism should be allowed?	3.8 (2.3)	6.1 (1.3)		
- Do you think there are offences where vigilantism should be allowed?	5.1 (2.2)	6.4 (1.3)		
- Should two offenders get the same punishment, if the victim of one of them is a criminal? [a]	2.6 (2.5)	1.8 (1.7)		
Mean	4.7(1.7)	6.1 (1.1)	.001	1.07

Note. Comparisons between Peruvians and Germans were done by comparing the means of each category with *t*-tests for independent samples. PE = Peruvians; DE = Germans; LS = Legal system.

[a] For better comparison, results for this item were recoded to its reverse polarity

To further test the effect of trust in the legal system on vigilantism and legal reasoning, it is nonetheless necessary to continue testing people from other countries whose inhabitants also differ in their attitudes towards their legal system. In a first attempt to do this, I am currently testing people from Russia. Russian participants should answer similarly to Peruvians, because according to the World Values Study (World Values Study Association, 2014, World Values Survey Wave 6: 2010-2014) only 32.5% of Russians have a great deal or quite a lot of trust in the courts and 31.7% in the police. First results support this hypothesis: Russian participants (N = 42) decided not to punish vigilantes in 28.6% of the cases of manslaughter, in 44% of the cases of bodily injury, and in 19% of the cases of defamation.

4.3. Summary and Implications of Chapter 4

In this second block of experiments it was investigated whether differences in a person's attitudes about offenders and offences affects legal conditional reasoning. It was assumed that if cultural or ideological norms of behavior support or condemn excusing offences, then their willingness to withdraw conclusions from legal rules should be enhanced or inhibited, respectively. The results of this Chapter are widely in accordance with these assumptions. In Experiment 4 differences between highly and less religious people were found in decision times and in punishment decisions for crime irrelevant circumstance information. The differences in decision times were difficult to interpret, but the tendency of highly religious participants to punish offenders even for little morally outraging offences indicates that the dogmatism usually linked to high religiosity affects legal reasoning. Especially Experiment 5 showed how different attitudes towards offences have an effect on punishment conclusions. While Peruvians excused offences in light of vigilante circumstances, Germans did not. That is, people's defeasible reasoning in light of vigilante circumstances varied between countries.

However, one question which requires further discussion is whether Peruvians and Germans, or highly and little religious people, differed only in their punishment decisions or also in their feelings of moral outrage. Robinson and Darley (2007) argue that people's negative reactions towards morally wrong offences and the desire to punish these offenders is universal and intuitive. Accordingly also the severity ratings of Peruvians and Germans, and highly and less religious people, did not differ much. But when there are no big differences in

moral outrage, how can one explain the differing punishment decisions? Robinson and Darley answer by saying that what is universal is people's wish to punish offenders, but which particular punishment one thinks is appropriate varies between people and cultures. In other words, the perception that, for instance, manslaughter is worse than theft is homogenous across countries. Robinson and Darley talk in this respect about a correlation of Kendalls W = .95 (see Robinson & Kurzban, 2007). Yet, whether manslaughter should be punished with death sentence, life prison or 15 years of jail differs across people. In this sense the participants in this Chapter were probably similarly morally outraged by the offences (as suggested by the severity ratings), but decided differently when it came to the implications of these feelings of moral outrage. For instance, while both Peruvians and Germans perceived bodily injury as a severe offence which requires punishment, only Peruvians concluded that a vigilante can deliver this punishment. This chapter's findings can thus be related to the literature on the role of emotions on behavior. Emotions evoked by certain events do not necessarily result in a specific behavior (Baumeister, DeWall, Vohs, & Alquist, 2010; Schwarz & Clore, 2007). Like fear can lead to fight or flight responses, feelings of moral outrage towards offences can lead to desire for different types of punishments.

Chapter 5: Linguistic Factors in Legal Conditional Reasoning

So far, the experiments I presented have shown that laypeople do not readily accept exculpatory circumstances as defeaters. But does this finding not conflict with the vast literature on conditional reasoning showing that people usually do engage in defeasible reasoning? As explained in Chapter 1, in everyday situations people do consider defeaters and exceptions during reasoning, sometimes even when instructed to ignore background knowledge and to reason solely on the basis of the premises (see Vadeboncoeur & Markovits, 1999). So why is it that the participants in this thesis have difficulties to withdraw previous conclusions when given additional information about exculpatory circumstances for an offence?

One possible reason was already considered in the hypotheses of this thesis, namely the importance of punishing offenders in our society. As explained in Chapter 1 and 2, people experience negative feelings of moral outrage when faced with offences, resulting in a desire of punishment (Alter et al., 2007; Carlsmith & Darley, 2008; Carlsmith et al., 2002; Darley, 2009; Darley & Pittman, 2003; also see Fehr & Gächter, 2002; Greene et al., 2001; Haidt, 2001; 2007; Keller, et al., 2010).

Another reason why participants have difficulties in accepting defeaters in legal reasoning might have something to do with the phrasing of the legal conditional. On the one hand, all legal conditionals were phrased deontically, asking about what *should* happen to an offender. On the other hand, legal conditionals were phrased as universal rules, stating implicitly that every time the antecedent is the case, the consequent follows. In the following I will present three experiments that consider the role of phrasing in legal conditional reasoning. More precisely, it is investigated whether the effect of one's own sense of justice on legal reasoning can be modulated by changing the way in which the legal conditional is phrased.

5.1. Experiment 6: The Modal Auxiliary

One reason why reasoners have ignored so far potential exculpatory circumstances might be the modal auxiliary used in the legal conditional. In the previous experiments legal conditionals were phrased with the modal "should", and not with "will" as it is the case in most of the literature on conditional reasoning (Kilpatrick, Manktelow, & Over, 2007). In legal theory, modals such as "should" or "ought" stand for the normative nature of legal rules (Bäcker, 2009; 2010). However, when presented in a MP inference, the modal "should" can have additional implications, as can be seen in the following example:

If a person kills another human, then the person should be punished for manslaughter.
A person killed another human.

Should the person be punished for manslaughter?

The "should" in the conclusion can be understood as asking for what should happen according to the deontic principles of this rule (see deontic possibilities; Johnson-Laird & Byrne, 2002; Quelhas & Byrne, 2003), or, one step further, according to the reasoner's own deontic standards. Participants might ignore exceptions because they answer in line with what is correct according to their own sense of justice. As already explained in Section 1.1.1.2.2, Over et al. (2004) argued that for deontic conditionals, the acceptance of conditional rules depends on the preference for the different outcomes such a rule can have. The more a reasoner prefers the outcome '*p* and *q*' (in our case: committing an offence and being punished for that) over '*p* and ¬*q*' (in our case: committing an offence and not being punished for that), the more the rule will be accepted. In the case of deontic legal conditionals, I expect the preference of '*p* and *q*' over '*p* and ¬*q*' to be correlated with moral outrage: the more morally outraged a reasoner is by the offence in the legal conditional, the more (s)he will prefer the outcome '*p* and *q*' (i.e., offence and punishment) over '*p* and ¬*q*' (i.e., offence and no punishment), and the more she or he will conclude that the offender should be punished.

But what happens if instead of "should", the legal conditional is phrased with the modal "will"? Consider following example:

If a person kills another human, then the person will be punished for manslaughter.
A person killed another human.

Will the person be punished for manslaughter?

Different from "should", the modal "will" in the conclusion suggests that the inference is asking about what happens in the real world, i.e., what is factually the case (see factual possibilities, Johnson-Laird & Byrne, 2002; Quelhas & Byrne, 2003). Therefore, when thinking about what happens in real world, the own sense of justice should lose importance, and the perceived frequency of exceptions should gain importance. The more frequently participants perceive exceptions to occur for a given offence, the less they will think *q* happens given *p*, and thus the less the rule will be accepted. In this case, when asked whether an offender *will* be punished, a reasoner might conclude that the offender will not be punished even though the offence is severe and morally outrageous.

The fact that different modals can have different implications – either by their meaning per se or by the context in which they are uttered – is known from linguistics (e.g., Groefsema, 1995) and has also received some psychological support (e.g., Johnson-Laird, 1978). Bell and Johnson-Laird (1998), for instance, showed that depending on which modal is used ("can" vs. "must"), inferences are drawn differently fast: when asked about what *can* be the case participants take longer to answer affirmatively than to answer negatively, but when asked about what *must* be the case it is the other way around. Further, Ferguson and Sanford (2008) showed that modals in counterfactual conditionals can affect interpretation of subsequent information. In fact, the difference between asking about deontic states or factual states is widely known in the psychological literature (e.g., Beller, 2008a; Bucciarelli & Johnson-Laird, 2005). However, most of the research about the difference between factual and deontic reasoning has been done with Wason's selection task (Wason, 1968; see also Beller, 2008a; Kilpatrick et al., 2007; Manktelow & Over, 1991; see Cheng & Holyoak, 1985; Cosmides, 1989) where participants have to choose cards in order to falsify or violate a rule (for details and problems on the Wason's selection task see Section 7.1.4 in the General Discussion). All in all however, the role of modals in the consideration of exceptions in inference tasks has not received much attention to date. To fill this gap, Experiment 6 aims to investigate the effect of modals on the consideration of exceptions in legal conditional reasoning. Because of the implications of asking about "should" or "will" I predict different answer patterns depending on how the legal conditional is phrased. If the legal conditional has the modal "should" (deontic legal conditionals), just like in the previous experiments

reported, I expect that people base their conclusions on their own sense of justice and feelings of moral outrage, ignoring the existence of exceptions. This should reflect in the participants' acceptance of a rule, given their preference of '*p* and *q*' over '*p* and ¬*q*' (see Over et al., 2004). However, if the legal conditional has the modal "will" (factual legal conditionals), people are expected to base their conclusions on their knowledge about what happens in real world, leaving their own sense of justice aside and considering exceptions. This consideration of exceptions should be reflected in the perceived frequency of cases of '*p* but ¬*q*' (cf. Geiger & Oberauer, 2007; see Section 1.1.1.1).

Hypothesis 1: When the legal conditional is phrased deontically with the modal *should*, then reasoners' conclusions should depend on their sense of justice and thus on their preference of *pq* over *p*¬*q*. The frequency of exceptions should be irrelevant.

Hypothesis 2: When the legal conditional is phrased factually with the modal *will*, then reasoners' conclusions should depend on the frequency of exceptions; that is, how often they think *p*¬*q* happens in the real world. Their own sense of justice should be irrelevant.

5.1.1. Methods

5.1.1.1. Participants

42 participants took part in the experiment. Two participants were excluded from the analysis because they afterwards reported to have specialized knowledge of logic or about the German penal code. The final sample consisted of 40 participants (20 female), with a mean age of 23.7 years (*SD* = 3.0). Half of the participants received deontic legal conditionals, the other half factual legal conditionals.

5.1.1.2. Material and Design

The materials were selected through a large (*N* = 298) preliminary study via SoSci Survey (Leiner, 2014). In the first half of the preliminary study participants' acceptance of legal rules was measured. Based on the proposal of Over et al. (2004), participants were confronted with

legal rules together with their outcomes '*p* and *q'* and '*p* and *not q'* (e.g., "A person kills another human and is punished for manslaughter" and "A person kills another human and is not punished for manslaughter"). In total I tested $N = 92$ legal rules but each participant received only 14 - 16 of them. The participants' task was to indicate their preference for each of these two possible outcomes on a scale from 1 (*completely in favor*) to 7 (*completely against*). A participant's acceptance of a rule was computed by dividing the ratings she or he gave for the outcome '*p* and *not q'* by the ratings for the outcome '*p* and *q'*. The higher this quotient, the more a participant accepted a rule.

In the second half of the preliminary study, participants were asked to rate the frequencies of exceptions for each rule. For this they were confronted once more with the same rules and asked to rate in how many of 100 cases they think *p* (i.e., the offence) occurs, but without the following *q* (i.e., the punishment; e.g., "A person kills another human, but the person is not punished for manslaughter. In how many of 100 cases do you think this is the case?").

The offences finally used for the conditionals in the actual experiment were selected depending on the ratings obtained in both parts of the preliminary study. I selected 8 offences: two with high acceptance rates and high frequency of exceptions, two with high acceptance and low frequency of exceptions, two with low acceptance and high frequency of exceptions, and two others with low acceptance and low frequency of exceptions. The assignment to each category was corroborated statistically. The list of the legal conditionals used in the experiment together with their ratings from the preliminary study can be found in Table 15.

Table 15

Legal conditionals used in Experiment 6, together with means (and standard deviations) for rule acceptance (RA) and frequency of exceptions (EX) from the preliminary studies.

Items	RA (SD)	EX (SD)
High RA, high EX (high-high)		
1. If a person downloads child pornography, then the person should/ will be punished for possession of child pornography.	6.3 (1.5)	56.6 (27.9)
2. If a person pollutes the soil and thereby harms animals, then the person should/ will be punished for soil pollution.	5.4 (2.2)	70.9 (24.9)

High RA, low EX (high-low)		
1. If a person abducts a human being in order to coerce a third person to commit an act, then the person should/ will be punished for hostage taking.	6.5 (1.3)	19.9 (17.4)
2. If a person kills another human, then the person should/ will be punished for manslaughter	5.8 (2.1)	22.9 (22.1)
Low RA, high EX (low-high)		
1. If a person downloads music from the internet without allowance, then the person should/ will be punished for breaching the copyright law.	1.9 (2.2)	69.9 (34.7)
2. If a person participates in an illegal game of chance, then the person should/ will be punished for illegal gambling.	2.1 (1.8)	62.2 (25.5)
Low RA, low EX (low-low)		
1. If a person kills another human because of the explicit and earnest request of the person killed, then the person should/ will be punished for homicide upon request.	2.0 (2.3)	14.8 (21.6)
2. If a shop-owner opens his/ her shop without allowance on a Sunday, then the person should/will be punished for breaching the Shop Closing Act.	1.9 (2.0)	26.5 (26.3)

For the experiment, each offence was phrased as a legal conditional and presented once in an MP and once in an MT inference, creating a total of 16 problems. The conclusion was phrased as a question. Thus the inference problems consisted of 1) a legal conditional rule, 2) the fact p or $\neg q$ (for MP and MT inferences, respectively), and 3) the question about the conclusion, asking whether *q should* or *will* follow (for MP inferences) or whether p is the case (for MT inferences). Contrary to the previous experiments, exculpatory circumstances were not presented explicitly as part of the inference task. This was not necessary because they were already tested implicitly by the preliminary study on the frequency of exceptions. Half of the participants got the problems with the modal "should" and the other half with "will". See Table 16 for an illustration. Thus the experiment followed a 2 (modal: should vs. will) x 2 (rule acceptance: high vs. low) x 2 (frequency of exceptions: high vs. low) x 2 (inference: MP vs. MT) mixed design. The modal was varied as a between subjects factor, all other factors were varied within individuals.

Table 16

Structure of the problems used in Experiment 6 illustrated by the legal conditional of manslaughter. R=Conditional rule; F=Fact; C=Conclusion.

	Modus Ponens		Modus Tollens	
Should	R:	If a person kills another human, then the person should be punished for manslaughter.	R:	If a person kills another human, then the person should be punished for manslaughter.
	F:	A person kills another human.	F:	A person should not be punished for manslaughter.
	C:	Should the person be punished for manslaughter?	C:	Did the person kill another human?
Will	R:	If a person kills another human, then the person will be punished for manslaughter.	R:	If a person kills another human, then the person will be punished for manslaughter.
	F:	A person kills another human.	F:	A person is not punished for manslaughter
	C:	Will the person be punished for manslaughter?	C:	Did the person kill another human?

I also created a generation and an evaluation task. The generation task served as an additional measure for the availability of exceptions. Participants were presented with the same eight offences from the inference task and were asked to generate reasons of why somebody who committed those offences *should*/ *will* not be punished ("A person kills another human, but this person should/ will not be punished for that"). The modal used in the generation task was the same the participants had received in the inference task. Following previous studies (see Cummins, 1995; Cummins et al., 1991; De Neys et al., 2003a; 2003b) participants had 1.5 minutes to write down as many reasons as they could think of. After 10 seconds of inactivity, the 1.5 minutes ended prematurely.

In the evaluation task participants were confronted once more with the same eight offences and asked to indicate on a 7-point-Likert scale how morally outraged they were by each (1= *no moral outrage*, 7= *high moral outrage*). Since feelings of moral outrage towards offences are an essential part of people's sense of justice (Carlsmith et al., 2002; Carlsmith & Darley, 2008; Darley, 2009; Darley & Pittman, 2003), I expected the moral outrage ratings to correlate with the acceptance ratings from the preliminary study.

5.1.1.3. Procedure

Participants were tested individually on a computer. The experiment was presented with Cedrus Superlab © 4.5 and was introduced as an experiment on reasoning in law. The participants were told that they will be confronted with general rules that are embedded in specific situations and that their task is to decide whether this rule *should* (for participants in the deontic legal conditionals condition) or *will* (for participants in the factual legal conditional condition) be applied in the given situation. Apart from the modal used for describing the task, instructions were kept constant across conditions. After one practice trial consisting of two items with the legal conditional for tax aversion (once as MP, once as MT), participants were left alone in the experimental room. The two premises were presented on subsequent screens and participants could move to the next screen by pressing the space bar. The conclusion was always phrased as a question and written in red. After reading this question, participants had to answer on a 7 point Likert scale ranging from *yes, very certainly* to *no, very certainly not.* The polarity of the scale was reversed for half of the participants. Between each problem participants had the opportunity to take a break. After the inference task, the two supplementary tasks were presented in random order. After the experimental session participants were asked in an open interview about their knowledge about logic and law.

5.1.2. Results

5.1.2.1. Inference Task

For the MP inferences "yes, very certainly" answers were scored with 0 points and "no, very certainly not" answers with 6 points. The ratings in-between were scored respectively with 1 to 5 points. For the MT inferences the opposite was done. "No, very certainly not" were scored with 0 points and "yes, very certainly" with 6 points. Again, the ratings in-between were scored respectively with 1 to 5 points. These scores were averaged separately for MP and MT inferences and indicate the degree of rejection of the logically valid conclusion. I will call this the "rejection rating". The higher the rejection rating, the less the MP or MT

inferences were accepted. The rejection ratings for MP and MT inferences can be found in Figure 7.

A 2 (modal: should vs. will) x 2 (rule acceptance: high vs. low) x 2 (frequency of exceptions: high vs. low) x 2 (inference: MP vs. MT) ANOVA on rejection ratings showed a main effect of frequency of exceptions, $F(1,38) = 14.06$, $p = .001$, $\eta_p^2 = .270$, a main effect of rule acceptance, $F(1, 38) = 23.05$, $p < .001$, $\eta_p^2 = .378$, and a main effect of rule, $F(1,38) = 21.69$, $p < .001$, $\eta_p^2 = .363$. However, the ANOVA also showed that the kind of inference (MP vs. MT) interacted significantly with rule acceptance, $F(1, 38) = 30.03$, $p < .001$, $\eta_p^2 = .441$, with the frequency of exceptions and the modal auxiliary, $F(1, 38) = 6.73$, $p = .013$, $\eta_p^2 = .15$, and marginally with the frequency of exceptions, $F(1, 38) = 3.08$, $p = .087$, $\eta_p^2 = .075$, and with the modal auxiliary, $F(1,38) = 3.128$, $p = .085$, $\eta_p^2 = .076$. Therefore, I decided to analyze the data from MP and MT inferences with two separate 2 (modal: should vs. will) x 2 (rule acceptance: high vs. low) x 2 (frequency of exceptions: high vs. low) ANOVAS. As can be seen in Figure 7, results for MP and MT indeed differed.

For MP inferences the ANOVA showed main effects for modal auxiliary, $F(1, 38) = 5.78$, $p = .021$, $\eta_p^2 = .132$, for rule acceptance, $F(1, 38) = 43.988$, $p < .001$, $\eta_p^2 = .537$, and for frequency of exceptions, $F(1, 38) = 20.021$, $p < .001$, $\eta_p^2 = .345$. These effects were explained in terms of the expected interactions between the modal auxiliary and rule acceptance, $F(1, 38) = 7.62$, $p = .009$, $\eta_p^2 = .167$, and between the modal auxiliary and the frequency of exceptions, $F(1, 38) = 12.51$, $p = .001$, $\eta_p^2 = .248$. On the one hand, the interaction between the modal auxiliary and frequency of exceptions shows that the frequency of exceptions had a strong effect on factual legal conditionals, but did not have any effect on deontic legal conditionals: whereas rejection ratings were higher for factual conditionals with high frequency of exceptions ($M = 2.34$; $SD = 1.43$) than for those with low ($M = 1.16$; $SD = 0.60$), $t(19) = 4.91$, $p < .001$, $d = .81$, rejection ratings for deontic conditionals with high ($M = 1.13$; $SD = 1.05$) and low frequency of exceptions ($M = 0.99$; $SD = 0.83$) did not differ, $t(19) = 0.49$, $p = .428$, $d = 0.14$ (Bonferroni adjusted alpha level: $\alpha = .025$). On the other hand, the interaction between modal auxiliary and rule acceptance shows that the effect of rule acceptance on rejection ratings was higher for deontic legal conditionals than for factual legal conditionals: although rejection ratings for conditionals with low rule acceptance were always higher than rejection ratings for conditionals with high rule acceptance, this effect was higher for deontic ($M = 1.88$; $SD = 1.48$ vs. $M = 0.24$; $SD = 0.43$; respectively), $t(19) = 5.58$, $p < .001$, $d = 1.24$, than for factual conditionals ($M = 2.09$; $SD = 1.13$ vs. $M = 1.41$; $SD = 0.95$;

respectively), $t(19) = 3.58$, $p = .002$, $d = 0.64$ (Bonferroni adjusted alpha level: $\alpha = .025$). All other effects were not significant (all $F \leq 1.6$, $p \geq .21$).

The ANOVA for the MT inferences did not show any significant effects at all (all $F \leq 1.60$, $p \geq .214$).

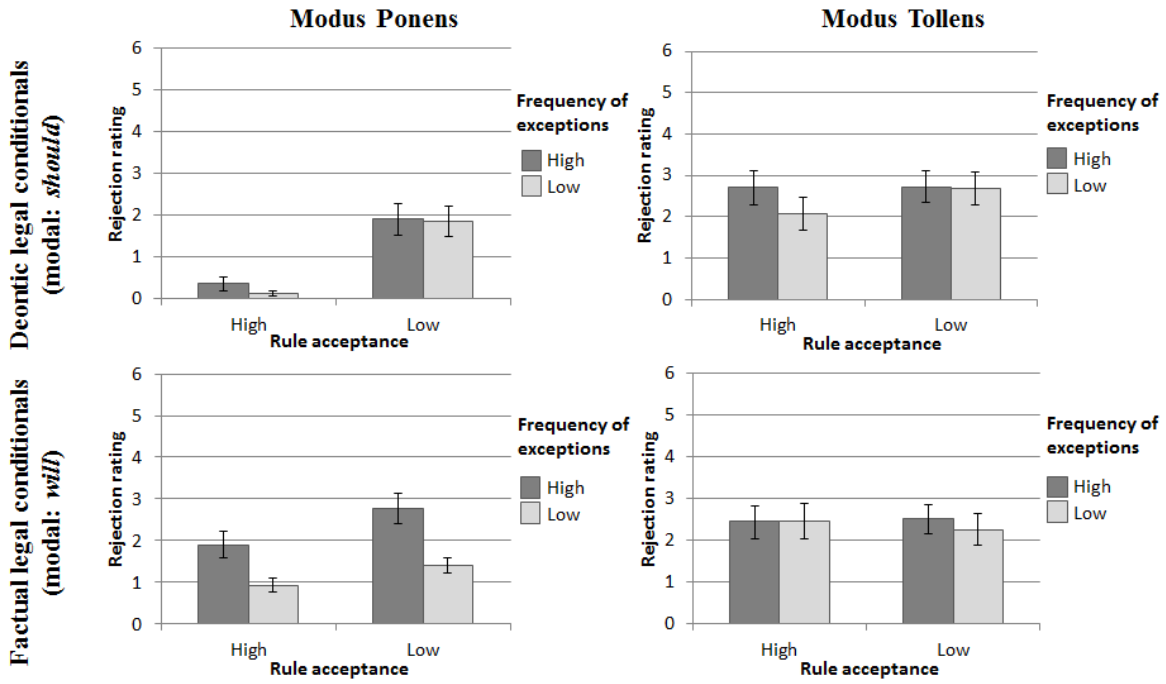

Figure 7. Rejection ratings (0 - 6) for Modus Ponens and Modus Tollens inferences for deontic and factual legal conditionals in Experiment 6. Error bars represent standard errors.

5.1.2.2. Generation Task

Two raters counted independently the amount of exceptions generated by participants and also the quality of each exception ($\rho = .98$ for the amount of exceptions in general, $\rho = .98$ for amount of factual, and $\rho = .92$ for the amount of deontic exceptions). Exceptions describing cases of malpractice (e.g., not being caught, not being accused, influences, etc.) were counted as factual exceptions. Exceptions describing cases were an offender should not be punished as a matter of principle (e.g., lack of criminal liability) were counted as deontic exceptions. The amount of exceptions generated was analyzed within a 2 (modal: should vs. will) x 2 (rule acceptance: high vs. low) x 2 (frequency of exceptions: high vs. low) ANOVA. Only a main

effect for the modal auxiliary, $F(1, 38) = 4.55$, $p = .04$; $\eta_p^2 = .107$, and a main effect of the amount of exceptions was found, $F(1, 38) = 5.25$, $p = .028$; $\eta_p^2 = .121$. Participants in the factual condition generated more exceptions than participants in the deontic condition (M = 2.53; SD = 1.28 vs. M = 1.83; SD = 0.69, respectively). Despite not being a big difference descriptively, participants also generated more exceptions for offences classified as having a low frequency of exceptions than for offences classified as having high frequency of exceptions (M = 2.29; SD = 1.16 vs. M = 2.06; SD = 1.08, respectively). In addition, the quality of exceptions differed depending on whether the participant was confronted with deontic or factual legal conditionals: while participants in the factual condition generated in 53% of the cases factual exceptions, participants in the deontic condition did this only in 20% of the cases. Along the same lines, deontic exceptions were more frequent in the deontic condition than in the factual condition (77% vs. 43%, respectively).

5.1.2.3. Moral Outrage

The moral outrage ratings given to offences in the inference task were correlated with the corresponding rule acceptance ratings for the same offences from the preliminary study. As expected, the mean moral outrage ratings correlated with the mean rule acceptance ratings. This was the case for both, the moral outrage ratings given in the "should", condition, $r = .95$, $p < .001$, and the ones in the "will" condition, $r = .97$, $p < .001$. The more the participants accept a legal rule, the more moral outrage they feel when this rule is broken.

5.1.3. Discussion

The results of Experiment 6 show that the modal auxiliary used in legal conditionals affects MP inferences. If the conditional is phrased with the modal auxiliary "should", participants seem to rely on their own sense of justice: whereas for conditionals about severe offences participants show very low rejection ratings (i.e., favoring that the offender should be punished), for conditionals about mild offences - which they do not accept - they show higher rejection ratings (i.e., favoring that the offender should not be punished). This supports the Hypothesis 1 and replicates the findings of Experiments 1-5. However, the effect is attenuated for conditionals phrased with the modal "will". For factual conditionals also the frequency of

exceptions matters. Whereas for conditionals with a low frequency of exceptions participants show small rejection ratings (i.e., favoring that the offender will be punished), for conditionals with a high frequency of exceptions they show higher rejection ratings (i.e., favoring that the offender will not be punished). Since this was the case for both severe and mild offences, the modal "will" seems to make people think that factual information, and not deontic principles, are asked for. This is in accordance to Hypothesis 2 and is corroborated by the generation task, where factual exceptions were much more present in the factual condition than in the deontic one. Yet, an unexpected result was that the conclusions drawn from factual conditionals were not only influenced by the frequency of exceptions, but also by the participants' acceptance of the rule. One probable explanation is that participants still followed their own sense of justice to some extent. This is in accordance with the results of Experiment 2, where laypeople were still affected by their feelings of moral outrage even though instructed to decide like a real judge.

Contrary to the results for MP inferences, those for MT were not so straightforward. I expected to find the same pattern of results as for MP inferences, but this was not the case. None of the factors influenced the conclusions selected for MT inferences. This is not the first time MT produces unexpected results (e.g., De Neys et al., 2003a; Johnson-Laird & Byrne, 1991; Singmann et al., 2014). A common explanation is that the negation in the MT inference makes the task more difficult (De Neys et al., 2003a; see Chapter 1 for further explanations). In this thesis, this higher difficulty might have led to more logical errors, which might have covered the effects of modals, exceptions, and rule acceptance. In fact, participants showed higher rejection ratings for MT (M = 2.48; SD = 1.45) than for MP inferences (M = 1.40; SD = 0.97), $t(39) = 4.54$, $p < .001$, $d = 0.86$.

An alternative explanation is that the higher rejection ratings for MT inferences occurred because participants were in some way led to assume that p was the case although q was not. According to Sperber and Wilson's (1995) principle of relevance, people make inferences about the messages they get assuming that all the information is relevant to them. In this way, the information given in MT inferences of somebody being not punished (i.e., $\neg q$) is only relevant if there are reasons to assume that the person could have been punished, e.g., because the person actually committed the offence. So, if people think that someone actually committed an offence, but find out that the person is not punished, then they should implicitly consider that some exception occurred. That is, the utterance denying that person is punished is only relevant if there are reasons for thinking that the offender could be punished but some exception occurred (e.g., not being caught, having some legal justifications).

The idea that the negation in MT inferences suggests that the antecedent is actually true, but that some exception happened was already tested empirically by Bonnefon and Villejoubert (2007; see also Oaksford and Chater, 2013). To test if this explanation also applies for legal conditionals I conducted an online study where participants (N = 112) were confronted with negated consequents of legal conditionals (e.g., Sven is not punished for bodily injury) and were asked to write down why somebody would utter this sentence. While in only 7% of the cases participants said that ¬*q* was uttered because of ¬*p*, in 37% of the cases they said ¬*q* was uttered because *p* happened but some exception occurred. In the remaining cases participants did not provide any concrete reasons for this utterance, but for instance, described situations when such sentence could be uttered leaving unclear whether they thought *p* was the case or not (e.g., "after a trial"). Notwithstanding these results, it is still premature to conclude that conversational implicatures are responsible for the high rejection ratings of MT inferences. Further studies are necessary, especially to understand why frequencies of exceptions did not affect MT inferences.

A last point I want to make is that participants generated slightly more exceptions in the generation task for conditionals with a low frequency of exceptions compared with conditionals with a high frequency of exceptions. Although testing the relationship between frequency and amount of exceptions was not a primary aim of this experiment, this mismatch between amount and frequency of exceptions is not surprising. As explained in Chapter 1, already Geiger and Oberauer (2007) showed that, although often correlated, the amount of exceptions and the frequency of exceptions do not always lead to the same effects (cf. Fernbach & Erb, 2013). There may be exceptions which happen often (e.g., self-defence or necessity) and exceptions which happen less often (e.g., being coerced by threats to life and physical integrity). It is important to keep this difference in mind especially when dealing with legal conditionals. In legal reasoning, when asking whether an offender will be punished or not, it is necessary to know how often exceptions occur and not how many different exceptions there may exist. In other words, it is important how often an offence will be punished, and not how many different causes may exist in principle for not punishing somebody. As Fernbach and Erb (2013) proposed, the power of the antecedent in predicting the consequent does not only depend on the amount of exceptions, but also on how relevant they are, and in this way - maybe - also on how often they occur. If further studies corroborate the results found in this paper, then in a next step one could let participants assign probabilities to the premises and conclusions of legal conditionals, and test in how far

probabilistic accounts can explain this thesis' results (cf. Evans, Thompson, & Over, 2015; Oaksford & Chater, 2007; Singmann et al., 2014).

5.2. Experiment 7: The Modal Auxiliary and Closeness with the Offender

In Experiment 6 it was shown that people follow their own sense of justice when reasoning with deontic legal conditionals, but that this can be attenuated by changing the modal auxiliary into will. The aim of Experiment 7 was to test one further implication of using different modals in legal reasoning. If "should" prompts following one's own deontic preferences and "will" prompts using one's knowledge about what happens in the real world, then it should be possible to introduce experimental manipulations that affect only inferences for either deontic or factual legal conditionals. Along these lines, in Experiment 7 the offenders were presented as people we love (i.e., family members or best friends). People are usually more forgiving and caring with those they are close with and whom they like (e.g., Fehr, Gelfand, & Nag, 2010; McCullough et al. 1998; Sprecher & Fehr, 2005), and do not want them to be hurt or in danger (e.g., Buckley, Chapman, Sheehan, & Cunningham, 2012; see also Swann et al., 2014). Consequently presenting the offender as someone close should affect deontic legal conditionals by making people less strict than when deciding about a random person (as they did in Experiment 6). However, it should not affect factual conditionals, because the relationship one has with an offender does not influence what actually happens in the real world (e.g., events involving police, judges, penal code, etc.).

Hypothesis 1: When the legal conditional is phrased deontically with the modal *should*, then framing the offender as a close or beloved person should affect inferences, making conclusions less punitive.

Hypothesis 2: When the legal conditional is phrased factually with the modal *will*, then framing the offender as a close or beloved person should not affect inferences. Inferences should depend on the frequency of exceptions.

5.2.1. Methods

5.2.1.1. Participants

42 students participated in the experiment. All participants with academic knowledge about law or formal logic were excluded. The final sample thus consisted of 40 participants (20 female). Half of the participants received deontic legal conditionals, the other half factual legal conditionals.

5.2.1.2. Material, Design, and Procedure

Experiment 7 was constructed as Experiment 6, with the only difference that the offender in the second premise was specified as being one's mother, father, best friend (either female – in German: "beste Freundin", or male – in German: "bester Freund"). Which offence was paired with which relative was selected randomly for MP and MT inferences. It was made sure that family members and friends are distributed uniformly among all offence categories. As an illustration:

If a person kills another human, then the person should/will be punished for manslaughter.
Your father killed another human.

Should/Will your father be punished for manslaughter?

As in Experiment 6, participants had to solve in addition to the inference task a generation task and a moral outrage task. The offenders in these supplementary tasks were also family members and friends.

5.2.2. Results

5.2.2.1. Inference Task

The conclusion ratings were transformed as in Experiment 6 into rejection ratings. The rejection ratings for MP and MT inferences can be found in Figure 8. These were analyzed in

a 2 (modal: should vs. will) x 2 (rule acceptance: high vs. low) x 2 (frequency of exceptions: high vs. low) x 2 (inference: MP vs. MT) ANOVA. The ANOVA showed main effects of rule acceptance, $F(1,38) = 20.93$, $p < .001$, $\eta_p^2 = .355$, and of frequency of exceptions, $F(1,38) = 32.93$, $p < .001$, $\eta_p^2 = .464$, but also interactions between inference and rule acceptance, $F(1,38) = 4.74$, $p = .036$, $\eta_p^2 = .111$, between inference, rule acceptance and modal, $F(1,38) = 11.51$, $p = .002$, $\eta_p^2 = .232$, and between inference, rule acceptance, and frequency of exceptions, $F(1,38) = 8.00$, $p = .007$, $\eta_p^2 = .174$. Because of these interactions I continued analyzing the data in two separate 2 (modal: should vs. will) x 2 (rule acceptance: high vs. low) x 2 (frequency of exceptions: high vs. low) ANOVAs for MP and MT inferences.

For MP inferences the ANOVA revealed a main effect of rule acceptance, $F(1, 38) = 22.95$, $p <. 001$, $\eta_p^2 = .373$, a main effect of frequency of exceptions, $F(1, 38) = 9.99$, $p = .003$, $\eta_p^2 = .208$, and an interaction between rule acceptance and modal auxiliary, $F(1, 38)=7.25$, $p = .01$, $\eta_p^2 =.160$. The interaction between modal auxiliary and rule acceptance shows that rule acceptance affected deontic legal conditionals, but not factual legal conditionals: whereas rejection ratings for deontic conditionals were higher for low ($M = 2.44$; $SD = 1.39$) than for high acceptance rules ($M = 1.04$; $SD = 0.95$), $t(19) = 4.25$, $p < .001$, $d = 1.16$, for factual conditionals this effect did not reach the Bonferroni adjusted alpha level of .025, $t(19) = 2.13$, $p = .046$, $d = 0.44$ ($M = 1.96$; $SD = 0.95$ vs. $M = 1.58$; $SD = 0.79$; respectively). In fact, the rejection ratings found for deontic legal conditionals were generally higher than the corresponding ones from Experiment 6, showing that participants tended to punish less when offenders are close relatives. This observation was confirmed by comparing the mean rejection rating of deontic MP inferences in Experiment 7 ($M = 1.74$; $SD = .93$) with that in Experiment 6 ($M = 1.06$; $SD = .87$), $t(38)=2.39$, $p = .022$, $d = 0.76$. All other effects, including the interaction between modal auxiliary and frequency of exceptions, were not significant (all $Fs \leq 1.51$, $ps \geq .227$).

The ANOVA for MT inferences showed a main effect of rule acceptance, $F(1, 38) = 7.827$, $p = .008$, $\eta_p^2 = .171$, a main effect of frequency of exceptions, $F(1, 38) = 24.22$, $p < .001$, $\eta_p^2 = .389$, and an interaction between both factors, $F(1, 38) = 10.632$, $p = .002$, $\eta_p^2 = 219$. All other effects were not significant ($Fs \leq 1.85$, $p \geq .181$). The interaction shows that frequency of exceptions affected conditionals with highly and little accepted rules differently. For highly accepted rules frequency of exceptions did not affect inferences: participants showed low rejection ratings for offences with high ($M = 1.41$; $SD = 1.16$) and low ($M = 1.05$; $SD = 1.47$) frequency of exceptions, concluding that if an offender is not punished (not *q*), then this offender probably did not commit an offence (not *p*), $t(39) = 1.49$, $p = .144$, $d = 0.27$.

However, for little accepted rules the frequency of exceptions mattered: if the offence had only a low frequency of exceptions, participants showed lower rejection ratings (M = 1.00; SD = 1.15) than when the offence had a high frequency of exceptions (M = 2.40; SD = 1.82), $t(39) = 5.81$, $p < .001$, $d = 0.87$ (Bonferroni adjusted alpha level: $\alpha = .025$). For better comparisons I plotted the rejection ratings for deontic and factual legal conditionals together (since there was no effect of modals) and compared them with the corresponding rejection rates for MT in Experiment 6. A 2 (frequency of exceptions: high vs. low) x 2 (rule acceptance: high vs. low) x 2 (experiment: 6 vs. 7) ANOVA confirmed the initial analysis: a main effect of experiment shows that participants indeed accept MT inferences more when the offender is a close relative compared to when the offender is not specified, $F(1, 78) = 12.66$, $p = .001$, $\eta_p^2 = .14$, and a three-way-interaction between all factors confirms that the interaction between frequency of exceptions and rule acceptance is unique for Experiment 2, $F(1, 78) = 6.316$, $p = .014$, $\eta_p^2 = .075$ (all other effects were explained by the three-way interaction or were not significant).

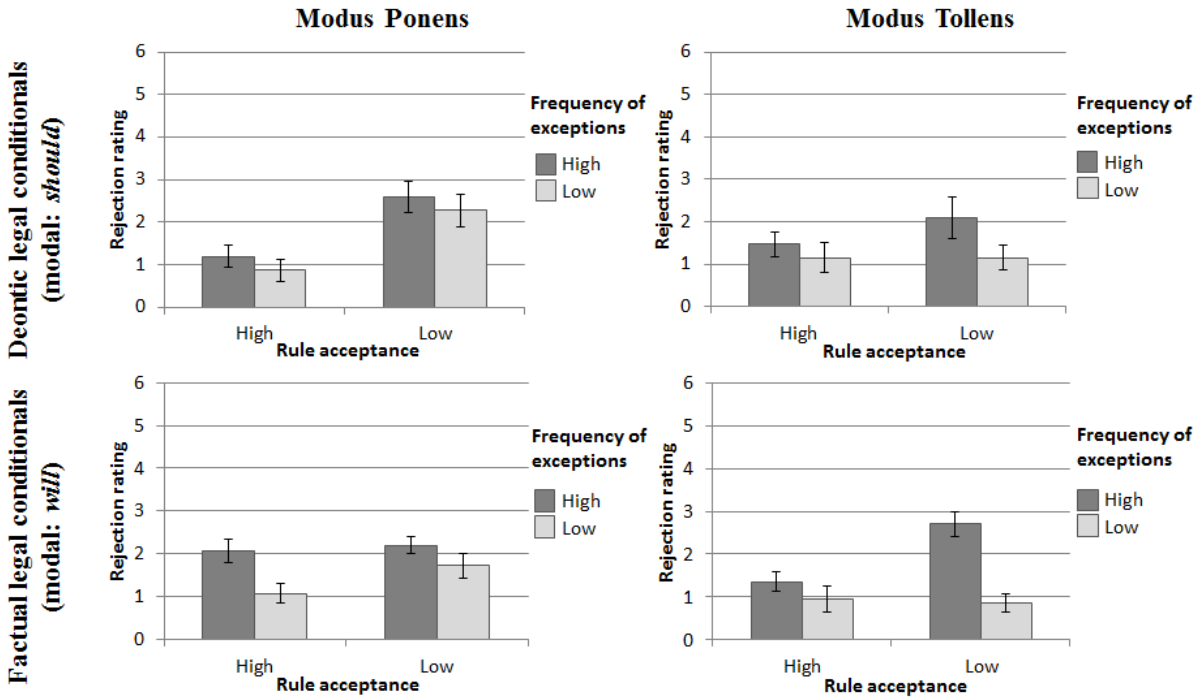

Figure 8. Rejection ratings (0 - 6) for Modus Ponens and Modus Tollens inferences for deontic and factual legal conditionals in Experiment 7. Error bars represent standard errors.

5.2.2.2. Generation Task

Two independent raters coded the amount and quality of exceptions generated (ρ = .99 for the amount of exceptions in general, ρ = .96 for amount of factual, and ρ = .92 for the amount of deontic exceptions). The amount of exceptions was analyzed within a 2 (modal: should vs. will) x 2 (rule acceptance: high vs. low) x 2 (frequency of exceptions: high vs. low) ANOVA. Only an interaction between modal auxiliary and rule acceptance was found, $F(1, 38) = 9.48$, $p = .004$, $\eta_p^2 = 0.2$. Participants in the deontic condition generated slightly more exceptions for low ($M = 2.03$; $SD = 0.99$) than for high ($M = 1.73$; $SD = 0.88$) acceptance rules, while participants in the factual condition showed a trend in the opposite direction ($M = 1.9$; $SD = 0.77$; $M = 2.29$; $SD = 1.10$; respectively). However, pairwise *t*-tests did not reach the Bonferroni adjusted alpha level of α = .025 ($t(19) = 2.40$, $p = .027$, $d = 0.32$ and $t(19) = 2.09$, $p = .050$, $d = 0.38$; respectively). Yet, as in Experiment 6, participants in the factual condition generated in 65% of the cases factual exceptions and only in 33% of the cases deontic exceptions. For participants in the deontic condition it was the other way around (75% deontic exceptions and 22% factual exceptions).

5.2.2.3. Moral Outrage

Moral outrage ratings given to offences in the inference task correlated with the corresponding rule acceptance ratings for the same offences from the preliminary study ($r = .91$, $p = .002$ for the correlation between the moral outrage ratings given in the "should", condition, and $r = .93$, $p = .001$ for the ones in the "will" condition).

5.2.3. Discussion

The results once more show that the modal auxiliary affects the conclusions drawn from legal conditionals. For MP inferences, when asked whether an offender should be punished or not, people generally decide that severe offences should be punished and minor offences not. However, this effect is attenuated when the modal auxiliary is changed to "will". In addition, also the relationship with the offender affects the impact of rule acceptance on deontic

punishment decisions. People are less severe when the offender is a close relative, probably because they do not want that somebody they like to be punished.

When planning Experiment 7 it was thought that describing offenders as close people should only affect inferences for deontic legal conditionals, but not for factual legal conditionals. Yet, the present results show that the relationship with the offender also influenced inferences for factual legal conditionals. As shown in Figure 8, different to the results found in Experiment 6, the frequency of exceptions only affected factual legal conditionals with highly accepted rules. When the conditional contained a little accepted rule, participants did not differentiate between offences with high or low frequency of exceptions. On the contrary, they also decided that for offences with a low frequency of exceptions the offender will not be punished. Although unexpected, this result can be explained if it is considered that frequency of exceptions was calculated in the preliminary study of Experiment 6 by asking participants about the frequency of exceptions for an unspecified person ("A person kills another human, but the person is not punished for manslaughter. In how many of 100 cases do you think this is the case?"). It is possible that when participants are told that it was a close relative who committed the offence – and it is a mild and therefore understandable offence – they calculate the probability of being punished differently, assuming that the relative's offence will be one of the few exceptional cases were no punishment follows. As Fernbach and Erb (2013) postulated, the power of an antecedent in predicting the consequent depends on the context. This could also explain why in the generation task of Experiment 7 the mismatch between frequency of exceptions and amount of exceptions found in Experiment 6 could not be replicated. Given that the frequency of exceptions estimations were calculated in the preliminary study by asking about an unspecified person, and the generation task asked to generate exceptions for one specific close person (e.g., Your best friend killed another person but should/ will not be punished for that. Why?), it could be possible that this mismatch covered any advantage of frequency over amount of exceptions.

Contrary to Experiment 6, participants in Experiment 7 only seldom rejected MT inferences, suggesting that when the offender is someone close, the negated consequent stated in MT inferences does not suggest the existence of exceptions. This result can be explained by remarking that people normally do not believe that beloved people commit regrettable actions. Because of this, when people hear that a beloved person is not punished, they do not conclude that he or she committed an offence and that some exception happened. Instead, they simply conclude that he or she did not commit the offence. Only if the offence is mild and people

know that a lot of exceptions exist, people can imagine that the offence was committed but that an exception occurred (as suggested by the elevated rejection rates for conditionals with little accepted rules and high frequency of exceptions).

5.3. Experiment 8: Quantifier[5]

Another way in which changes in legal conditionals can enhance the consideration of defeaters is by suggesting the existence of exceptions through the phrasing of the legal conditional. In all experiments so far, legal rules have been presented as conditionals, describing the relationship between the offence and the punishment by an if-then relation, for example: "If a person kills another human, then the person should be punished for manslaughter." But what happens if instead of these if-then relations, legal rules are phrased with quantifiers? Quantifiers allow statements about the amount or quantity of objects in the domain of discourse (Knauff, 2006). They can thus be either universal (i.e., All As are Bs) or existential (i.e., Some As are Bs). This peculiarity of quantifiers, that is to describe either universal relations or relations which allow exceptions, can be used to manipulate people's consideration of exceptions in legal reasoning. Imagine for instance a legal rule phrased universally (e.g., "All persons that kill another human should be punished for manslaughter") and compare it to a rule phrased existentially (e.g., "Some persons that kill another human should be punished for manslaughter"). How would you decide when faced with the fact that an offence has been committed? As far as I know, all research on the consideration of exceptions in defeasible reasoning has been done with conditionals with the classical logical connective if-then. However, based on the logical and linguistic implications of universal and existential quantifiers, existential quantifiers should trigger the consideration of exceptions. Reasoners should defeat more conclusions when the rule is phrased with an existential compared to a universal quantifier. In other words, people's defeasible reasoning should be enhanced by using existential quantifiers. As a consequence, reasoners should also take longer to reach a conclusion from existential compared to universal rules. When reasoning with universal legal rules reasoners can simply answer according to their own sense of justice

[5] Data for this experiment was partially gained during the Bachelor Thesis of Katharina Horn (Horn, 2014), which I supervised.

guided by feelings of moral outrage. However, when people reason from existential rules, then the exceptions such a rule implies might conflict with one's own sense of justice, needing more time to draw a conclusion and making people less certain about their conclusions – at least for highly morally outraging offences.

Hypothesis 1: When legal rules are phrased as universal rules, then reasoners make more punishment decisions than when the rule is phrased as an existential rule.

Hypothesis 2: When legal rules are phrased as existential rules, then reasoners take longer to draw conclusions than when the rule is phrased as a universal rule – at least for highly morally outraging offences.

Hypothesis 3: When legal rules are phrased as existential rules, then reasoners are less certain about their conclusions than when the rule is phrased as a universal rule – at least for highly morally outraging offences.

5.3.1. Methods

5.3.1.1. Participants

43 participants took part in the experiment. Two participants had to be excluded because of technical problems and another one because the participant afterwards reported to study law. The remaining 40 participants were on average 23.08 years old (*SD* = 3.24).

5.3.1.2. Materials and Design

For Experiment 8 I selected from a pilot study (*N* = 87 and *N* = 82) 6 high and 6 low morally outraging offences. These offences were embedded into quantified legal rules by either adding an "All" or a "Some" in the beginning of each statement (for an illustration see Table 17).

Table 17

Rephrasing high and low morally outraging legal rules as universal or existential rules in Experiment 8

Rule	Moral Outrage of Offence	
	High	Low
Legal Rule	Whoever kills a human being, without being a murderer, is punished for manslaughter with imprisonment for not less than five years	Whoever organizes without governmental permission a game of chance or provides the facilities for this, is punished with imprisonment up to two years or with fine.
Universal	All persons that kill another human should be punished for manslaughter.	All persons that organize a game of chance without governmental permission should be punished for unauthorized organization of games of chance.
Existential	Some people that kill another human should be punished for manslaughter.	Some persons that organize a game of chance without governmental permission should be punished for unauthorized organization of games of chance.

The kind of quantifier was varied between individuals (n = 19 got the rules with universal, and n = 21 with existential quantifiers) and each quantified rule was presented twice, once as an MP and once as an MT inference. Examples are given in Table 18. After each inference, participants were told to rate their certainty on a 3-point-Likert scale (uncertain – neutral – certain). In total, participants were confronted with 24 problems. Experiment 8 thus followed a 2 (moral outrage: high vs. low) x 2 (inference: MP vs. MT) x 2 (quantifier: universal vs. existential) mixed design.

Table 18

Structure of the problems used in Experiment 8 illustrated by the legal conditional of manslaughter. R=Conditional rule; F=Fact; C=Conclusion.

	Modus Ponens		Modus Tollens	
All	R:	All persons that kill another human should be punished for manslaughter	R:	All persons that kill another human should be punished for manslaughter
	F:	A person kills another human.	F:	A person is not punished for manslaughter.
	C:	Should the person be punished for manslaughter?	C:	Did this person kill another human?
Some	R:	Some persons that kill another human should be punished for manslaughter.	R:	Some persons that kill another human should be punished for manslaughter.
	F:	A person kills another human.	F:	A person is not punished for manslaughter
	C:	Should the person be punished for manslaughter?	C:	Did this person kill another human?

5.3.1.2. Procedure

The experiment was programed in Superlab 4.5 from Cedrus Cooperation. Participants were tested individually. In the instructions participants were told that they will be confronted with statements describing legal cases, including some general rule about the offence and that they have to decide for each case whether they would apply the initial legal rule or not. Each statement (the quantified legal rule, fact, and conclusion) was presented on a separate screen. Participants could switch to the next screen by pressing the space bar. They gave their answer about the conclusion – which was written in red font – by either pressing a "Y" (yes) or "N" (no) key on the keyboard. The certainty ratings were given by pressing one of three keys from the numerical pad. Dependent variables were participants' conclusions, the decision times, and the certainty ratings. Participants were told to answer intuitively and that right or wrong answers do not exist. The 24 inference problems were presented in a random order after a short practice trial.

5.3.2. Results

Three separate 2 (moral outrage: high vs. low) x 2 (inference: MP vs. MT) x 2 (quantifier: universal vs. existential) mixed ANOVAs were conducted; one for the conclusions, one for decision times (corrected for sentence length as in Experiments 2 and 4), and one for certainty ratings. Descriptive statistics are given in Table 19.

For the analysis of the conclusions, I computed the percentage of logically "incorrect" responses per category. That is, for MP inferences percentages of no-punishment conclusions were computed, and for MT inferences the percentage of conclusions stating that the offence was committed. This measure therefore resembles the "rejection rating" from Experiments 6 and 7. The ANOVA showed a main effect of moral outrage, $F(1, 38) = 28.37$, $p < .001$, $\eta_p^2 = .43$, but also an interaction between moral outrage and inference, $F(1, 38) = 34.87$, $p < .001$, $\eta_p^2 = .48$. For MP inferences, when the offence was of high moral outrage, participants selected significantly less often to conclude that the offender should not be punished ($M = 5.83$; $SD = 10.37$) compared to when the offence was of low moral outrage ($M = 35.83$; $SD = 26.30$), $t(39) = 7.03$, $p < .001$, $d = 1.47$. In contrast, for MT inferences moral outrage did not affect the participants' conclusions ($M = 18.75$; $SD = 30.71$; $M = 16.25$; $SD = 27.60$; for high and low moral outrage respectively), $t(39) = 0.80$, $p = .430$, $d = 0.08$. All other effects were not significant ($F \leq 2.75$, $p \geq .106$).

The ANOVA for decision times showed a main effect of inference, $F(1, 38) = 25.82$, $p < .001$, $\eta_p^2 = .41$. Participants needed more time to decide for MT inferences ($M = 3.95$s; $SD = 1.56$) than for MP inferences ($M = 2.80$s; $SD = 1.08$). Also an interaction between inference and moral outrage was found, $F(1, 38) = 4.68$, $p = .037$, $\eta_p^2 = .11$. For MP inferences participants were descriptively faster when the offence was of high moral outrage ($M = 2596$ms; $SD = 1155$) than of low moral outrage ($M = 2993$; $SD = 1280$), but for MT it was the other way around ($M = 4070$; $SD = 1489$; $M = 3821$; $SD = 2050$; respectively). However, both post hoc *t*-tests did not reach the Bonferroni adjusted alpha level of .025 ($ts \leq 2.22$, $p \geq .033$). All other effects were not significant ($F \leq 3.31$, $p \geq .077$).

The ANOVA for certainty ratings revealed a main effect of moral outrage, $F(1, 38) = 8.48$, $p = .006$, $\eta_p^2 = .18$, a main effect of inference, $F(1, 38) = 47.92$, $p < .001$, $\eta_p^2 = .56$, and an interaction between moral outrage and offence, $F(1, 38) = 35.25$, $p < .001$, $\eta_p^2 = .48$. For MP inferences participants were more certain when the rule contained a high moral outrage offence ($M = 2.87$; $SD = 0.17$) compared to when it contained a low moral outrage offence (*M*

= 2.50; *SD* = 0.35), $t(39) = 6.08$, $p < .001$, $d = 1.34$. For MT inferences, however, differences in certainty ratings for offences with high and low moral outrage did not reach the Bonferroni adjusted alpha level of .025 (*M* = 2.05; *SD* = 0.63; *M* = 2.17; *SD* = 0.50; respectively), $t(39) = 2.14$, $p = .039$, $d = 0.21$. All other effects were not significant ($F \leq 2.01$, $p \geq .164$).

Table 19

Percentage of no-punishment/ offence-committed conclusions (C), decision times in seconds (DT), and certainty ratings (CR) for high and low morally outraging universal and existential rules for Modus Ponens (MP) and Modus Tollens (MT) inferences in Experiment 8. Standard deviations are shown in brackets

Rule	High Moral Outrage		Low Moral Outrage	
	MP	MT	MP	MT
Universal				
C	7.89 (11.6)	15.79 (26.3)	39.47 (25.6)	10.53 (17.8)
DT	2.9 (1.1)	4.4 (1.6)	3.2 (0.8)	4.3 (2.3)
CR	2.9 (0.2)	2.1 (0.5)	2.4 (0.3)	2.2 (0.3)
Existential				
C	3.97 (9.0)	21.43 (34.6)	32.54 (26.3)	21.43 (33.8)
DT	2.3 (1.1)	3.8 (1.4)	2.8 (1.6)	3.4 (1.7)
CR	2.9 (0.2)	2.0 (0.7)	2.6 (0.4)	2.1 (0.6)

5.3.3. Discussion

The aim of Experiment 8 was to show that people's defeasible reasoning can be enhanced when legal rules are phrased as existential rules instead of universal rules. However, the results do not support my hypotheses. No effects of quantifiers were found, not even in decision times or certainty ratings. There are several reasons why the quantifier used in the legal rule may not have affected inferences. One reason is that the legal rule was still phrased deontically, with the modal auxiliary "should". It is possible that the effect of the modal "should" was stronger than the subtle variation in the quantifier. In other words, it might be that reasoners simply ignored the quantifier used in the rule. This maybe happened because

the quantifier was varied between subjects. In this experiment I decided to work with a between subjects design to avoid artefacts because of demand characteristics. Further studies could therefore test if the quantifier is still ignored in within designs. If in within designs people do consider the quantifier, then this would show that people know the differences between both quantifiers but that they do not care about these differences spontaneously.

Another explanation related to the previous one is that the manipulation of the quantifier only affected the antecedent of the general legal rule, but not the conclusion. The question about the conclusion in both – universal and existential – conditions was still the same, namely whether the offender should be punished (MP) or whether the offence was committed (MT). It is possible that in the moment of deciding which conclusion to draw, people only look at this specific question about the conclusion. Which in the case of being deontically, asking about what should happen, evokes the reasoners' own sense of justice as showed in Experiments 6 and 7. The way the previous premises were written are ignored or forgotten, only their main message (e.g., somebody killed another human and this requires punishment) is extracted; especially if the premises are presented one after the other – without the possibility to scroll back – as it was the case in this experiment.

There are of course other explanations for the missing effects in this study, such as that the dichotomous answer modality did not differentiate sufficiently (see Markovits, Forgues, & Brunet, 2010). It could also be possible that people do not represent the logical differences between universal and existential quantifiers mentally, or at least not in a way it could affect inferences. Newstead (1989) already showed that reasoners do not understand quantifiers in the same way logic does. However, it is important to notice that in this study no logical interpretation of quantifiers was necessary. The differences between universal and existential quantifiers used in everyday language were thought to be enough to evoke different conclusions. The distinction we draw in our everyday life between all and some should have evoked or at least made participants aware that some includes exceptions. In fact, it is the logical interpretation which actually allows us to use *some* even if no exceptions exist. In this way the explanation of not understanding the meaning of *all* and *some* cannot explain the results of Experiment 8.

In sum, varying the kind of quantifier did not moderate the effect of one's own sense of justice on legal reasoning. However, the results of Experiment 8 are still interesting for cognitive psychology. The results show that people defeat conclusions also when reasoning with quantified legal rules. When quantified legal rules described low morally outrageous offences and were framed in MP inferences, participants defeated the conclusion to punish the

offender more often than when the offence was of high moral outrage (and were less certain in their judgment). Yet, when quantified legal rules were framed in MT inferences, moral outrage did not affect conclusions. Moral outrage likely affected MP but not MT inferences because only the conclusion of the former activates personal values by asking what should happen to an offender. Contrary to MP, the structure of MT inferences is less emotionally charged, because MT conclusions only ask whether an offence was committed or not. This is in accordance with the findings from Experiments 6 and 7: also Experiment 8 shows how important the specific phrasing of the conclusion is. The conclusions people draw depend on what is asked for and how it is asked for it.

5.4. Summary and Implications of Chapter 5

The results from Chapter 5 show the significance of linguistic factors in reasoning. Experiments 6 and 7 show that the way a conditional rule is phrased affects which kind of information is used for reasoning and which conclusions are drawn. People's own sense of justice is only highly predictive for legal reasoning when the legal conditionals are phrased deontically, asking for what *should* happen to an offender. If instead the legal conditional is phrased factually, asking for what *will* happen to an offender, then reasoners consider exceptions and one's own sense of justices loses importance. It was also expected that the universality of a legal rule should affect legal reasoning. However, Experiment 8 showed that changing the legal conditional into a universal or existential rule did not affect inferences. Probably because changes which only alter the way the legal rule is phrased are not as notorious as those which also alter what the conclusion asks for.

I am not the first to highlight the importance of linguistic factors in the psychology of reasoning. For example, Schmeltzer and Hilton (2014) showed that the pragmatic implications of the antecedent influence the conclusions participants draw. There are many different kinds of conditionals in the literature, e.g., causal conditionals, conditionals describing threats, tips, and promises (see Dieussaert et al., 2002) or consequential conditionals describing outcomes (Bonnefon & Hilton, 2004), making clear how important it is to track their different formulations when making predictions about how people reason with them. The distinction between deontic and factual conditionals is especially relevant. As I have shown for legal

reasoning, deontic and factual conditionals activate different systems of rules: while the former activates the moral system, the latter activates the knowledge about real world. Some might argue that our factual conditionals were still deontic in certain way because they still represented some kind of rule. I agree with that. However, instead of weakening my arguments, this criticism supports the hypothesis about the relevance of modals in reasoning: although the content was in both cases somewhat deontic, the different modals used in order to express the rule made participants draw different conclusions.

One last point that needs further investigation are the results for the MT inferences. In all three experiments in this Chapter, factors that affected MP inferences did not affect MT inferences. For instance, while the preference of *pq* over *p*¬*q* (Experiments 6 & 7) or moral outrage (Experiment 8) affected MP inferences (e.g., by enhancing punishment conclusions), the endorsement of MT conclusion were not affected by people's sense of justice. As argued in Experiment 8, one main reason is probably that the structure of MT inferences is less emotionally charged than the structure of MP inferences. While MP inference ask for what should happen (or will happen in Experiments 6 and 7), in MT inferences participants are only asked whether an offence was committed or not. They are therefore required to make a rather objective analysis of the matters of facts. However, another reason may be the one presented in Experiment 6. In Experiment 6 I argued that according to the principle of relevance (Sperber & Wilson, 1995) the categorical of MT inferences ("The person should/will not be punished for offence X") makes people aware of exceptions. If this conversational implicature (see Grice, 1975) of legal MT inferences is correct, then it is necessary to 1) re-interpret existing data on MT endorsement, and 2) conduct further studies to identify moderators of this effect. According to Experiment 7 a possible moderator could be for instance a person's personal commitment to the conditional. The denial of the consequent enhances people's thought of exceptions only if there are no personal pragmatic implications in conflict with it. When people do not want to believe that the antecedent is true (here: someone close doing something they dislike), then hearing that the consequent is not the case fulfills their expectations making them to conclude that the antecedent is not true. In other words, the motivation to conclude that the antecedent is not the case might bring people to ignore exceptions and thereby endorse the classically correct conclusion to MT. All in all, further studies are necessary to fully understand the role of conversational implicatures on reasoning and especially on MT.

In sum, the results corroborate the complexity of reasoning. Contrary to the assumptions of classical logic, the way people interpret and reason with conditionals often

depends on its content, on the personal relevance the rule has, as well as on its linguistic phrasing. In some cases people may consider their own inner values when drawing conclusions, but in other cases factual information about exceptional situations take precedence.

Chapter 6: The Defeasibility of Negative Emotions in Everyday Scenarios

In the experiments presented in the previous Chapters, participants had difficulties withdrawing from the conclusion of punishing an offender. Moral outrage theory explains this finding by arguing that people experience negative intuitive emotions of moral outrage and a desire for punishment when faced with offences. Accordingly, people's difficulties in considering exculpatory circumstances in defeasible reasoning can be also interpreted as a difficulty of leaving aside the negative emotions (i.e., moral outrage) evoked by the offences. The aim of this Chapter is to investigate whether this difficulty in withdrawing negative emotions is something particular for the legal domain, or whether it also exists in everyday situations. In our daily life we often have to make inferences about the consequences of emotionally charged situations. For instance, we might know that "If I fail an exam, then I am sad". In the case we actually fail an exam we can thus conclude that we will be sad. Similar to the punishment conclusions in legal reasoning, this conclusion of being sad is the result of a conditional framing a situation which one considers to be negative. Now, if withdrawing conclusions evoked by negative emotions is not particular to legal reasoning but also exists for everyday situations, then it should also be difficult to withdraw negatively-charged conclusions in everyday situations. In the example above this implicates that people should have difficulties in withdrawing the conclusion of being sad after having failed in an exam. This difficulty in defeating negatively-charged conclusions should bring participants to still conclude being sad even when defeaters (e.g., positively-charged information, such as going to the cinema with a friend) are available.

In the following I present two experiments on the defeasibility of negative emotions in everyday scenarios. The first experiment investigates the withdrawal of conclusions from highly negative situations (analogous to the case of high moral outrage) and the second the withdrawal of conclusions from less negative situations (analogous to the case of low moral outrage).

6.1. Experiment 9: Defeating Strong Negative Emotions

Experiment 9 explored the role of strong negative emotions in defeasible reasoning. It was investigated whether people's reluctance to defeat conclusions from legal conditionals describing strongly negative-charged situations – such as high moral outrage offences – also has a correspondence in everyday scenarios. In the previous experiments, very severe offences such as manslaughter or maltreatment of wards were linked to negative emotions like moral outrage. However, strong negative emotions can also be evoked in everyday situations, as for example in the case of loss of a beloved person. Consider for instance the following inference problem:

If my mother dies, then I am sad.
My mother dies.
I get my dream job.

Am I sad?

How would you answer? Just like legal conditionals described an offence in the antecedent and the corresponding punishment in the consequent, also this conditional describes a negatively-charged situation in the antecedent, and the corresponding emotion in the consequent. Further, just like the inference tasks on legal reasoning contained information about defeaters (i.e., exculpatory circumstances), this inference problem also contains a defeater: a positively-charged situation, which evokes emotions contrary to those triggered by the situation in the antecedent. Therefore, if the findings from legal reasoning have a correspondence in everyday situations, then reasoners should feel attracted to draw MP inferences and conclude that one is sad after a negative situation; even in light of defeaters.

Evidence in this direction can be found in the psychological literature on attitudes and emotions. Several researchers have shown the existence of a so-called *negativity bias* (see Cacioppo, Gardner, & Berntson, 1997; Peeters & Czapinski, 1990). People "respond more strongly to very negative stimuli than to matched positive stimuli" (Norris, Larsen, Crawford, & Cacioppo, 2011, p. 100; see also Ito, Cacioppo, & Lang, 1998). As Vaish, Grossmann, and Woodward (2008) review, in comparison to positive information, negative information is observed longer (e.g., Fiske, 1980), weighted more heavily when making decisions and evaluations (e.g., Atthowe 1960; Kahneman & Tversky, 1983; Kanouse & Hanson, 1972), and considered more when making inferences about other people's traits (e.g., Aloise, 1993; Skowronski & Carlston, 1989; Wyer & Hinckle, 1976; for reviews on the negativity bias see

Peeters & Czapinski, 1990; Vaish et al., 2008). Negative information is therefore more informative than positive information (Peeters & Czapinski, 1990; Vaish et al., 2008).

In Experiment 9 the amount of defeated conclusions drawn from conditionals describing negatively-charged situations is compared with the amount of conclusions drawn from conditionals describing positively-charged situations. For instance:

If I get my dream job, then I am happy.
I get my dream job.
My mother dies.

Am I happy?

This comparison serves on the one hand as a control condition to show that negative emotions are indeed difficult to defeat. On the other hand it also provides a measure of how difficult it is to defeat negatively-charged conclusions compared to positively-charged ones.

Hypothesis: People defeat conclusions drawn from strongly negatively-charged conditionals less often than conclusions drawn from strongly positively-charged conditionals

6.1.1. Methods

6.1.1.1. Participants

30 participants (15 female) took part in the experiment. All but one participant were students. The mean age was M = 24.83 years (SD = 3.82).

6.1.1.2. Material and Design

36 conditional inference problems were constructed, 24 of them with additional information like the problems in the introduction of this experiment (i.e., experimental items), and the remaining 12 problems without any additional information (i.e., control items; for a similar design see Hilton et al., 1990). The problems therefore consisted of 1) an initial conditional rule, 2) the fact that the antecedent of this rule is given, 3) an additional information (omitted for control items), and 4) the question about the conclusion. I experimentally varied the

valence of the initial conditional rule and the valence of the additional information by using situations rated as positive or negative by $N = 255$ participants in an online preliminary study via SoSci Survey (Leiner, 2014). In this preliminary study participants had to rate how they feel about $n = 31$ positive and $n = 31$ negative situations on a 7-point-Likert scale (1= *very sad*, 7= *very happy*). For the experiment the six most positive ($M = 6.40$; $SD = 0.16$) and the six most negative ($M = 1.50$; $SD = 0.26$) situations were selected and brought into a conditional form (i.e., "If [the situation], then I am sad/happy."). Each conditional rule was presented three times, once with additional information of the same valence, once with additional information of the opposite valance, and once without any additional information. As additional information I used the same six positive and six negative situations that were used to create the conditionals but without framing them into a conditional form. Hereby, I alternated which situation was presented as the conditional rule and which information was presented as the additional information. As illustrated in the two examples in the introduction of this experiment, in half of the cases a specific situation A was phrased as the conditional rule, together with situation B as the additional information. In the other half of the cases, however, it was the other way around: the situation B was now presented as the conditional rule, and situation A as the additional information. It was made sure that the pairs of positive and negative situations used in one inference problem were matched in strength, i.e., ratings for positive and negative situations were equally distant from the neutral scale midpoint. In a second step, in order to guarantee that the pairs of situations in each inference problem were not only matched in the strength of their emotions, but also in their associative strength, a second validation study was conducted. Via SoSci Survey (Leiner, 2014) I presented each of the 12 situations (and 12 filler situations) in a conditional form paired once with the positive and once with the negative consequent (e.g., "If your mother dies, are you then sad?"). Participants ($N = 67$) had to answer on a seven point Likert scale ranging from 1 = *Definitely not* to 7 = *Definitely yes*. Results corroborated that positive situations were as strongly associated to the "I am happy" consequents ($M = 6.67$; $SD = 0.40$) as were the negative situations to the "I am sad" consequents ($M = 6.56$; $SD = 0.47$), $t(66) = 1.63$, $p = .109$, $d = 0.23$).

The experiment followed a 2 (Valence of the conditional: positive vs. negative) x 3 (Valence of the additional information: positive vs. negative vs. none) within subjects design.

6.1.1.3. Procedure

The experiment was conducted with Cedrus SuperLab © 4.5 on a desktop computer. Participants were tested individually and instructed to answer spontaneously to the question about the conclusion in the end of each problem. They were told that there exists no right or wrong answer. Participants gave their answers by either pressing a "Y" (yes) or a "N" (no) key on the keyboard. The order of these keys was counterbalanced. The spacebar was used to switch from one premise to the next one. The question at the conclusion was written in red, the premises in black. After each problem participants had the possibility to take a break. Before starting the experiment participants completed two practice problems.

6.1.2. Results

I computed the amount of defeated conclusions in percent for each category of problems separately. One participant had to be excluded from the computations because he or she deviated over 4 SD from the mean in two of the four experimental conditions with additional information.

6.1.2.1. Control Problems with no Additional Information

An important prerequisite for the experiment was that when no additional information is presented, participants make the logically valid conclusion. Therefore, it was analyzed how often participants defeated the logically valid conclusion of control problems. Without any additional information participants almost never defeated the conclusion suggested by the conditional rule, neither for positively (M = 0.0%; SD = 0.0) nor for negatively (M = 3.45%, SD = 11.25) charged conditionals. In fact, a t-test for repeated measures shows that the participants' acceptance of positively-charged and negatively-charged conclusions did not differ from each other, $t(28) = 1.65$, $p = .110$, $d = 0.43$.

6.1.2.2. Problems with Additional Positive or Negative Information

To test people's defeasible reasoning from emotionally-charged conditionals, I conducted a 2 (Valence of the conditional: positive vs. negative) x 2 (Valence of the additional information: positive vs. negative) ANOVA on the percentages of defeated conclusions of the experimental problems. Descriptive results can be found in Figure 9. As expected, the ANOVA showed a main effect for the valence of the conditional, $F(1, 28) = 20.64$, $p < .001$, $\eta_p^2 = .424$, a main effect for the valence of the additional information, $F(1, 28) = 41.90$, $p < .001$, $\eta_p^2 = .599$, and an interaction between both factors, $F(1, 28) = 204.79$, $p < .001$, $\eta_p^2 = .880$. This interaction was analyzed with post-hoc *t*-tests and a Bonferroni adjusted alpha level of .025. As can be seen in Figure 9, conclusions from positively-charged conditionals were defeated by negative additional information ($M = 74.71\%$; $SD = 28.04$) more often than were conclusions from negatively-charged conditionals by positive additional information ($M = 33.33\%$; $SD = 27.46$), $t(28) = 5.64$, $p < .001$, $d = 1.49$. When the conditional rule was paired with additional information of the same valence, participants did not differ in the percentage of defeated conclusions ($M_{neg\text{-}neg} = 4.02\%$; $SD = 11.49$; $M_{pos\text{-}pos} = 0.57\%$; $SD = 3.09$), $t(28) = 1.54$, $p = .136$, $d = 0.42$.

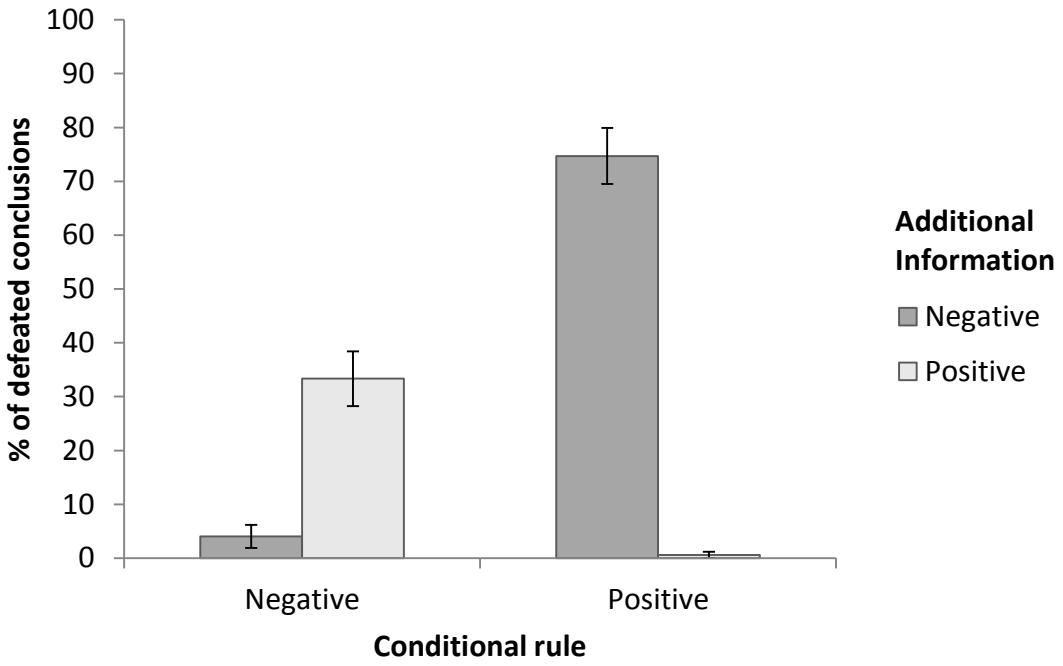

Figure 9. Percentage of defeated conclusions as a function of the valence of the conditional rule and additional information in Experiment 9. Error bars show standard errors.

6.1.3. Discussion

The results support my hypothesis and show that it is more difficult to withdraw conclusions drawn from negatively-charged conditionals than conclusions from positively-charged conditionals. This finding is in accordance with people's reluctance to defeat conclusions in legal reasoning. However, people's reluctance to defeat conclusions in legal reasoning was limited to highly morally outraging offences. When offences were only of low moral outrage, then people did defeat conclusions. In order to test whether the findings from legal reasoning indeed have a correspondence in everyday situations it is therefore necessary to test if people's defeasible reasoning from negatively-charged conditionals is enhanced when these emotions are only mild in strength. This will be tested in Experiment 10.

6.2. Experiment 10: Defeating Mild Negative Emotions

In Experiment 9 conditionals were constructed that described strongly negative situations and their corresponding emotion. As for legal reasoning, reasoners had difficulties to defeat conclusions from such negatively-charged conditionals. Experiment 10 aims to test whether people's reluctance to withdraw conclusions from negatively-charged conditionals persists when the situations described in the conditionals are only mildly negative. According to the findings in legal reasoning – where in cases of low moral outrage people did consider exculpatory evidence – conclusions from conditionals describing situations that evoke only mild negative emotions should be defeated to a higher extent than conclusions from strongly negative conditionals. As in Experiment 9, people's defeasible reasoning from negative emotions is tested by comparing the amount of defeated conclusions drawn from negatively-charged conditionals with the amount of defeated conclusions drawn from positively-charged conditionals.

Hypothesis: People defeat conclusions from weakly negatively-charged conditionals more often than conclusions drawn from strongly negatively-charged conditionals

6.2.1. Methods

6.2.1.1. Participants

31 participants took part in the experiment. One participant did not understand the task and was removed from the analysis. The final sample consisted thus of 30 participants (15 female). All but one participant were students. Their mean age was 23.90 years (*SD* = 2.98).

6.2.1.2. Material, Design, and Procedure

The problems used in Experiment 10 followed the same structure as the problems used in Experiment 9. Hence, they consisted of 1) a conditional rule, 2) a fact, 3) additional information (omitted for control problems), and 4) the question about the conclusion. The valence of the conditional and the additional information was varied exactly as in Experiment 9. However, contrary to Experiment 9, this time I did not choose highly negative and positive situations, but selected only situations of mild positive (e.g., "I see something beautiful on offer") and mild negative (e.g., "My favorite band splits") strength. These mild positive and mild negative situations were gained with the help of two online preliminary studies on SoSci Survey (Leiner, 2014). In these online studies *N* = 259 and *N* = 245 participants had to rate how they feel about different positive and negative situations on a 7-point-Likert scale (1= *very sad*, 7= *very happy*). In total *n* = 115 positive and *n* = 113 negative situations were created, yet each participant was only confronted with a subset of them (28-31 positive and 28-31 negative situations). For the experiment six positive and six negative situations with ratings around the scales mean of 4 were selected. The mean rating of the mild positive situations was *M* = 4.91 (*SD* = 0.15) and the mean rating for the mild negative situation was *M* = 3.04 (*SD* = 0.12). Again it was made sure that each pair of situations was matched in emotional strength. As in Experiment 9 a second validation study was conducted to test for associative strength (*N* = 43). The procedure was the same as in Experiment 9. Results corroborated that positive situations were as strongly associated to the "I am happy" (*M* = 5.03; *SD* = 0.93) consequents as were the negative situations to the "I am sad" consequents (*M* = 4.87; *SD* = 0.84), $t(42) = 1.47$, $p = .149$, $d = 0.18$).

The design and procedure was the same as in Experiment 9.

6.2.2. Results

As in Experiment 9, I computed the amount of defeated conclusions in percent for each category of items separately.

6.2.2.1. Control Problems with no Additional Information

Participants only seldom defeated the conclusion from problems without additional information (M_{pos} = 7.22%, SD_{pos} = 14.31, and M_{neg} = 18.33%, SD_{neg} = 22.47). Nonetheless, participants' acceptance of positively-charged and negatively-charged conclusions differed from each other, $t(29) = 2.76$, $p = .01$, $d = 0.56$. This will be considered in further computations.

6.2.2.2. Problems with Additional Positive or Negative Information

People's defeasible reasoning from conditionals describing situations only mild in emotional strength was investigated in two steps. First the percentages of defeated conclusions for positive and negatively-charged conditionals were analyzed. In the second step the results from this experiment were compared to the ones of Experiment 9, with the aim to discover whether people defeat more often conclusions from negatively-charged conditionals if the emotions are only mild of strength. Descriptive data can be found in Figure 10.

A 2 (Valence of the conditional: positive vs. negative) x 2 (Valence of the additional information: positive vs. negative) ANOVA on the percentage of defeated conclusions of the experimental items revealed no main effect of valence of the conditional, $F(1, 29) = 0.02$, $p = .882$, $\eta_p^2 = .001$, but a main effect of valence of the additional information, $F(1, 29) = 12.43$, $p = .001$, $\eta_p^2 = .300$, which is explained by an interaction between both factors, $F(1, 29) = 161.69$, $p < .001$, $\eta_p^2 = .848$. Participants defeated positive conclusions more often when they were followed by negative compared to positive additional information ($M_{pos\text{-}neg}$ = 60.56%, $SD_{pos\text{-}neg}$ = 21.66; $M_{pos\text{-}pos}$ = 6.11%, $SD_{pos\text{-}pos}$ = 15.46; $t(29) = 10.92$, $p <.001$, $d = 2.90$), and they defeated negative conclusions more often when they were followed by positive compared to negative additional information ($M_{neg\text{-}pos}$ = 48.89%, $SD_{neg\text{-}pos}$ = 30.30; $M_{neg\text{-}neg}$ = 16.11%, $SD_{neg\text{-}neg}$ = 21.66; $t(29) = 7.82$, $p < .001$, $d = 1.18$). Yet, as can be seen in Figure 10, the percentage of conclusions drawn from positively-charged conditionals defeated by

negative information did not differ significantly from the percentage of conclusions drawn from negatively-charged conditionals defeated by positive information, $t(29) = 1.47$, $p = .153$, $d = 0.45$. However, since negative and positive conditionals already differed significantly in the control problems (without additional information), this finding is difficult to interpret. To control for possible confounds I compared a second time the amount of withdrawn conclusions from conditionals with conflicting additional information, but this time with a corrected measure, eliminating all problems where participants disagreed with the initial conditional (i.e., the problems where they answered "no" in the control condition). Again, no differences were found ($M = 45.83$, $SD = 31.70$ for negative conditionals with positive additional information, $M = 58.67$, $SD = 26.09$ for positive conditionals with negative additional information, $t(29) = 1.50$, $p = .145$, $d = 0.44$).

This first analysis thus shows that when the emotion described in the conditional is only mild in strength, then people do not differ in how often they defeat conclusions from positive and negatively charged conditionals. However, the statistics used do not allow the interpretation of null-effects. To test whether people's higher reluctance to defeat conclusions drawn from negatively-charged conclusion is particular for conditionals of high emotional strength, the results from Experiment 10 were compared to the ones of Experiment 9. Compared to Experiment 9, participants in Experiment 10 (corrected measure) defeated more negative conclusions by positive additional information (33.33% vs 45.83%; respectively), and less positive conclusions by negative additional information (74.71% vs. 58.67%; respectively). Consequently, the difference between negatively-charged conclusions defeated by positive information and positively-charged conclusions defeated by negative information is significantly smaller in Experiment 10 than Experiment 9, $t(57) = 2.52$, $p = .014$, $d = 0.66$. In other words, people's reluctance to defeat conclusions drawn from conditionals describing negative situations diminishes when the emotions are only of mild emotional strength.

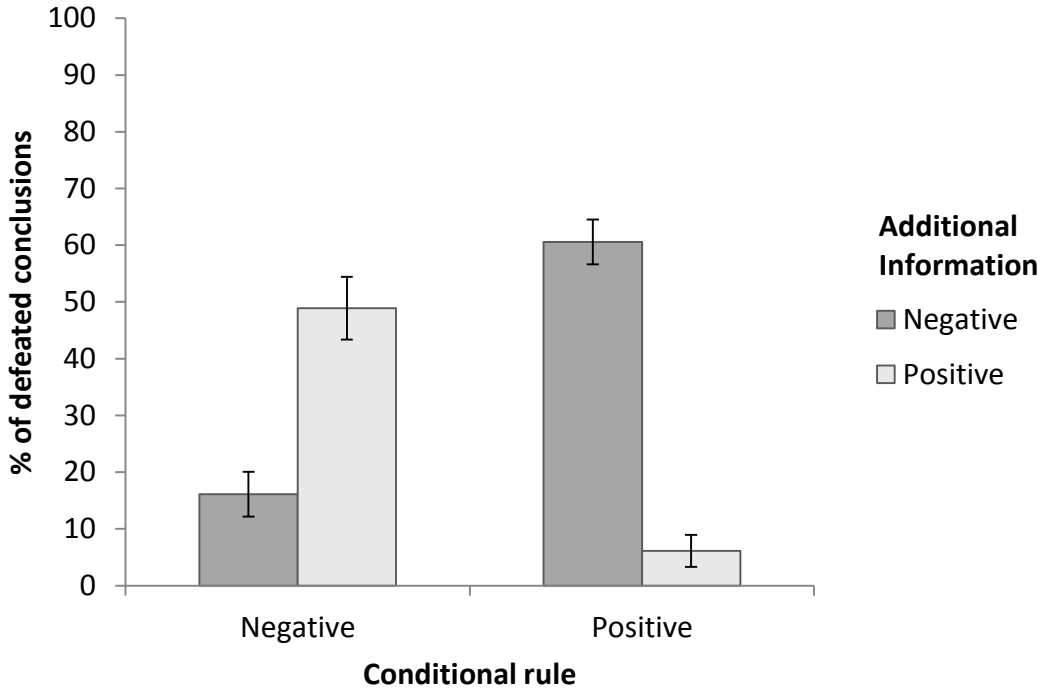

Figure 10. Percentage of defeated conclusions as a function of the valence of the conditional rule and additional information in Experiment 10. Error bars show standard errors.

6.2.3. Discussion

The results show that when the situation described in the conditional is only mild in negative emotional strength, then people do defeat initially drawn conclusions. More precisely, the amount of defeated conclusions drawn from negatively-charged conditionals does not differ anymore from the amount of defeated conclusions from conditionals describing positive situations. Compared to Experiment 9, the current results show that when the situation described in the conditional is only of mild emotional strength, people's problems in defeating negatively-charged conclusions diminishes. This pattern of results supports my hypothesis and resembles thus the pattern of results found when investigating people's legal reasoning, because here, too, participants mostly had problems excusing highly morally outraging offences, but not offences of low moral outrage.

6.3. Summary and Implications of Chapter 6

The results of Experiment 9 and 10 show similarities to the findings on legal defeasible reasoning. In legal reasoning people showed difficulties in withdrawing conclusions stating that offenders of severe offences should be punished. Similarly, in Experiment 9 people showed difficulties in withdrawing conclusions drawn from conditionals describing very negative situations. Moreover, as in legal reasoning where people's reluctance to defeat punishment conclusions was restricted to high moral outrage offences, in Experiment 10 – where only situations of mild emotional strength were presented – people's reluctance to defeasible reasoning diminished. In legal reasoning these findings were explained by arguing that people are influenced by their own sense of justice. Since people usually wish severe offences to be punished, they also endorsed such conclusions very often and refused to excuse such offences. When offences were not severe, this wish to punish offenders diminished and people engaged in defeasible reasoning. But how can the results from Experiment 9 and 10 be explained?

Differences in associative strength cannot account for the results because in both experiments positive and negative information were as strongly associated to their respective conclusions of being sad or being happy (see De Neys et al., 2003b; Markovits et al., 1998; Quinn & Markovits, 1998). As already pointed out in the introduction of this Chapter, negative information seems to be more informative than positive information (i.e, the negativity bias). But why is negative information more informative than positive one? Possible reasons for its superiority is that negative information can be more diagnostic (e.g., Skowronski & Carlston, 1987; 1989; for an overview see Lewicka, Czapinski, & Peeters, 1992), or less expected than positive information (Ikegami, 1993). Another explanation is provided by the Evaluative Space Model (ESM), which explains the negativity bias by assuming two separable underlying motivational systems functioning at different levels of evaluative input (Cacioppo & Berntson, 1999; Cacioppo et al., 1997; Cacioppo, Gardner, & Berntson, 1999; Ito & Cacioppo, 2005). It is argued that by weighing highly negative information stronger than positive information, one guarantees survival chances because strong negative events are often threat to life. ESM actually also explains why this superiority of negative information is only observable when the information is of high emotional strength. It is argued that when the emotional strength is only mild, not negative but positive information is more informative. This is called the *positivity offset* (Cacioppo et al., 1997; Ito

& Cacioppo, 2005). When the evaluative input is only of mild emotional strength, then positive information gains weight because ignoring mild negative events in favor of positive events enhances exploration behavior facilitating food finding and expansion (Cacioppo et al. 1997; 1999; Peeters & Czapinski, 1990; Vaish et al., 2008). Accordingly, other researchers as well have observed how at low levels of evaluative input (i.e., information of little emotional strength) positive information is actually weighed stronger than negative one (e.g., Amster, 1964; Anisfeld & Lambert, 1966, Boucher & Osgood, 1969; Ito et al., 1998). However, it is important to notice that in Experiment 10 positive information was not weighed stronger than negative information; no positivity offset was found. Further studies are therefore necessary to understand the cognitive mechanisms behind defeasible reasoning with emotional content outside of the domain of legal reasoning.

Chapter 7: General Discussion

The aim of this thesis was to investigate how people reason with legal rules and which factors influence their defeasible reasoning. In a series of experiments legal rules were presented as legal conditionals and embedded in inference tasks. In addition to the legal conditional and the categorical premise these tasks also included information about potential exculpatory circumstances. Exculpatory circumstances served as defeaters and were either presented explicitly as a third premise (Experiments 1, 2, 4-5), captured implicitly via preliminary studies (Experiments 6-8), or generated by participants (Experiment 3, without embedding the offences into inferences). Participants' task was to decide whether the offender described in the inference task should be punished (Experiments 1-8) or will be punished (Experiments 6-7). That is, participants had to decide whether it is appropriate to defeat the conclusion of punishing the offender in light of exculpatory circumstances. The first three experiments were designed to show that contrary to lawyers, laypeople's own sense of justice affects their willingness to withdraw conclusions from legal rules. The remaining experiments investigated what moderates the effect of one's own sense of justice: Experiments 4 and 5 by testing differences in people's attitudes towards offences and offenders, and Experiments 6 to 8 by varying the phrasing of the legal conditionals. Finally, this thesis also investigated whether the findings from Experiments 1-8 are specific to legal reasoning, or if they also have correspondence in everyday scenarios (Experiments 9 and 10).

The results show that legal reasoning is often defeasible, not only for lawyers whose defeasible reasoning followed from the defeasibility of the penal code, but also for laypeople. However, laypeople's defeasible reasoning differed from lawyers'. Laypeople's punishment decisions depended on their own sense of justice: their acceptance of exculpatory circumstances as defeaters depended on how morally outraging the offence described in the antecedent was. If the offence was highly morally outraging, it was more difficult to decide not to punish the offender than if the offence only evoked low moral outrage. In cases of high moral outrage, laypeople even had difficulties in retrieving exculpatory evidence from memory. As a consequence, differences in people's attitudes about offences and offenders were shown to also affect legal reasoning. For instance, Experiment 4 showed that highly religious persons confronted with an offence which is usually not considered morally outraging decided to punish the offender anyway. Also cultural differences affected the way people deal with their feelings of moral outrage. Experiment 5 linked moral outrage to the

acceptance of vigilantism: when offences were morally outraging, people from Peru defeated the conclusion of punishing the offender in light of vigilante circumstances more often than people from Germany. However, the effect of a person's own sense of justice on legal reasoning also depended on the way in which the conditionals were phrased. Experiments 6 and 7 showed that peoples own sense of justice affected inferences when they were asked whether the offender *should* be punished for a given offence. But when reasoners were asked whether an offender *will* be punished, the effect of one's own sense of justice was diminished allowing for consideration of exceptions. However, not all variations in the phrasing of legal conditionals affected inferences. Variations which only affected the legal conditional but not the specific question about the conclusion did not affect inferences. In Experiment 8 it did not matter whether the rule stated that *all* or *some* offenders deserve punishment. Given that the conclusion always asked whether the offender *should* be punished, conclusions only depended on reasoners' moral outrage; the original phrasing of the rule was ignored. Finally, Experiments 9 and 10 showed that people's reluctance to defeat conclusions that are evoked by negative emotions – such as moral outrage – is not limited to the legal domain. Negatively-charged conclusions were also difficult to defeat when conditionals described everyday situations: conclusions such as being sad when ones mother dies, were much harder to defeat than conclusions such as being happy when getting one's dream job.

The findings are novel in the psychological literature on reasoning. The idea of combining the fields of conditional reasoning and legal reasoning to investigate people's defeasible reasoning with legal rules is promising and has implications for psychology, law and society. However, due to its novelty there are also aspects of this thesis that deserve further discussion and need further investigation. Hence, in the following sections I report the findings in more detail and discuss their consequences and implications for cognitive psychology, law, social psychology, and society. Finally, also the limitations and further perspectives of this work are discussed.

7.1. Consequences for Cognitive Psychology

The current findings are relevant for cognitive psychology. First, they show the importance of domain knowledge for reasoning in general, and for defeasible reasoning in particular. Second, they also show that people often introduce their own preferences into reasoning tasks: factors people use in their everyday life to decide upon offenders and offences are also considered during legal reasoning. Relatedly, the current findings also show that the emotional attachment people have to the rule described in conditionals affects their inferences. In the following sections I discuss these three aspects in more detail. I start by discussing how, in legal reasoning, in absence of domain knowledge, people introduce their own preferences and attitudes to decide about the defeasibility of a given legal rule. Regarding these effects, I will also discuss how applying these preferences and attitudes when drawing conclusions can be seen as a kind of decision making, where reasoners consider the utility of punishing an offender for their society. Then I discuss how emotions affect legal reasoning and how the investigation of emotions and defeasibility can further our understanding of human cognition. Having discussed these three main aspects of legal reasoning (defeasibility, domain knowledge, and decision making), I evaluate which of the current reasoning theories best accounts for these factors and therefore best describes people's legal reasoning. Afterwards, I explain why it is important to be cautious when relating the findings of this thesis to the usual literature on deontic reasoning, and finally I discuss how this work can be related to the question about human rationality.

7.1.1. Defeasibility, Domain Knowledge, and Decision Making

The experiments presented in this dissertation show the importance of domain knowledge on defeasible reasoning. People defeat previously drawn conclusions in light of additional information, but what counts as a defeater for some is not necessary a defeater for others. When reasoners are confronted with a conditional and a potential defeater, they must decide if this defeater is strong enough to withdraw a conclusion. In everyday life – and in most of the experiments found in the literature – such a decision is easy. People are usually confronted with conditionals describing everyday situations where they have background knowledge about which defeaters are plausible and which not. However, if a conditional is *domain*

specific – such as legal conditionals – then the picture changes: inferences made by people who do have background knowledge about this specific domain differ from the inferences drawn by people who do not. If domain knowledge is available, this knowledge can be used to decide about defeaters. But, if such knowledge is not available, then reasoners must use other cues in order to decide the role assigned to a potential defeater. Simply ignoring the potential defeater is not an option because people usually assume that information uttered to them is relevant (Grice, 1975). Consequently, they cannot simply ignore potential defeaters but have to decide in which respect it may be relevant for the inference. But how? In the beginning of this thesis it was assumed that reasoners compensate their lack of knowledge by other criteria which might differ from the ones used by experts, but which are related to the domain in question. Experiments 1 to 5 showed that in legal reasoning, when asked to conclude what *should* happen to an offender, those criteria are the reasoners' own sense of justice, driven by feelings of moral outrage.

Using one's own sense of justice when reasoning with legal rules implies that people decide about defeaters according to their own preferences of what should happen to offenders. Given that legal conditionals describe in the antecedent an offence and in the consequent the corresponding punishment, people who perceive that an offence deserves punishment will endorse the corresponding conclusion to punish an offender more strongly than people who do not. Consequently, a defeater is rejected or not considered strong enough to defeat the conclusion to the extent that the reasoner considers the offence inexcusable. For instance, if people feel highly morally outraged by an offence and think that such an offence deserves punishment, then exculpatory circumstances that may be considered relevant in other situations are not considered strong enough to excuse this offence (Experiments 1-3). Similarly, if reasoners have a close relationship to the offender and do not wish this offender to be punished, then this attitude will affect the inferences they draw (Experiment 7). Or, if people are taught that offences deserve punishment irrespective of their severity, then this cultural norm will make reasoners to conclude that even harmless offences should be punished (Experiment 4). Finally, if people believe that vigilantism is an acceptable response for offences, then this attitude can bring people to consider vigilantism as a defeater (Experiment 5). All these results show that people who do not have objective measures for deciding about defeaters (i.e., domain knowledge) make inferences according to their preferences for the different outcomes a conditional can have (e.g. preferring pq over $p\neg q$; see Manktelow et al., 2000).

People's consideration of their own preferences for the different outcomes a conditional rule can have also highlights the importance of utilities in legal reasoning (see Section 1.1.2.5). Factors considered while making decisions or judging actions, for example when giving one's opinion about an offender, also influence reasoning. Just like Bonnefon (2009) and Manktelow et al. (2000) showed that utilities often influence inferences, especially in the domain of legal reasoning it is difficult and perhaps unnecessary to separate reasoning from decision making. Already in law theory the idea of illustrating legal reasoning with the help of the judicial syllogism has the main function of justifying a decision (e.g., Alexy, 1983). This is one of the reasons why in the Results sections of the reported experiments I sometimes talked about "punishment decisions". In fact, according to law theorists, legal reasoning should be considered as kind of practical reasoning (MacCormick, 1998), whose function it is to decide which actions have to be taken when an offence has been committed. This could explain why in some cases the lawyers in Experiments 1 and 2 did not defeat conclusions in light of exculpatory circumstances. Perhaps they knew that the defeater was in principle exculpatory, but nonetheless started to think about the possible negative consequences of excusing that particular offence (e.g., recidivisms). Especially laypeople's conclusions seem to follow some kind of utility maximizing strategy. Laypeople's preference to punish offenders and their reluctance to excuse offences mirrors findings from social psychology showing that punishment is often perceived as more beneficial than not punishing. For instance, it is known that in order to feel secure, people have several mechanisms primarily concerned with detaching oneself of being a victim in the future. For example, people like to believe in a just world (Lerner, 1970; Lerner & Miller, 1978) where people get what they deserve and where bad things only happen to bad people. Similarly, terror management theory (Greenberg, Pyszczynski, & Solomon, 1986) proposes that when a person's own mortality is made salient – which may have happened in the experiments through the presentation of severe offences – people feel motivated to reaffirm cultural values by for instance, endorsing higher punishment of offenders (Rosenblatt, Greenberg, Solomon, Pyszczynski, & Lyon, 1989). In that way, laypeople's punishment conclusions are of high utility, since they restore the status quo and our feeling of security. In contrast, deciding not to punish offenders would not only go against our beliefs in a just world, but would also highlight the fact that everybody is a potential victim.

However, Experiments 6 and 7 showed that people only use their own sense of justice and preferences when asked about what *should* happen to an offender. If reasoners are instead asked whether an offender *will* be punished, they also use their knowledge about how often

different offences are punished, that is, their factual knowledge. The results of Experiments 6 and 7 thus show that the way people compensate their lack of knowledge in legal reasoning – and probably also when reasoning about other domains of which they have no elaborated knowledge – depends on what reasoners think they are asked for. Utilities are important when the task asks what *should* happen, that is, when the conditional is deontic and appeals to the reasoners' attitudes and preferences. If instead the task asks about what *is* the case, that is, about what actually happens in the world, then reasoners compensate their lack of domain knowledge by their knowledge about what happens in the world. Yet, even in the case of factual conditionals people are probably not using objective measures as for example police statistics to estimate the frequencies of *pq* or *p*¬*q* cases. Instead, it makes more sense to assume that people use subjective frequencies of *pq* and *p*¬*q* cases, which might be biased by the media, social networks, or other experiences. In further studies one could thus compare reasoning performance for factual legal conditionals and test to what extent the conclusions drawn by people with and without knowledge of police statistics differ.

One question which still needs to be discussed is whether reasoners are aware of the factors used to compensate their lack of domain knowledge when reasoning with legal conditionals. In other words, do reasoners apply consciously their own sense of justice or their subjective frequencies of exceptions when reasoning? The experiments reported here do not allow a clear answer. What we should acknowledge, however, is that some experimental manipulations might have encouraged participants to reason according to their sense of justice and personal experiences. The use of own preferences or exception estimations when reasoning was probably enhanced by the fact that participants were instructed to reason as a judge (Experiment 1 and 2) or even explicitly according to their own sense of justice (Experiment 4 and 5). Contrary to most of the literature on deductive reasoning, participants were not instructed to consider the premises as true and to indicate what necessarily follows from the premises. Instead they were instructed to reason spontaneously (Experiments 6 to 8). In some experiments they were even told that no right or wrong answer exists (Experiment 8 to 10). This was done because I was interested in the natural understanding of legal conditionals. As others argued before (see Cummins, 1995; De Neys et al., 2003a; 2003b), when interested in everyday reasoning it is not necessary to instruct reasoners to assume the truth of the premises, or to respond with what follows with logical necessity from them. In everyday life people constantly draw inferences, but are never told to assume the information as true. On the contrary, people know that in real life hardly anything is necessarily true. Instead, things are true to a certain degree. Along the lines of the new psychology of

reasoning (Evans, 2012; see Section 1.1) my aim was to understand how people reason naturally with conditionals, and not to see if they reason as they should. How people *should* reason with legal rules is already regulated in the penal code. Instead, how people *do* reason is not regulated and is therefore an interesting topic of research.

In sum, the findings of this thesis show that defeasibility, domain knowledge, and decision making are closely linked. People's defeasible reasoning depends on their domain knowledge and on what reasoners think is relevant for the inference. If people think the inference task asks about what they prefer, then their own preferences are used in reasoning, and the conclusions are influenced by their attitudes about what is described in the conditional, for instance, one's own sense of justice in legal reasoning. However, if reasoners think they are asked about matters of facts, then these preferences can be ignored and the answers are according to what they think is the case in the real world. This complex interplay between domain knowledge, preferences and interpretation raise attention to theoretical oversimplifications. Reasoning theories should aim to consider all factors used in drawing conclusions. That is to say, not only those factors explicitly provided in the tasks instructions should be considered, but also those that reasoners themselves bring along, for example own preferences and personal experiences. Just as Bonnefon (2009; 2010) argued, everything we usually use during reasoning in everyday life *leaks* into the experimental setting.

7.1.2. Emotions in Defeasible Reasoning

Most research focused on how emotions influence deductive reasoning followed the standards of classical logic. It was usually found that emotions hinder people from drawing logically valid conclusions, probably because emotions impair working memory and sometimes trigger heuristic strategies (Blanchette, 2014; Blanchette & Amato, 2014; Blanchette & Richards, 2010; Quraishi & Oaksford, 2014). This thesis expands these findings by suggesting that emotions affect defeasible reasoning as well. The current findings suggest that when the antecedent and the consequent of conditionals describe something one emotionally approves, inferences are easily drawn and conclusions are difficult to defeat. This is the case not only when the content of the conditional is law-related (Experiments 1-8), but also when the antecedent describes emotional situations from everyday life and the consequent the corresponding emotion (Experiments 9-10). However, the possibility that emotions can

encourage people to draw conclusions by ignoring defeaters conflicts with the usual findings from the literature on emotions in deductive reasoning. As just mentioned, many studies showed that emotions hinder people from drawing logically valid conclusions. That is, people refuse to endorse the valid conclusions of MP or MT, and make more AC and DA fallacies. This thesis instead suggests that emotions inhibited people's consideration of defeaters, thereby enhancing their willingness to draw MP conclusions. Is it possible that emotions affect deductive reasoning and defeasible reasoning in different ways? I do not think so. A more likely explanation for these conflicting results is the methodology used in the different experiments. As Blanchette and Richards (2010) argued, studies differ in how emotions are induced (see also Gangemi, Mancini, & Johnson-Laird, 2014). Some researchers induce emotions externally, by showing a sad or a happy movie before the experiment, by conditioning the content of the conditional to one emotion, or by giving participants negative feedback about their reasoning capacities (e.g., Blanchette, 2006; Blanchette & Leese, 2010; Jung, Wranke, Hamburger, & Knauff, 2014). Results from studies inducing emotions externally are the ones which usually show that emotions hinder deductive reasoning performance (e.g., Blanchette & Campbell, 2005; see Gangemi et al., 2014). In my experiments, however, emotions were induced internally by the *content* of the conditional. The antecedent described a situation which induced an emotion (e.g., moral outrage) and the consequent described an action enhanced by this emotion (e.g., punishment), therefore triggering MP conclusions. Other researchers have also induced emotions internally through the content of the conditional and showed that those can enhance reasoning. Blanchette, Richards, Melnyk, and Lavda (2007) for instance found that people living near the place where terrorists attacked London in 2005 made less logical errors when reasoning with syllogisms concerning terrorism than people living further away from London (e.g., Some terrorist attacks are murders. All terrorist attacks are heinous. Therefore, some murders are heinous.). Similar effects were found with veterans when reasoning with war related syllogisms (Blanchette & Campbell, 2005). Emotions can therefore influence reasoning in distinctive ways, both hindering and improving performance. Depending on how the emotion is induced, the endorsement of conclusions is sometimes enhanced or inhibited (see Gangemi et al., 2014).

The fact that emotions affect reasoning differently depending on how an emotion is induced shows how complex the relationship between emotions and reasoning is. Further studies are therefore necessary to gain deeper insights into when emotions enhance reasoning performance and when not, especially in defeasible reasoning. In this thesis, emotions seem to

have enhanced valid MP inferences by inhibiting the effect of defeaters. The same could happen when, for example, people with dog phobia are asked to think about defeaters for the conditional "If a dog barks at you, then he will bite you" (cf., De Jong & Vroling, 2014). Another avenue for future research is to construct problems where emotions do not inhibit, but enhance the consideration of defeaters. One could formulate conditionals where the antecedent evokes an emotion which is not captured by the consequent. Imagine for instance you are a person with social anxiety and you are confronted with the conditional: "If I have to speak in front of a huge audience, then I can show how much I know." People with social anxiety will certainly feel negative emotions when imagining speaking in front of big audiences. These negative emotions would therefore not trigger the conclusion of showing how much one knows, but would instead trigger the conclusion of getting red or embarrassing oneself. That is, the emotion evoked by the antecedent (i.e., fear of public speaking) conflicts with the emotion implied in the consequent (i.e., proudness of showing one's skills). By triggering an opposite emotion to the one described in the conditional, defeaters can be made salient. For instance, in this last example social fears could make reasoners aware that when speaking in front of audiences one can make mistakes, have a black out, or receive negative feedback – all possible defeaters. As some researchers argue, being emotionally involved by the emotional content of conditionals can affect people's motivation to think deeper about possible alternatives or disabling conditions (cf. Gangemi et al., 2014).

Finally, another task for the future is to understand in greater detail *how* emotions influence defeasible reasoning. In other words, what are the mental processes that underlie the effect of emotions on reasoning? There are different explanations for the inhibiting effect of *externally* induced emotions (for a review see Blanchette & Richards, 2010). Some researchers explain people's errors in reasoning by arguing that some emotions impede analytic thinking and enhance heuristic thinking (e.g., Blanchette & Amato, 2014; Quraishi & Oaksford, 2014). Others suggest that externally induced emotions are irrelevant for the task, thus being a burden for working memory (e.g., Blanchette, 2014; Richards, French, Keogh, & Carter, 2000). Oaksford, Morris, Grainger, and Williams (1996), for instance, showed that secondary working memory tasks inhibited reasoning performance similarly to mood inductions, suggesting that emotions trigger the spontaneous retrieval of mood congruent memories which in turn burden working memory. Yet others say emotions hinder reasoning by moving attention away from task-relevant information (Blanchette & Richards, 2010; see also Schimmack, 2005). However, how *internally* induced emotions enhance people's drawing of MP and MT conclusions still requires further research. In legal reasoning

emotions could guide inferences by enhancing or reducing the perceived sufficiency of the antecedent for its consequent. Even though necessity and sufficiency relations were not measured in this thesis, the results suggest that an antecedent is considered highly sufficient to its consequent when the relationship described in the antecedent and the consequent corresponds to the emotions evoked by the conditional. This happens in cases of highly morally outraging offences. When the antecedent describes a highly morally outraging offence, people's feelings of moral outrage evoke a desire to punish the offender, which is described precisely in the consequent. Instead, an antecedent loses sufficiency when its consequent is not in accordance with the experienced emotions. This happens in cases of low morally outraging offences. Given that low moral outrage is only linked to a little desire to punish the offender, the corresponding consequent is not perceived as necessary. Sufficiency and necessity relations can also be used to explain the effect of emotions on reasoning with conditionals describing everyday scenarios. The antecedent of the conditional "If I fail an exam, then I am sad" is probably perceived more sufficient for its consequent than the antecedent in "If a write a bad grade, then I am sad". However, the explanation based on emotions affecting the sufficiency and necessity relations of conditionals cannot fully explain the findings from Experiments 9 and 10. In Experiments 9 and 10 negative emotions were weighed stronger than positive ones, even though the associative strength between antecedent and consequent did not differ between the two kinds of emotions. Further, sufficiency and necessity relations are also unable to explain why negative emotions are weighed stronger when emotions are strong, but not when they are mild. Alternative explanations are therefore necessary. In Chapter 6 it was argued that emotions can influence reasoning through the negativity bias: people weigh strong negative emotions more heavily than strong positive emotions. Yet, it is not clear why this negativity bias exists. The evolutionary explanation provided by ESM is tempting but probably not supported by all researchers because evolutionary explanations are often considered circular, untestable, or "just so" stories (see Siegert & Ward, 2002). Instead, the explanation of negative emotions being more diagnostic than positive ones can seem more plausible. However, diagnosticity cannot explain why the negativity bias only exists for strong emotions and not weak ones. Alternative explanations are necessary. I can imagine that with the help of physiological measures one can gain insights into why negative emotions are sometimes perceived as more important than others. Along these lines it could be interesting to analyze the role of utilities in emotional reasoning. As Blanchette (2014) argues, emotions can influence the calculation of utilities. Perhaps strong negative emotions (such as sadness or high moral outrage) affect people's utilities

calculation in such a way that endorsing MP inferences (e.g., drawing punishment conclusions) is perceived as more beneficial than withdrawing such conclusions.

7.1.3. Mental rules, Mental Models, Probabilities, or Utilities?

So far I have discussed how domain knowledge, preferences, and emotions influence the conclusions people draw or withdraw in legal reasoning. But which are the reasoning processes behind these effects? The experiments in this study were not specifically designed to test between the different competing reasoning theories in the field. It is therefore difficult to test the single assumptions of each theory with my data. What can be done instead is to weigh the pros and cons of each theory when trying to understand the different effects found in this study. This is the aim of this section.

Mental rules theories argue that people apply abstract general reasoning rules to draw inferences, and are thus often labeled syntactic theories (Knauff, 2006). In this thesis, however, laypeople's conclusions depended highly on the content of conditionals. Inferences with identical syntactic structure were rated differently depending on the content of conditionals, that is, whether it contained a highly morally outraging offence (Experiments 1-5) or whether it contained an offence with a high frequency of exceptions (Experiments 6 and 7). Mental rules theories could account for these content effects by arguing that content only affects the comprehension component and that the inference schemas act on the output of those comprehension processes (Braine & O'Brien, 1991). For instance, Braine and O'Brien (1991) argue that due to pragmatic principles certain contents can affect the semantic representation of conditionals. However, it is not clear how exactly pragmatic principles can account for people's introduction of their own preferences and attitudes on legal reasoning, or how they consider the frequencies of exceptions in their conclusions. When explaining how exceptions may affect comprehension, Braine and O'Brien (1991) refer to Byrne's (1989) study and argue that defeaters are integrated into the antecedent. But as explained in Section 1.2.1 this is not plausible due to the limitations of working memory and the impossibility of exhaustively enumerating defeaters. Consequently, mental rules theories have problems in accounting for laypeople's reasoning performance. Mental rules theories could be nonetheless used to explain lawyers' legal reasoning. Contrary to laypeople, lawyers have background knowledge about the penal code and about which circumstances count as defeaters and which

not. Their main task is therefore that of external justification (Alexy, 1983; see Section 1.2.1). That is, lawyers have to decide whether the circumstance information provided is regulated as exculpatory circumstance in the penal code. After having decided whether a fact is exculpatory or not, lawyers' actual inference process (the internal justification) requires no additional weighing, but can be described as following some clearly *prescribed rules*: if the circumstance information is not considered exculpatory according to penal code, they conclude that the offender should be punished. But if the circumstance information is considered exculpatory, they withdraw the conclusion of punishing the offender. This way one could conclude that lawyers apply something like mental rules. Perhaps not the mental rules proposed by Braine (1978; Braine & O'Brien, 1991) or Rips (1994) since those do not allow defeasible reasoning, but still some kind of abstract rules which can be applied once the structure of the inference has been extracted.

To explain laypeople's reasoning performance we need a theory which considers the semantic content of conditionals. One such candidate is mental models theory (Johnson-Laird & Byrne, 1991; 2002), which proposes that people construct mental models of the premises to derive conclusions. According to the principle of pragmatic modulation (Byrne et al., 1999; Johnson-Laird & Byrne, 2002) when the task requires, people can also include their general knowledge about the content of the task in mental models. This principle of pragmatic modulation could be used to illustrate how people reason with legal conditionals. When reasoners are confronted with a legal conditional, they could first construct the following models from the premises:

offence punishment

…

The additional information of exculpatory circumstances can be integrated in these models:

offence exculpatory circumstance punishment

If this exculpatory circumstance is stored in memory as a defeater, then reasoners should be able to construct the following explicit mental models based on their general knowledge:

exculpatory circumstance ¬punishment
¬exculpatory circumstance ¬ punishment
¬exculpatory circumstance punishment

When reasoners are confronted with exculpatory evidence as part of the inference task, the first possibility is integrated into the model of the conditional rule:

offence exculpatory circumstance ¬punishment punishment

According to the principle of pragmatic modulation general knowledge has priority over the information given in the premises, yielding the following mental model (Johnson-Laird & Byrne, 2002):

offence exculpatory circumstance ¬punishment

As a consequence, reasoners should conclude that, when an offence has been committed in exculpatory circumstances, no punishment should follow.

Mental models theory can also be used to explain why laypeople differed from lawyers in the first two experiments. The different inference patterns of laypeople and lawyers can be a result of different mental models: depending on reasoners' prior knowledge, different mental models from general knowledge are built, and different circumstances are recognized correctly as exculpatory, and used as defeaters in further inferences. This idea is strengthened by the findings of Experiment 3, which suggested that laypeople's and lawyers' different acceptance rates of exculpatory circumstances relies – at least in part – on different amounts of exculpatory circumstances stored in memory. Is it also possible to explain lawyers' legal reasoning by mental models? In my opinion mental models theory is only suitable to explain lawyers' external justification, i.e., the decision of whether specific circumstance information is a defeater or not, by comparing the given premises with their knowledge about what counts as a defeater. The internal justification – the actual reasoning process by which a punishment conclusion is reached (see Section 1.2.1) – is in my opinion better described by mental rules theories. Given that lawyers know from the penal code what to infer from a given set of law-related premises, they do not need to construct mental models to read off from them what follows. Instead, they can simply apply the rules they know: either punishing or not punishing, depending on whether defeaters are available or not. Notwithstanding this argumentation, mental models might still be applied to predict lawyers' legal reasoning in more complex and less defined legal cases. One such example are cases for which the penal code does not prescribe how precisely to decide, but different premises would have to be evaluated and weighed (e.g., in higher instances such as in the Supreme Court).

In sum, mental models theory can be used to illustrate how laypeople arrive at conclusions and how lawyers decide whether a potential exculpatory circumstance counts as a

defeater. However, there are some problems. First, mental models theory has difficulties in explaining defeasible reasoning from deontic legal conditionals. The principle of pragmatic modulation was introduced to explain defeasible reasoning from factual conditionals, and not from deontic ones. That is, pragmatic modulation is indicated to describe how single defeaters are integrated in mental models, but it is difficult to explain how laypeople's preferences and attitudes about offences and offender are represented in these models. Arguing that reasoners' own sense of justice is reflected in the amount of defeaters available in memory is not enough. The results of this thesis show that although reasoners are able to at least generate some exculpatory circumstances (Experiment 3), they nonetheless conclude that offenders should be punished, indicating that they have problems in considering these as defeaters in reasoning (see Experiment 2, 4 to 8). A second problem of mental models theory is that pragmatic modulation only describes how single possible defeaters are considered during reasoning, but not how the frequencies of exceptions are considered, i.e., the mental models built from general knowledge do not represent how often different defeaters occur or how often exceptions occur in general. This is problematic, because Experiment 6 showed that the frequency of exceptions and not the amount of defeaters best predicted inferences. It is not clear how the frequency of exceptions is computed within pragmatic modulation. It is imaginable to introduce frequencies to the mental models built from general knowledge by representing the proportion of models in which each possibility holds (cf. Johnson-Laird et al., 1999; Stevenson & Over, 1995). However, first empirical studies question the suitability of such adapted mental models (see Fernbach & Erb, 2013).

To understand how people reason with factual legal conditionals a theory is necessary which describes how frequencies of exceptions are considered. Probabilistic theories can do this through the computation of conditional probabilities. According to most probabilistic accounts, people calculate the conditional probability by assuming that *p* holds and then computing the ratio between the instances were *q* happens and those in which *q* does not happen. The higher the perceived probability of P(*pq*) is compared to P(*p*¬*q*), the higher the conditional probability is and the more likely MP conclusions are endorsed (Evans & Over, 2004; Geiger & Oberauer, 2010; see also Oaksford & Chater, 2003b; Oaksford et al., 2000). People's consideration of frequencies of exceptions in legal reasoning is thus directly captured by the conditional probability, which – as just explained – contains the frequency of *p*¬*q* cases (here: offence but no punishment). Of course further studies are necessary to compute the concrete conditional probabilities for legal conditionals. For example, neither in Experiment 6 nor 7 the frequency of *pq* cases was measured which is actually necessary to

compute conditional probabilities. Nonetheless, in my opinion, probabilistic accounts are still the most appropriate ones to predict reasoning with factual legal conditionals, because contrary to the other theories described in this thesis, they can best predict the effect of frequency of exceptions in legal reasoning.

Nevertheless, probabilistic theories have problems accounting for the conclusions drawn from deontic legal conditionals as well. On the on hand, it is difficult to explain lawyers' legal reasoning because in law a legal rule can be either applied or not applied but not more or less applied (Dworkin, 1977). In fact, in one experiment conducted by my colleagues and me, advanced law students and graduate lawyers had to indicate on a Likert-scale the degree to which an offender should be punished (similar to Experiment 6 and 7). Many lawyers reported after the experiment to have felt confused by this response scale, and mentioned that "yes" or "no" responses would be legally more appropriate. On the other hand, also laypeople's reasoning from deontic legal conditionals is difficult to explain by probabilistic approaches. As Over et al. (2004) noticed, conditional probabilities are not predictive for reasoning with deontic conditionals (see Section 1.1.1.2.2). Instead, decision theoretic calculations of people's preference of *pq* over *p*¬*q* are better predictors of their conclusions (Over et al., 2004). This was indeed corroborated by Experiments 6 and 7, where frequencies of exceptions only predicted conclusions for factual but not deontic legal conditionals. Inferences from deontic legal conditionals depended on the reasoners' sense of justice, measured by their preference of *pq* over *p*¬*q*.

People's introduction of their own preferences and attitudes when reasoning with deontic legal conditionals is directly related to utilities and hence also to decision making. The importance of utilities for legal reasoning was already discussed in the beginning of this Chapter. In short, the idea that conclusions reflect people's preference of *pq* over *p*¬*q* may well be the result of utility calculations resulting from feelings of moral outrage. Accordingly, in Experiment 6 the preference of *pq* over *p*¬*q* correlated highly significantly with moral outrage. This utility based explanation is also appropriate for the finding that laypeople decided to rely on the legal conditional rule (Experiments 1 and 2) even if they could actually think of at least one defeater (Experiment 3). In fact, this thesis' main hypothesis that people's own sense of justice influences inferences implies some kind of utility computation: when people decide to punish offenders due to moral outrage, they are probably weighing the benefits of punishing an offender more heavily than the ones of acquitting the offender. However, besides Bonnefon's folk axioms (2009), there is no concrete description of how utilities are concretely used in reasoning, and even less in defeasible reasoning. There may be

some decision theoretic explanations for the cards turned in the Wason's selection task (but see Section 7.1.4), but further studies are needed for conditional inference tasks. One solution could be to expand Bonnefon's et al. (2013) justice template to describe how utilities affect laypeople's defeasible legal reasoning. Although the justice template suggests that bad things should happen to bad people, people may refuse to conclude that something bad should happen to a person who committed an offence, if circumstance information questions the utility of this action. Imagine for instance a case of manslaughter. The legal conditional says "If a person kills another person, then the person should be punished for manslaughter". According to the justice template people should endorse the MP conclusion. However, if additional information questions the utility of punishing the offender, then people should withdraw this conclusion, for instance when it is known that the family of the offender will take revenge.

In sum, this analysis suggests that there is no single reasoning theory which can account for all the findings in this thesis. Lawyers' legal reasoning could be explained by theories based on mental rules (internal justification) or mental models (external justification), but laypeople's calls for a theory which captures the introduction of one's own preferences and utilities, as well as the consideration of the frequency of exceptions. Especially the differences found between lawyers and laypeople suggest that dual process theories may account for people's legal reasoning (Evans, 2003; 2006; Klauer et al., 2010; Verschueren et al., 2005). Given that exculpatory circumstances are regulated in the penal code, it is possible that lawyers decide about what counts as defeaters in an analytic, deductive way, whereas laypeople decide in a more heuristic way, based on their preferences for certain outcomes or utilities. This is in accordance with the automatic, intuitive nature of feelings of moral outrage (Darley, 2009). However, although the proposal that there are two different processes underlying laypeople's and lawyers' decisions is tempting, this thesis is only the tip of the iceberg unraveling these processes. For instance, it is not clear if behind these two different responses are really different cognitive processes, or perhaps the same but one of them simply being biased by emotions. It is necessary to conduct more research to fully understand how people reason with legal conditionals.

One potential avenue for further research is using non-monotonic logics for well-structured and formally clear, yet flexible psychological theories of defeasible reasoning. There are many logics that could be implemented in cognitive psychology. For instance, Stenning and van Lambalgen (2005) showed how logic constructs, e.g., the closed-world assumption, can be integrated to psychological theories to explain people's performance in

defeasible reasoning. Also Pfeifer combined insights from philosophy and psychology to derive his mental probability logic (Pfeifer & Kleiter, 2005; 2010). Pfeifer (2013) argues that the inferences people draw should be probabilistically coherent and that the uncertainty of the premises is transmitted to the conclusion in a deductive way. Another candidate could be System P by Kraus et al. (1990). System P was developed in Artificial Intelligence and provides rules that a non-monotonic consequence relationship should satisfy (Da Silva Neves et al., 2002). Another System is the Rational Closure, which includes System P together with a rational monotony postulate (Da Silva Neves et al., 2002; Kraus et al., 1990). Already Da Silva Neves and colleagues (2002) showed that System P plus Rational Monotony are appropriate to predict human non-monotonic reasoning. They nonetheless acknowledged that these postulates should be not understood as rules people apply during reasoning, but as emerging proprieties "from a spreading activation process operating directly on knowledge structures" (p. 117).

7.1.4. Legal Reasoning and Deontic Reasoning

The majority of the legal conditionals used in this study were deontic, the consequents describing what should happen to an offender. This deontic formulation was selected following the law theoretic literature on the judicial syllogism. However, until now I have purposefully not talked much about the psychological literature on deontic reasoning, nor have I tried to relate my findings to these studies. In the following I present the main tenets with respect to deontic reasoning, and explain why it may be problematic to relate the literature on deontic reasoning to the defeasibility of legal rules investigated in this thesis.

In deontic reasoning people are usually asked to make inferences about which actions are forbidden or allowed, obligatory or not obligatory given a deontic rule (Beller, 2010). Reasoners have to decide about the deontic adequacy of actions (e.g., If a person has a ticket, then the person may enter. A person has no ticket. What follows? See Beller, 2010). Which actions are permissible or not given a deontic rule are depicted in the deontic square (Figure 11; e.g, Beller, 2008b; 2010; McNamara, 2014):

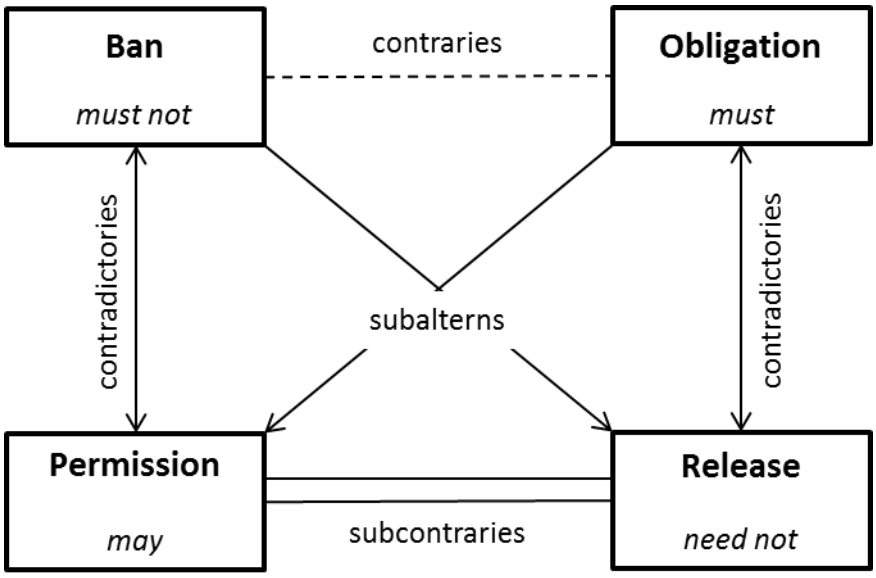

Figure 11. Deontic square. Figure based on Beller (2010).

For instance, if a deontic rule states that A obligates B ("If a person has no ticket, then the person must stay outside"), and A is the case (i.e., a person has no ticket), then B is permissible (i.e., it is permissible to stay outside without a ticket), but it is a violation of the rule not to do B (i.e, not to stay outside, that is to enter without a ticket; see Beller, 2010).

Deontic reasoning has been investigated either directly, by asking participants which actions are allowed, permissible or prohibited from given deontic conditional rule (e.g., Beller, 2008a; Bucciarelli & Johnson-Laird, 2005), or with the help of the deontic version of Wason's (1968) selection task. In the deontic selection task participants are confronted with a rule (e.g., "If a person is drinking beer, then that person must be over 18 years of age"; Cox & Griggs, 1982) and have to decide when this rule is violated. For this they must choose between four cards representing *p* (here: drinking beer), ¬*p* (here: not drinking beer), *q* (here: over 18 years), or ¬*q* (here: not over 18 years). The task is to indicate which cards have to be turned in order to test whether the rule was violated. The correct answer is to turn the *p* and ¬*q* cards, because a person under 18 years drinking beer is a case that violates the rule. Participants are usually quite good in selecting those cards; much better than when the rule is not deontic but abstract (e.g., "If there is a vowel on one side, then there is an even number on the other side"). In the abstract selection task (the one used in Wason's seminal experiment from 1968) people have to turn cards to falsify the rule. The correct cards to turn are also the *p* and ¬*q* cards, because both are necessary to falsify the two valid inferences according to

material implication: MP and MT. However, performance in the abstract selection task is very poor: reasoners turn the correct cards in only 4% of the cases (Johnson-Laird & Wason, 1970; see also Oaksford & Chater, 1995b). Several theories have been proposed to explain superior performance in the deontic compared to the abstract selection task. According to Cheng and Holyoak's (1985) pragmatic reasoning schema theory, people apply domain specific inference schemas in reasoning. Among these schemas people are though to possess permission and obligations schemas (see Holyoak & Cheng, 1995a; 1995b), which facilitate reasoners' correct resolution of the deontic selection task. Cosmides (1989) goes one step further arguing that people have some kind of cheater detection system, which originated through evolution and helps people in the selection task to pick the right cards. Also decision theoretic approaches have been developed to explain people's performance in the deontic selection task. Oaksford and Chater (1995b; 2007; 2009) argue that people decide which cards to turn according to their expected utility for violation detection. Finally, an explanation from mental models theory for the high level of performance in the deontic selection task comes from Bucciarelli and Johnson-Laird (2005). They argue that people construct mental models of the permissible situations according to the evaluated rule, and that these models help to infer which actions are forbidden.

Although tempting, linking the selection task to the experiments in this study is problematic. First, the selection task and the conditional inference tasks measure different things with different task requirements. In the selection task participants are not asked which conclusion should follow given some rule and fact. Instead, the selection task requires reasoners to either search for violations or to engage in hypothesis testing, applying Popper's falsification theory (Popper, 1959; see e.g., Oaksford & Chater, 1995a). Second, and perhaps related to these different task requirements, different factors influence performance in the two tasks. Thompson (2000) showed in a multivariate principal component analysis that performance on conditional inference tasks depends on necessity and sufficiency relations, but performance on the Wason task not. Which cards are turned in the selection task depends instead on whether the conditional is deontic or factual. Conditional inference tasks and the Wason task evoke different representations of the conditional relationship (Thompson, 2000). As a consequence, Thompson (2000) concludes that the interpretation of conditionals in not only content specific, but also task dependent.

These methodological issues are, however, not the only problems when trying to relate the findings from the literature on deontic reasoning to the findings in this thesis. The main problem is the underlying assumption of what a deontic rule represents: the deontic square

used to decide what correct deontic reasoning is, is not applicable to legal reasoning. Why so? According to the deontic square legal rules can be classified as obligation rules[6], for instance: "If a person kills another human, then the person must be punished". Now imagine the following inference:

If a person kills another human, then the person must be punished.
A person kills another human

The person needs not to be punished.

According to the deontic square, this conclusion is contradictory. However, in law it is not necessarily so: people need not to be punished after killing another human if exculpatory evidence exists. Now consider this inference:

If a person kills another human, then the person must be punished.
A person kills another human

The person is not punished.

According to the permission and obligation relations captured by the deontic square, this conclusion violates the deontic rule: If an action (here: punishing) is not taken although it is obligatory, then the rule is violated. However, in law this is not necessarily a violation. The rule of manslaughter is only violated if somebody kills another human and is not punished due to malpractice (e.g., errors in the police investigation, police corruption). But if instead a person is not punished because of exculpatory circumstances, then it is not a violation. The penal code includes several instances which actually permit cases of p and $\neg q$. Therefore, if somebody is not punished for manslaughter because of, e.g., self-defense, it is only an exception to the manslaughter rule. Legal rules are defeasible, and even though they state that some action deserves punishment, in light of exculpatory evidence it is permissible to not apply this rule anymore. Hence, defeasibility is not captured by the deontic square. Beller (e.g., Beller, 2008a; 2010) tried to find a solution to this problem by arguing that all relevant preconditions necessary for the deontic regulation should be included in the deontic rule

[6] On a first sight one could think that laws are prohibition laws, e.g., "It is prohibited to kill another human". However, the way legal rules are exposed in the penal code corresponds to obligation rules: they describe what law practitioners are obligated to do when an offence was committed.

exhaustively, or that they are exhaustively represented in the mental representation of the rule. Applied to legal reasoning this would mean that all exculpatory evidence should be captured by the antecedent of the legal conditional. However, already in Section 1.2.1 it was explained why this is impossible in legal reasoning (and in reasoning in general). Another solution could be to say that legal rules are not obligation rules but only permission rules, to be read as "If a person kills another human, then the person may be punished". Yet, this is also legally inadequate, this time because of being too weak for a legal rule. To say that a person may be punished would permit a judge not to punish offenders even when no exculpatory circumstances exist. This leaves too much room to subjectivity and thereby does not capture the spirit of legal rules. Legal rules are obligation rules, but defeasible obligation rules: they state which actions must be taken after an offence, but also allow defeating these decisions in light of exculpatory evidence. $P\neg q$ instances do not invalidate or violate the rule, but only defeat the conclusion.

In sum, the legal conditionals used in this thesis describe deontic rules, but it is difficult to relate legal reasoning to the psychological literature on deontic reasoning. First, in this thesis participants were not asked to indicate when a conditional rule was violated. Second, deontic and legal reasoning have different understandings of "correct inference", which have to do with defeasibility: the common literature on deontic reasoning does not consider the defeasibility of rules, which is crucial in law. Relating legal reasoning to the literature on deontic reasoning is thus in my opinion confusing. Maybe in future studies one could try to develop deontic logics for legal reasoning, for instance recently non-monotonic versions of deontic logic have been developed (e.g., Beirlaen & Straßer, 2013; Nute, 1997; Ryu, 1995). Future studies could also try to adapt the experimental paradigm and see how allowances and prohibitions are weighted in legal reasoning, for instance by trying to implement Oaksford and Chater's (2007) decision theoretic approach on deontic reasoning.

7.1.5. What does it mean to be rational?

One question that arises from the current investigation is whether the conclusions reached by participants were rational or not. Until now, I have abstained from using the term "error" when discussing the experiments' results. Nevertheless, what is a rational and what is an irrational conclusion in legal reasoning? From a legal point of view laypeople are clearly

doing wrong when ignoring exculpatory circumstances as reasons to void punishment. If the penal code states that a given circumstance is exculpatory and should therefore lead to the exoneration of the offender, then people who do not exonerate offenders are making "errors". However, laypeople who do ignore potential defeaters and do not withdraw the MP valid conclusions are actually reasoning deductively correct because they preserve the monotonicity of inferences; such behavior has been considered in many psychological papers rational (e.g., Johnson-Laird & Byrne, 1991; Markovits, 1986; Markovits & Vachon, 1990; Rumain et al., 1983; Taplin, 1971; see Evans, 2002 for an overview). That means: had I conducted the very same experiments in a deductive monotonic framework, then laypeople's conclusions that ignore defeaters would have been labeled correct and rational. In fact, lawyers' conclusions of not punishing offenders in light of exculpatory circumstances would have been wrong. So, who is right and who is wrong?

My suggested answer is that neither lawyers nor laypeople are wrong, they are simply using distinct criteria for reasoning: those prescribed by the penal code, and those imposed by their own sense of justice, respectively. The ambiguity of reasoning "correctly" shows how important it is to clarify beforehand what is being tested – and even more importantly – what the reasoners think they should do. As Stenning and van Lambalgen (2004; 2005; 2008) argue, before evaluating responses in light of a rational norm (e.g., be that classical logic, probabilities, or some non-monotonic logic), and declaring those right or wrong, it is essential to understand how reasoners interpreted the task (see also Thompson, 2000). Researchers ought to consider what different interpretations a task allows. As Experiments 6 and 7 showed, when reasoners interpret an inference task as asking about their deontic preferences, they reach different conclusions than when they think they have to answer according to their factual knowledge about the world. Along the same lines, researchers should also consider that people with different background knowledge might interpret problems differently. Just as laypeople and lawyers interpreted potential exculpatory circumstances differently, also other individual differences can affect interpretation. For instance, a person with training in classical logic would probably ignore content and answer according to the logical structure of a problem. I am not saying that researchers should wait until after the experiment to decide post-hoc which interpretation participants took. Instead, similar to Anderson's rational analysis (1991), researchers should carefully examine their experimental paradigm and consider the reasoners' goals, preferences, and background knowledge to predict which interpretation is most probable. Only if we can predict how a problem is being interpreted by those called to resolve it, we can say which conclusions rightly follow, and attempt to specify

how they might have been reached. Therefore, depending on how a problem has been interpreted, how consistently, in what context, and for what purpose, different conclusions may end up being right or wrong.

Due to the complexity of understanding how reasoners interpret a problem it is also difficult to equal rationality to only one logical norm. I have just described how conclusions considered rational under the assumptions of classical logic can be considered wrong from the perspective of defeasible legal reasoning, and vice versa. Moreover, even within one monotonic framework the very same inferences can be sometimes considered errors and sometimes not, depending on their context. For instance, as contained in Grice's (1975) conversational implicatures, sometimes AC and DA inferences – which are fallacies according to material implication – can be well justified, for example in conversational situations. Consider the following conditional: "If you pay me, I give you the product". Even though there may be alternatives for someone handling the product without paying (e.g., an assault), in normal conversations, it is appropriate to conclude that if somebody does not pay, they do not get the product. In fact, in law AC and DA inferences are actually valid. The legal conditional "If somebody kills another person, then the person should be punished for manslaughter" has no alternative antecedents resulting in the same consequent. If somebody does not kill another human, then the person should not be punished for manslaughter, and if somebody should be punished for manslaughter, it is because the person killed another human. Legal conditionals thus bare a biconditional interpretation.

The difficulty of classifying responses as correct or wrong, challenges the necessity of using norms in cognitive psychology. If everything depends on interpretation, then which norm can be used to describe human rationality? Some researchers have argued that it is not the task of psychologists to decide which norm should be applied to describe human reasoning (see Evans, 2012; Elqayam & Evans, 2011). Instead, psychologists should concentrate on *how* people actually reason. How people *should* reason is a concern for philosophers (Evans, 2012). This is of course a very simple solution for the problem in defining what rationality is, but I think it is not the only one. Instead of psychology giving up completely normative research, it could be more fruitful to allocate more resources to investigate how reasoners interpret problems, and which different norms may be appropriate for different interpretations. Many psychologists equal "logic" with classic propositional logic, probably for historical reasons. However, there are very different logics, especially for non-monotonic reasoning (e.g., default logic or epistemic logic). Therefore, in order to find a solution to the problem of the right norm for human reasoning, it is necessary that

psychologist collaborate with philosophers and logicians. Philosophers and logicians can suggest to psychologists various formal frameworks, whose assumptions can be tested experimentally, and psychologists can provide philosophers and logicians with empirical data about how people interpret reasoning tasks so that philosophers and logicians can suggest which logics may be the suitable norm for a given task or interpretation.

7.2. Consequences for Law

The differences between laypeople and lawyers found in this study are probably not surprising for law theorists. It is quite obvious that someone who has no elaborated knowledge about the penal code reasons differently with legal rules that somebody who does. Not for nothing lawyers have to study several years at university before obtaining the credential to apply law correctly. However, at a second sight, the results of this thesis have implications for the law theoretic debate about the moral correctness of law. In law there have been several discussions regarding the relationship between law and morals. Most law practitioners and law theorist agree that law and morals are not always the same (Hilgendorf, 2001; see also Alexy, 2008). In fact, law theorists following the school of legal positivism argue that it is necessary to distinguish and separate law from morals. They argue that law does not need a moral justification (e.g., Raz, 2009). However, this idea of justifying law by itself is often related to the injustices happening in the Third Reich. According to Radbruch (1946) the separation of law and morals allows people to carry out actions which might be legal, yet unjust according to one's morality. In other words, it allows legalizing actions which should not be carried out by moral standards. Radbruch argues that law cannot be law if it does not guaranty justice. Accordingly, he appeals to judges to refuse applying laws which are extremely unjust (i.e., morally wrong; see Alexy, 2008), because unjust laws are simply not laws. This in turn allows sanctioning persons who acted according to some clearly unmoral law, as for example the leaders of concentration camps or border guards in the German Democratic Republic (Hilgendorf, 2001). The appeal to decide against immoral laws is also defended by other law philosophers such as Hart (1949). However, Hart still recognized immoral laws as laws, but ones which need to be refused by law practitioners. To sum up, the debate about moral correctness in law shows that although law theorists recognize that law

and moral are not equivalent, it may be necessary to think about which laws are morally acceptable and which not. But what is a morally acceptable rule? Hilgendorf (2001) argues that in the past religion defined what was morally right or wrong. But since nowadays religion has lost importance, he argues that new measures of morality are necessary. In this way, empirical studies – such as this thesis – can help to answer the question about how a morally acceptable rule should be. For instance, similarly to this study (but in a more ecologically valid context) one could create scenarios where agents carry out actions which are then rated as just or unjust, or as justified or unjustified. In addition, such studies could help to uncover further discrepancies between law and morality and to discuss their societal consequences (see Section 7.4).

Another way in which this thesis is relevant for law is that it shows that psychological investigations can help to test theoretical assumptions from law theory. I showed here that the judicial syllogism, which was developed as a theoretical construct, can be tested empirically. Such experiments can thus help to test to what extent other theoretical constructs are suitable for practice as well. In this regard, Experiments 1 and 2 suggest that lawyers have no problems in applying the judicial syllogism when reasoning with legal rules. After the experiments no lawyer complained about the format of the task. On the contrary, they seemed to be familiar with reasoning about legal rules in a syllogistic way. Further, they also did not complain about the exculpatory circumstances that were presented to them, showing that defeating conclusions in light of exculpatory circumstances is something natural to them. There are plenty other law theoretical constructs besides the judicial syllogism that can be tested with psychological methods, for example the weighing of fundamental rights, something known in legal theory as *balancing*. The problems presented in this thesis were actually quite simple from a legal point of view. In order to defeat a conclusion, lawyers had only to decide whether a given circumstance information is exculpatory according to the penal code. However, when deciding between two or more fundamental rights, there are no written rules on how to decide. In principle, all fundamental rights have to be guaranteed. There is nothing like a penal code prescribing what to do when two fundamental rights are in conflict. Judges have to decide case by case which fundamental right deserves in this particular dispute more importance. For instance, imagine a case where a celebrity is photographed by a journalist during a candle light dinner. On the one hand the celebrity has the right to privacy, so he or she could prohibit the journalist to publish the picture. On the other hand the journalist has the right to freedom of the press, which allows a journalist to publish information and opinions considered relevant. How would you decide? Because there are no

concrete rules on how to decide, some legal theorists argue that such decisions cannot be made rationally (e.g., Habermas, 1992). However, Alexy (2003) argues that this is possible by proposing the *Weight Formula* (= Gewichtsformel):

$$G_{i,j} = \frac{I_i \times G_i \times S_i}{I_j \times G_j \times S_j}$$

The Weight formula considers the degree of detriment of the first fundamental right *i* and compares it to the importance of satisfying the competing fundamental right *j* (Alexy, 2003). This is done by considering the case specific (I_i and I_j) and abstract weights (G_i and G_j) of the rights in conflict, and how certain the effects of favoring or disfavoring the one or the other fundamental right are (S_i and S_j). Depending on this quotient, the first or the second right is preferred over the other. This rather abstract formula can be tested empirically in psychological investigations. For instance, one could construct cases of conflicting fundamental rights where all variables of the formula get a specific value. One could then compare participants' preferences of fundamental rights to the predictions of the Weight formula. Such an empirical investigation of the Weight Formula might play a role in the debate about the rationality of balancing.

7.3. Consequences for Social Psychology

This study is important for social psychology in several ways. On a practical level it helps to understand people's perception of justice. The literature on social justice and sentencing philosophies usually shows that people react negatively to offences, with a desire to punish offenders (e.g., Darley, 2009; Darley & Pittman, 2003; Carlsmith & Darley, 2008; Tetlock et al., 2000). People perceive that offences deserve punishment proportional to their severity (e.g., Alter et al., 2007; Carlsmith et al., 2002; Darley, 2009; Darley et al., 2000; Gromet & Darley, 2009). This thesis corroborates these findings by emphasizing the conditions under which people indeed have problems in excusing offences. In accordance with retributive sentencing philosophies, this difficulty was shown to depend on the perceived wrongfulness of the offence, even when asked to reason like real a judge (Experiment 2). Along these lines

– as already suggested in Section 7.1.1 – the reluctance to excuse offences can be also linked to the vast social psychological literature on people's attempts to have control over a stable world. For example, the literature on "Belief in a Just World" (Lerner, 1970; Lerner & Miller, 1978) shows that people are motivated to believe that everybody gets what they deserve. People therefore show a tendency to punish offenders or to derogate victims more harshly depending on how severe the offence was (see e.g., Walster, 1966). This can be linked with the findings reported in this thesis: the increased reluctance to accept exculpatory circumstances for severe offences can also be a result of the attempt to restore status quo. In sum, this thesis shows how deeply engrained people's wish to punish offenders is. This has consequences when deliberating about the adequacy of sentencing reforms in legal systems. For instance, in recent times it has been argued that retributive punishment is not always necessary: punishing because of retribution does not necessarily help offenders to regenerate and to reincorporate later to society (e.g., Bazemore, 1998). Instead, law should pursue rehabilitation and especially restauration, which helps both – the victim and the offender – to understand their respective perspectives and necessities (Wenzel, Okimoto, Feather, & Platow, 2008). However, as long as people experience the wish to punish offenders – at least for severe offences – it remains open whether the population will accept such innovative sentencing methods (for further details see Robinson & Darley, 2007). If rehabilitation and restauration are not perceived as a proper punishment, people may feel dissatisfied with their legal system. The potential consequences of such dissatisfaction are described in Section 7.4.

On a theoretical level this study helps to further understand the theory of moral outrage. For instance, Darley and Pittman (2003) proposed that mitigating circumstances, such as negligence, should lower feelings of moral outrage. They argue that mitigating circumstances usually show that offenders did not want to behave wrongfully, so that the perceived intentionality of the offence is lessened and consequently the moral wrongfulness of offences should be lowered. In this thesis, however, people had problems in considering exculpatory circumstances for severe offences. Although severity ratings were lowered by exculpatory evidence, participants still decided to punish the offender. This suggests that feelings of moral outrage impeded the consideration of exculpatory evidence as defeaters. How can this finding be integrated to the theory of moral outrage? One reason could be the way the offences and exculpatory circumstances were presented. In this thesis participants were first presented with the fact that an offence was committed and only afterwards they got the information about exculpatory circumstances. Both kinds of information were presented in separated screens, one after the other. It could thus be that the initial presentation of the

offence already evoked such strong feelings of moral outrage, that the subsequent presentation of exculpatory circumstances could not have high impact. This could be tested in further experiments where the order of the premises is reversed. If first exculpatory circumstances are presented (e.g., Bob is a person with a psychological disorder) and then the offence (Bob kills another human), then people might be more open to excuse this offence. Such primacy effects are long known in psychology (e.g., Anderson, 1965; Page & Norris, 1998) and can also be relevant in law, for instance when planning the order in which to present case relevant information during a trial (cf. Spiecker & Worthington, 2003). Moral outrage theory should therefore specify more clearly what is exactly meant when arguing that mitigating circumstances should lower feelings of moral outrage. It is not clear whether low feelings of moral outrage should actually exonerate offences or only mitigate them. Such as in this thesis, Robinson and Kurzban (2007) already suggested that even for low morally outraging offences people might expect at least some kind of mild punishment.

Another implication for social psychology, in fact also for psychology in general, is methodological. This study shows that the conditional inference task – originally designed to measure reasoning performance – can also be used in other disciplines. Here it was put to use in testing theoretical constructs from law theory (see Section 7.2), but also in measuring people's attitudes and preferences. In all experiments people's conclusions mirrored their attitudes about offenders. Moreover, Experiments 4 and 5 showed how the conditional inference task can even be used to measure cultural differences in legal reasoning. Further still, the conditional inference task could be also used to infer which knowledge structures participants have, for example by analyzing how people weigh potential defeaters. I propose to continue using the conditional inference task in other domains at the interface between psychology and the social sciences. The paradigm has many advantages over other more typical methods in social psychological research. For instance, with the conditional inference task one can systematically manipulate the conditional rule and circumstance information without the need to add unnecessary filler information, which may confound the results and make comparisons between experimental conditions difficult. One can focus on the relevant factors without adding unnecessary noise to the task. Also, if the rule, fact, circumstance, and conclusion are presented on separate screens, the conditional inference task allows the measurement of decision times, which is more cumbersome with other methodologies such as vignettes. In this thesis measuring decision times provided insights into the cognitive processes of legal reasoning. They showed that deciding contrary to feelings of moral outrage is difficult, irrespective of whether this implies to exonerate severe offences or to punish low

morally outraging offences (Experiment 2). In Experiment 4, decision times even allowed showing differences between highly and less-religious participants that were not visible in the mere responses. This may actually explain why many studies on the effect of religiosity on punishment decisions that do not measure decision times sometimes have difficulties in finding such effects (see McCullough & Worthington, 1999). The bottom line proposal is not that traditional social psychological measures should be abolished. Instead I propose that the conditional inference task could be an interesting supplement to traditional social psychological measures, especially when generating hypotheses about the factors that influence particular decisions.

7.4. Consequences for Society

The discrepancies found between law and feelings of justice discussed in the previous sections also bear on our society. The reported experiments show that investigating reasoning with legal conditionals is of interest beyond the rather abstract investigation of defeasible reasoning. This thesis' results help to understand why people often are annoyed when they hear about offenders released on parole or when they hear that offenders "only" get a hospital treatment order: when offences are of high moral outrage, laypeople have difficulties in accepting exculpatory circumstances. This difficulty does not necessarily mean that laypeople do not know that specific circumstance information may be exculpatory. Instead, this difficulty shows that people have problems in withdrawing the punishment of offenders when this withdrawal conflicts with their own sense of justice. Accordingly, in Experiment 2 laypeople accepted exculpatory circumstances as reasons for voiding punishment, but only when the offence was of low moral outrage.

The implications of laypeople rejecting situations that the penal code labels as exculpatory are problematic. The problems used in these experiments show that sometimes laypeople perceive offences differently than the legal system does, although they were fictitious and had limited external validity (see Section 7.5.5). Darley (2001; 2009) previously discussed the negative consequences these discrepancies can have for society. Besides dissatisfaction, such a mismatch between the law and people's own sense of justice can have other consequences: people might lose respect for the legal system and refuse to follow rules

they consider inappropriate (Darley, 2001). In other words, people might be willing to follow and respect a legal rule only if they perceive it as just and right. Darley (2009), for instance, describes how the prohibition laws in the beginning of the 20th century in the United States – which prohibited selling, producing, and transporting alcohol – were not followed by the people, primarily because alcohol consumption was considered as morally permissible. A similar behavior is observed when people illegally download music or stream TV shows. Downloading music illegally was used as a legal conditional in Experiments 6 and 7 and was the offence with the lowest moral outrage ratings and the lowest rule agreement ratings. This suggests that illegal download of music is not perceived as morally wrong, and could thus at least partly explain why so many people engage in such practices (Swash, 2009). Along these lines, also corruption and underground practices can result from people's perception of having to obey rules they do not consider morally necessary (Robinson & Darley, 2007). Police officers or other authorities are probably more willing to act against the law if they have not internalized as a moral duty following such rules.

Another serious consequence of people feeling unsatisfied with their legal system is vigilantism, which has been already treated in Experiment 5. Vigilantism is often linked to low trust in the legal system. The literature talks in this respect about people feeling that the police and courts are corrupt, that they do not do their jobs well and leave offenders free. This perception is not only linked to malpractice of police and courts. People can also feel low trust in police and courts if they consider that the rules of their justice system are not appropriate, e.g., too lenient (Robinson & Darley, 2007). In other words, people can also feel low trust in the legal system if they do not agree with the rules of the penal code. This implies that if the mismatches between legal system and people's sense of justice can be replicated outside the lab, then people's disagreement with exonerating offenders for severe offences can increase their sympathy for vigilantes. Interestingly, only two years after conducting the study on vigilantism (Experiment 5) a movement denominated "chapa tu choro" arose in Peru. "Chapa tu choro" is a colloquial way to say "catch your thief". This movement started as an online campaign in social networks, in response to the increasing criminality experienced in Peru in 2015. The campaign was adopted very fast by the Peruvian population (Doyle, Torres, & Judah, 2015). Within few weeks a lot of thieves were caught by their own victims or by witnesses of assaults. People even organized plans to capture thieves in flagrante. Captured thieves were insulted, humiliated, or beaten (Collyns, 2015; Roper, 2015), and in some cases delivered to the police ("Atrapan a sujeto robando," 2015). However, even though at first sight this movement was effective with respect to capturing offenders, one should not forget

that it is grounded on the idea of vigilantism, which is a crime. Even though most offenders caught by "chapa tu choro" were indeed guilty, one should not forget that everybody has a right to defend themselves and to get a just trial: everybody is innocent until the opposite is proven, but proving is not something that vigilantes invest much effort in.

In sum, it is important to continue investigating legal reasoning not only for basic research reasons, but also because it helps to gain insights in people's perception of law and their attitudes towards the legal system. We can only help improve our society if we understand how people perceive what is just, and what modulates their ways of dealing with injustices. Understanding these concepts is essential for constructing a desirable society where people do not follow its rules because of obedience, but because they perceive them to be just.

7.5. Limitations and Further Perspectives

The findings of the experiments presented in this thesis have helped to understand how people reason about legal rules, and on what their defeasible reasoning depends. In addition, they highlighted the importance of considering people's background knowledge during defeasible reasoning as well as the reasoners' attitudes and preferences about the conditional content. It is noteworthy as one of the first studies to investigate experimentally defeasible reasoning in law. As a consequence, there are still many open questions that require further work. Along these lines, the experiments presented in this thesis also have some methodological shortcomings to be considered in further research. In the following I will discuss these shortcomings and argue how these could be solved in future experiments.

7.5.1. Measuring Moral Outrage

In all experiments moral outrage was measured indirectly by preliminary studies. In online studies large samples of participants rated how morally outraged they felt by offences (N= 448 for Experiment 2 and 4, N = 568 for Experiment 5). The same was done for Experiments 9 and 10 when measuring the emotional valence of everyday conditionals. The main reason for measuring moral outrage indirectly was to avoid demand characteristics. If reasoners had

been asked in the beginning of the experiment to rate how morally outraged they feel by the different offences, then they might have become aware that the experiment was meant to measure some relation between moral outrage and reasoning. Asking participants after the experiment to rate their moral outrage was also not a good option because the problems in the experiment might have made participants aware that offences can be mitigated, which in turn could have affected moral outrage ratings.

One way to have a more direct measure of moral outrage in future experiments could be through the use of physiological measures, like electrogalvanic skin response or pupil dilatation (e.g., Blanchette & Leese, 2010; Prehn & van der Meer, 2014). Participants' electrogalvanic skin response, for instance, could be measured in preceding experimental blocks. Before resolving the inference tasks, participants could be confronted with the different offences (e.g., "A person kills another human") while measuring their electrodermal activity (EDA). One important benefit is that feelings of moral outrage would be measurable even without giving any explicit rating instruction. This way, measuring EDA has no demand characteristics because it allows recording participants' natural response to offences, without letting them note what is aimed to measure and thereby potentially leading to artifacts. A similar approach has been used by Blanchette and Leese (2010), who were interested in the in relation between skin conductance responses (SCR) and reasoning performance in a deductive inference task. The content of their conditionals was originally emotionally neutral, but after a classic conditioning trial these neutral conditionals were associated with certain emotions. Blanchette and Leese found that higher arousal to negative stimuli during the conditioning task was related to more logical errors in reasoning.

Another advantage of measuring moral outrage with EDA is that it allows testing for individual differences. People with high electrodermal responses (and therefore higher feelings of moral outrage) should be more prone to punish offenders and ignore potential defeaters. Physiological measures could therefore be an interesting addition to the literature on moral outrage, but also on the role of emotions in reasoning, especially because in most experiments emotions are only measured by self-reports, even though arousal is considered to be an important component of emotions (Blanchette & Richards, 2010). However, EDA also poses some methodological difficulties. SCR are very slow and time-lagged, making long stimulus intervals of 6 to 12 seconds necessary (Figner & Murphy, 2011). Measuring the arousal evoked by single offences would thus require long lasting pause slides between offences. This could make the experiment very tedious and participants might lose motivation

and attention. One solution would be to use block designs to measure EDA. This requires grouping offences of similar severity, but allows a more reliable measurement of SCR.

Semantic priming (e.g., Neely, 1977) is yet another reliable – though indirect – measure of moral outrage. When offences do evoke feelings of moral outrage, offences should enhance the classification of outrage related words compared to neutral words. For instance, in lexical decision tasks participants could be confronted with offences as primes, and then presented with either non-words or words which are unrelated or related to moral outrage, e.g., anger, injustice, outrage, immoral, disgust, etc. (see Rozin, Lowery, Imada, & Haidt, 1999). Depending on how fast participants classify moral outrage related words one could infer how highly morally outraged they were by the offence. Alternatively, upon presenting offences as primes, participants could be asked to classify ambiguous situations – perhaps pictures of social interactions – along different dimensions such as right/wrong, just/unjust, or even outraging/not outraging. Ratings as well as decision times could indicate how morally outraging participants were by the offence. Similar experiments have already been carried out to measure the effect of emotions on the interpretation of ambiguous stimuli, for example by the use of homophones, homographs, or ambiguous sentences (Richards & French, 1992; see also Calvo & Castillo, 1997; 2001; for a review see Blanchette & Richards, 2010).

There are many benefits of measuring moral outrage more objectively, besides the possibility of testing individual differences. For instance, one could also investigate in how far lawyers are morally outraged by offences. In this study lawyers' punishment decisions did not depend on the level of moral outrage of the offences. However, this does not mean that lawyers did not experience any emotions at all. It is not clear whether lawyers do not react emotionally to offences at all, or if they are simply able to control their emotions and inhibit their effect. For instance, Schleim, Spranger, Erk, and Walter (2011) found that even when lawyers report not to be emotionally involved when judging offences, brain regions involved in emotional processing did not differ in their activity from laypeople. Further studies including the measurement of arousal could thus help to understand what it actually means "to think like lawyers" (Goodenough, 2001, p. 41).

7.5.2. Improving the Experimental Paradigm

In this thesis defeasible reasoning was assessed in different ways. In some experiments it was measured by comparing inferences drawn from conditionals with and without exculpatory circumstances (Experiment 5, 9, and 10), or with irrelevant circumstances (Experiment 1 and 4). In other experiments defeasible reasoning was measured by comparing the amount of defeated conclusions drawn from conditionals containing differently morally outraging offences (Experiment 2 and 8) or different quantifiers (Experiment 8). Yet in others, measures of defeasible reasoning were derived from testing in how far the frequency of exceptions or people's preference of *pq* over *p*¬*q* predicted inferences (Experiment 6 and 7). All these methods were rather static. Only through comparisons between different conditions it was possible to test whether one manipulation encouraged people to defeat conclusions or not. In further studies it would be recommendable to try to measure defeasible reasoning more dynamically. For instance, one could ask participants to select a conclusion twice, first after presenting only rule and fact, and second, after introducing a potential defeater. This has several advantages. On the one hand it is a direct measure of defeasible reasoning since it shows how people actually defeat a previously drawn conclusion in light of additional information. On the other hand it helps to understand how defeaters are weighed, for instance by comparing the impact of different defeaters on the same legal conditional. This allows the calculation of the defeasible power of each defeater. The procedure is also advantageous because it allows introducing several defeaters to one problem. For instance, one could present a case of manslaughter and first present coercion as a defeater, and then – after a punishment has been conclusion – a second defeater, such as psychological disorders. This would indicate not only how high the impact of additional defeaters is, but also whether there are any order effects.

A similar method has been used by Dieussaert, Schaeken, Schroyen, and d'Ydewalle (2000). Dieussaert and colleagues constructed problems where participants first drew a conclusion on the basis of the initial conditional and a fact. Only in a second step they presented another conditional implying a defeater (cf. Byrne, 1989) and asked participants if they would change their previous conclusion in light of this new information. Dieussart's et al. (2000) results show that at least some people consider defeaters in two stages: first they draw a putative conclusion without the consideration of the disabling condition, which they then amend in light of defeaters. Manktelow et al. (2000) also used a similar procedure in

their experiments on road traffic violations (see Section 1.2.3). In one condition Manktelow and colleagues presented the information about offence category (speeding vs. drink-driving), degree of the offence (minor vs. major), and circumstance information (aggravating vs. mitigating) in serial mode. Participants indicated after each information which fine the offender deserves. Results corroborated the findings from their static version of the task.

Another interesting manipulation to the experimental paradigm would be to introduce time constraints. The literature on dual process accounts of reasoning shows that under time constraints people rely more on System 1 than on System 2 (Markovits et al., 2013). Therefore, it may be that time constraints enhance the effect of moral outrage. For instance, highly morally outraging offences could be punished to an even higher degree, whereas low morally outraging offences even less. In other words, time constraints could enhance people's usage of their own sense of justice during reasoning. Especially interesting would also be to test lawyers under time constraints. In the current thesis lawyers answered according to the penal code, but when given less time they could rely more on their own moral values. In fact, before the experiments of Chapter 3 some lawyers asked me whether they should answer as lawyers or as ordinary people. So time constraints could help to better understand the relationship between the theoretical knowledge of law and the moral understandings of lawyers.

7.5.3. Different Legal Conditionals

In the present study all legal conditionals had the same structure: the antecedent described an offence mainly from penal code, while the consequent indicated whether the offender should or will be punished. Keeping constant this structure allowed comparisons across studies. Further studies could nonetheless vary the kind of offence described in the legal conditional, as well as specify the kind of punishment an offence deserves.

7.5.3.1. Specifying the Punishment

In all experiments legal conditionals stated that the offender of a crime should (or will) be "punished". This was done following the judicial syllogism tradition from law theory (see Alexy, 1983; Bäcker, 2009; Koch & Rüßmann, 1982). However, the word "punishment" can

have different meanings: from a fine, to life imprisonment, or even to death penalty in some countries. One aim for the future could thus be to create legal conditionals where the consequent is more precise. For instance, one could pair offences with punishments which are in accordance to people's sense of justice, or which are not. If a deontic legal conditional is paired with a too lenient consequent (e.g., "If a person kills another human, then the person should be punished with a fine"), or with a too harsh one (e.g., "If a person steals, then the person should be punished with life imprisonment") people might refuse to draw MP conclusions although they are against the committed offence. The same could happen with factual legal conditionals: if the consequent does not describe what is believed to be the usual punishment for a given offence, then people might also refuse to endorse MP. For example, the factual legal conditional "If a person illegally downloads music, then the person will be punished with life sentence" is probably endorsed even less than the non-specified version "If a person illegally downloads music, then the person will be punished".

The punishment could also be specified by keeping the consequent constant, but by providing a Likert-scale which reflects the different degrees of punishment that exist. Such a Likert-scale could range from no-punishment to life imprisonment. One problem with such a Likert-scale is that it might hinder the measurement of defeasible reasoning. By having only one option for no-punishment (e.g., the left most extreme of the scale) but several options for different degrees of punishment, participants might feel artificially tempted to avoid this one extreme. It would thus be more appropriate to include such a question about how strong a punishment should follow after participants already made the choice whether punishment should/ will follow or not.

7.5.3.2. Varying the Offences

Besides specifying the kind of punishment, one could also vary the specific offences in the legal conditionals. For instance, further studies could investigate people's defeasible reasoning with rules from *civil law*. Contrary to the penal code, offences in civil code do not describe core offences against life and physical or mental integrity, but are thought to regulate private law, such as transactions between entities or companies, property, or family law. Accordingly, Robinson and Kurzban (2007) argue that only core offences evoke high desires to punish offences. Hence, it might be possible that people's consideration of defeaters with respect to the civil code is higher than for the penal code. For instance, people could be

confronted with the rule "If there was an offer and acceptance, then a valid contract was created" and the defeater "One of the parties was insane when accepting the offer" and asked whether under these circumstances a valid contract was created (example taken from the Dutch contract law, see Prakken, 2001). Also cross-cultural studies with offences from the civil code could be interesting. Given that civil law varies between cultures, one could test how intuitive their regulations are: depending on whether defeaters from civil code are considered justified in cultures where they do not exist, one could tell something about the suitability of such rules and defeaters.

One could also investigate what happens if participants have to decide about actions that are considered offences under some personal or societal norms, but not under penal or civil law. Consider abortion for example. Under German law, there are several exculpatory circumstances for abortion. But abortion is prohibited by several religions. When considering exculpatory circumstances in cases of abortion, highly religious participants should be highly reluctant to accept these exculpatory circumstances as valid exceptions from punishment. The same may be true for offences which only exist in law, but are permitted by certain religions. For example law prohibits circumcision due to religious doctrines whenever minors are not explicitly told what circumcision means (Zacharakis, 2013). In some religions, however, such practices are supported. Comparing people's reasoning about rules which only exist according to law or another norm (e.g., religion) can help to understand how the source of a rule affects defeasible reasoning. Along these lines future studies could also investigate people's legal reasoning with offences whose perceived severity differs from the amount of penalty assigned to them in the penal code. For instance, according to the penal code manslaughter is considered much more severe (not less than 5 years, up to life imprisonment) than sexual assault (not less than 2 years). However, when these offences are dramatized in movies or TV series, people's reactions are in conflict with the severity order proposed in penal code. While killing people is part of many movies, showing how somebody is raped is considered not appropriate. Consequently, both laypeople's and lawyers' reasoning about such offences should differ in predictable directions: laypeople may consider raping as more severe than manslaughter and consider thus more defeaters for the latter than for the former, whereas lawyers' severity perception and consideration of defeaters may be the other way around.

7.5.4. Expertise and Longitudinal Studies

In the first three experiments laypeople's legal reasoning was compared with lawyers'. However, these "lawyers" were in their majority advanced law students, and not graduates or highly experienced lawyers. Although this might appear at a first sight problematic, I do not think that this has severe consequences for the interpretations. The problems used in the experiments were quite simple from a legal point of view. The only challenge was to recognize a certain exculpatory circumstance correctly. Law students in Germany learn to do this in the middle of their studies, in such a way that the advanced law students tested should be familiar with the relevant details of the penal code. In fact, an initial comparison of the advanced law students and the graduate students in Experiments 1 and 2 did not show relevant significant differences. Furthermore, more experienced lawyers normally specialize in a particular legal domain (e.g., civil law), and may not remember all the details of the penal code. One exception would be for sure criminal court judges, but convincing them to come to the lab and participate in some psychological experiment would be too difficult anyway. One solution could be nonetheless to construct online studies on legal reasoning. By giving lawyers the possibility to participate in the experiment from their homes, it would be possible to test more people with expert knowledge, perhaps even people working in higher instances of the legal system. However, online studies allow for less experimental control, for example continuous concentration cannot be guaranteed. In addition, lawyers may ease their reasoning by looking up the relevant parts of the penal code.

Another way in which the effect of theoretical knowledge on legal reasoning could be tested is through longitudinal studies. It is imaginable that the effect of moral outrage on punishment decisions decreases especially in the first years of law studies. Such findings would not only corroborate my interpretations, but also help to gain more insight into the relationship between moral outrage and reasoning. For instance, it is possible that the effect of moral outrage is stronger the more uncertain participants are. By testing participants with different degrees of knowledge we can find out under which circumstances moral outrage mostly affects inferences. This can actually also be tested with laypeople as participants. In this study only participants that reported to have no elaborated knowledge of the penal code were tested. However, it is also possible to test laypeople with different degrees of knowledge, for example knowledge from the media or TV. Irrespective of whether this knowledge is correct according to the penal code, such knowledge could still enhance

reasoners' subjective feeling of certainty and thereby lower the impact of their own sense of justice on reasoning.

7.5.5. External Validity

One last limitation of this study is its external validity. By being one of the first studies combining the fields of legal reasoning and conditional reasoning it was necessary to start with a rather abstract and simple experimental paradigm. People's legal reasoning was investigated with an adapted version of the deductive paradigm from cognitive psychology. Consequently, the problems only contained sparse information: a legal conditional rule, a fact, potential exculpatory evidence, and the question about the conclusion. Even though the judicial syllogism itself contains just as little information, in real legal cases certainly more information is available. For instance, real cases contain more information about offenders and victims, such as age, criminal record, or intentions. Moreover, in actual legal cases more than one rule may be applicable, and much more circumstance information is available. It is a task for further studies to operationalize such variables experimentally in order to get better insight into people's defeasible reasoning with legal rules. One way in which this could be done is by providing participants with short background stories about the offence, offender, and victim just before presenting the legal inference task. Such a methodology has been already applied by e.g., Demeure, Bonnefon, and Raufaste (2009). An even more externally valid alternative would be to embed legal conditionals and defeaters in case reports or newspaper articles. Participants could be asked to reason about a real case, thus enhancing their motivation, allowing more case relevant information, and thereby encouraging more realistic reasoning.

7.6. Conclusion

This thesis investigated how people withdraw conclusions from legal rules. While lawyers decided about exculpatory evidence as prescribed in the penal code, laypeople's withdrawal of conclusions drawn from legal rules depended on their own sense of justice. When

laypeople were asked whether an offender should be punished, their consideration of exculpatory circumstances varied according to how morally outraging the offence was. Conclusions drawn from highly morally outraging offences were difficult to defeat, probably because of conflicts with people's desire to punish offenders. Only when laypeople were asked whether an offender *will* be punished, they were more willing to consider exceptions also for highly morally outraging offences. Along these lines also societal rules of behavior affected inferences. Depending on people's attitudes towards offences and offenders, sometimes legally irrelevant defeaters or legally impermissible situations (i.e., vigilantism) were considered as reasons to void punishment. Finally, people's reluctance to defeat conclusions one emotionally approves can also be observed outside the domain of legal reasoning, such as when reasoning about emotional situations from everyday life.

This study corroborates the complexity of human reasoning. For a long time now psychologists have focused on investigating people's competence to reason deductively from a given set of premises. Very insightful theories on human reasoning have emerged from those studies. However, conclusions are hardly ever irrefutable. In most domains of our daily lives we draw conclusions which we defeat in light of additional information. Just like in law, also when deliberating about what to conclude from the arguments of politicians, colleagues, or friends it is quite often necessary to engage in defeasible reasoning. In this process of defeasible reasoning, people make use of different kinds of information. Besides potential defeaters, this study showed that people may also take into account their preferences, attitudes and experiences during reasoning. All these factors must be considered when trying to understand and to predict human cognition. Not considering them could lead researchers to unjustifiably classify some conclusions as "errors". I use "unjustifiably" because the integration of defeaters, preferences, and experiences, that is, of both objective and subjective factors, may be so complex and coherent that, far from being an "error", it might appear more like a sign of rationality. It is certainly a task for future studies to continue investigating how people reason in context rich situations. Only this way it will be possible to develop complex enough reasoning theories to fit the complexities of human cognition.

Chapter 8: References

Adinkrah, M. (2005). Vigilante homicides in contemporary Ghana. *Journal of Criminal Justice, 33*, 413-427.

Alexy, R. (1983). *Theorie der juristischen Argumentation: Die Theorie des rationalen Diskurses als Theorie der juristischen Begründung.* Frankfurt am Main: Suhrkamp.

Alexy, R (2003). Die Gewichtsformel. In v. J. Jickeli, P. Kreutz, & D. Reuter (Eds.), *Gedächtnisschrift für Jürgen Sonnenschein* (pp.69-78). Berlin: De Gruyter Recht.

Alexy, R. (2008). On the concept and the nature of law. *Ratio Juris, 21*, 281-299.

Aloise, P. A. (1993). Trait confirmation and disconfirmation: The development of attribution biases. *Journal of Experimental Child Psychology, 55*, 177-193.

Alter, A. L., Kernochan, J., & Darley, J. M. (2007). Transgression wrongfulness outweighs its harmfulness as a determinant of sentence severity. *Law and Human Behavior, 31*, 319-335.

Amster, H. (1964). Evaluative judgment and recall in incidental learning. *Journal of Verbal Learning and Verbal Behavior, 3*, 466-473.

Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences,14*, 471-517.

Anderson, N. H. (1965). Primacy effects in personality impression formation using generalized order effect paradigm. *Journal of Personality and Social Psychology, 2*, 1-9.

Anisfeld, M., & Lambert, W. E. (1966). When are pleasant words learned faster than unpleasant words? *Journal of Verbal Learning and Verbal Behavior, 5*, 132-141.

Atrapan a sujeto robando videojuegos en Polvos Azules (2015, September 08). *El Comercio.* Retrieved from http://elcomercio.pe/lima/policiales/atrapan-gamer-robando-videojuegos-polvos-azules-noticia-1839420

Atthowe, J. M. (1960). Types of conflict and their resolution: A reinterpretation. *Journal of Experimental Psychology, 59*, 1-9.

Bäcker, C. (2009). Der Syllogismus als Grundstruktur des juristischen Begründens? *Rechtstheorie, 40,* 404-424.

Bäcker, C. (2010). Rules, principles, and defeasibility. In M. Borowski (Ed.), *On the nature of legal principles, ARSP-Beiheft 119* (pp. 79-91). Stuttgart: Franz Steiner Verlag.

Bäcker, C. (2012). *Begründen und Entscheiden. Kritik und Rekonstruktion der Alexyschen Diskurstheorie des Rechts* (2nd ed.). Baden-Baden: Nomos.

Batson, C. D., & Gray, R. A. (1981). Religious orientation and helping behavior: Responding to one's own or the victim's needs? *Journal of Personality and Social Psychology, 40,* 511-520.

Baumeister, R. F., DeWall, C. N., Vohs, K. D., & Alquist, J. L. (2010). Does emotion cause behavior (apart from making people do stupid, destructive things)? In C. R. Agnew, D. E., Carlston, W. G. Graziano, & J. R. Kelly (Eds.) *Then a miracle occurs: Focusing on behavior in social psychological theory and research* (pp. 119-136). New York: Oxford University Press.

Bazemore, G. (1998). Restorative justice and earned redemption: Communities, victims, and offender reintegration. *American Behavioral Scientist, 41*, 768-813.

Beirlaen, M., & Straßer, C. (2013). Non-monotonic reasoning with normative conflicts in multi-agent deontic logic. *Journal of Logic and Computation, 24*, 1179-1207.

Bell, V. A., Johnson-Laird, P. N., (1998). A model theory of modal reasoning. *Cognitive Science. 22*, 25-51.

Beller, S. (2008a). Deontic norms, deontic reasoning, and deontic conditionals. *Thinking & Reasoning, 14,* 305-341.

Beller, S. (2008b). Deontic reasoning squared. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 2103-2108). Austin, TX: Cognitive Science Society.

Beller (2010). Deontic reasoning reviewed: Psychological questions, empirical findings, and current theories. *Cognitive Processing, 11*, 123-132.

Benesh, S. C., & Howell, S. E. (2001). Confidence in the courts: A comparison of users and non-users. *Behavioral Sciences and the Law, 19,* 199-214.

Benson, P., Fischer, E.F., Kedron, T. (2008). Resocializing suffering: Neoliberalism, accusation, and the sociopolitical context of Guatemala´s new violence. *Latin American Perspectives, 35,* 38-58.

Black, D. (1983). Crime as social control. *American Sociological Review, 48*, 34-45.

Blanchette, I. (2006). The effect of emotion on interpretation and logic in a conditional reasoning task. *Memory & Cognition, 34*, 1112-1125.

Blanchette, I. (2014.). *Emotion and reasoning*. London: Psychology Press.

Blanchette, I. & Amato, J.-N. (2014.). Reasoning and emotion in the body. In I. Blanchette (Ed.), *Emotion and reasoning* (pp. 119-133). London: Psychology Press.

Blanchette, I., & Campbell, M. (2012). Reasoning about highly emotional topics: Syllogistic reasoning in a group of war veterans. *Journal of Cognitive Psychology, 24*, 157-164.

Blanchette, I., & Leese, J. (2010). The effect of negative emotion on deductive reasoning: Examining the contribution of physiological arousal. *Experimental Psychology, 58*, 235-246.

Blanchette, I., & Richards, A. (2010) The influence of affect on higher level cognition: A review of research on interpretation, judgement, decision making and reasoning, *Cognition & Emotion, 24*, 561-595.

Blanchette, I., Richards, A., Melnyk, L., & Lavda, A. (2007). Reasoning about emotional contents following shocking terrorist attacks: A tale of three cities. *Journal of Experimental Psychology: Applied, 13*, 47-56.

Bloechl, A. L., Vitacco, M. J., Neumann, C. S., & Erickson, S. E. (2007). An empirical investigation of insanity defense attitudes: Exploring factors related to bias. *International Journal of Law and Psychiatry, 30*, 153-161.

Bonnefon, J.-F. (2009). A theory of utility conditionals: Paralogical reasoning from decision-theoretic leakage. *Psychological Review, 116*, 888-907.

Bonnefon, J.-F (2010): Deduction from if-then personality signatures. *Thinking & Reasoning, 16*, 157-171.

Bonnefon, J.-F. (2012). Utility conditionals as consequential arguments: A random sampling experiment. *Thinking & Reasoning, 18*, 379-393.

Bonnefon, J.-F., Haigh, M., & Steward, A. J. (2013). Utility templates for the interpretation of conditional statements. *Journal of Memory and Language, 68*, 350-361.

Bonnefon, J.-F., Hilton, D. J. (2004). Consequential conditionals: Invited and suppressed inferences from valued outcomes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 28-37.

Bonnefon, J.-F., & Vautier, S. (2010). Modern psychometrics for the experimental psychology of reasoning. *Acta Psychologica Sinica, 42*, 99-110.

Bonnefon, J.-F. & Villejoubert, G. (2007) Modus Tollens, Modus Shmollens: Contrapositive reasoning and the pragmatics of negation. *Thinking & Reasoning, 13*, 207-222.

Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper, L. V. Hedges, & Jeffrey C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 221 – 235). New York: Russell Sage Foundation.

Boucher, J., & Osgood, C. E. (1969). The Pollyanna hypothesis. *Journal of Verbal Learning and Verbal Behavior, 8*, 1-8.

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: A lexical entry, reasoning program, and pragmatic principles. *Psychological Review, 98*, 182–203.

Braine, M. D. S., & O'Brien, D. P. (1998) *Mental logic*. N. J. Lawrence Erlbaum, Mahwah.

Braine, M. D. S., Reiser, B. J., & Rumain, B. (1984). Some empirical justification for a theory of natural propositional logic. In G. H. Bower (Ed.) *The psychology of learning and motivation* (Vol. 18, pp 313–371). NY: Academic Press.

Brewka, G. (1991). *Nonmonotonic reasoning: Logical foundations of commonsense.* Cambridge University Press.

Briceño-León, R., Camardiel, A., & Avila, O. (2006). Attitudes towards the right to kill in Latin American culture. *Journal of Contemporary Criminal Justice, 22*, 303-323.

Bucciarelli, M., & Johnson-Laird, P. N. (2005). Naive deontics: A theory of meaning, representation, and reasoning. *Cognitive Psychology, 50*, 159–193.

Buckholtz, J. W., Asplund, C. L., Dux, P. E., Zald, D. H., Gore, J. C., Jones, O. D., & Marois, R. (2008). The neural correlates of third-party punishment. *Neuron, 60*, 930-940.

Buckley, L., Chapman, R. L., Sheehan, M. C., & Cunningham, L. (2012) Keeping friends safe: A prospective study examining early adolescent's confidence and support networks. *Educational Studies, 38*, 373-381.

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

Byrne, R. M. J., Espino, O., & Santamaria, C. (1999). Counterexamples and the suppression of inferences. *Journal of Memory and Language, 40*, 347-373.

Cacioppo, J. T., & Berntson, G. G. (1999). The affect system: Architecture and operating characteristics. *Current Directions in Psychological Science, 8*, 133-137.

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1997). Beyond bipolar conceptualizations and measures: The case of attitudes and evaluative space. *Personality and Social Psychology Review, 1*, 3-25.

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1999). The affect system has parallel and integrative processing components: Form follows function. *Journal of Personality and Social Psychology, 76*, 839-855.

Calvo, M. G., & Castillo, M. D. (1997). Mood-congruent bias in interpretation of ambiguity: Strategic processes and temporary activation. *Quarterly Journal of Experimental Psychology, 50A*, 163-182.

Calvo, M. G., & Castillo, M. D. (2001). Selective interpretation in anxiety: Uncertainty for threatening events. *Cognition and Emotion, 15*, 299-320.

Carlsmith, K. M., & Darley, J. M. (2008). Psychological aspects of retributive justice. *Advances in Experimental Social Psychology, 40*, 193-236.

Carlsmith, K. M., Darley, J. M., & Robinson, P. H. (2002). Why do we punish? Deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology, 83*, 284-299.

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

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

Código penal (2014). *Código penal: Decreto legislativo N° 635 (08/04/1991)* [Peruvian penal code: legislative decrete N° 635 (08/04/1991)]. Lima: Jurista Editores.

Collyns, D. (2015, October 06). Peru's Facebook-driven vigilantes deal rough justice to the guilty – and innocent. *The Guardian.* Retrieved from http://www.theguardian.com/world/2015/oct/06/catch-your-thief-peru-facebook-vigilante-movement

Cook, A. (2006). *Individual vs. systemic justice: Using trust and moral outrage to predict reactions to vigilante murder* (Unpublished doctoral dissertation). University of Missouri-Columbia.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition, 31,* 187-276.

Cox, J. R., & Griggs, R. A. (1982) The effect of experience on performance in Wason's selection task. *Memory & Cognition, 10*, 496–502.

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 & Cognition, 19*, 274-282.

Da Silva Neves, R. Bonnefon, J.-F., & Raufaste, E. (2002). An empirical test of patterns for nonmonotonic inference. *Annals of Mathematics and Artificial Intelligence, 34,* 107-130.

Darley, J. M. (2001). Citizens' sense of justice and the legal system. *Current Directions in Psychological Science, 10,* 10-13.

Darley, J. M. (2009). Morality in the law: The psychological foundations of citizens' desires to punish transgressions. *Annual Review of Law and Social Science, 5*, 1-23.

Darley J. M., Carlsmith, K. M., & Robinson, P. H. (2000). Incapacitation and just deserts as motives for punishment. *Law and Human Behavior, 24*, 659-683.

Darley, J. M., Pittman, T. S. (2003). The psychology of compensatory and retributive justice. *Personality and Social Psychology Review, 7*, 324-336.

De Jong, P. J., & Vroling, M. (2014.). Better safe than sorry: Threat-confirming reasoning bias in anxiety disorders. In I. Blanchette (Ed.), *Emotion and reasoning* (pp. 44-64). London: Psychology Press.

De Neys, W. (2006a). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science, 17*, 428-433.

De Neys, W. (2006b). Automatic-heuristic and executive-analytic processing in reasoning: Chronometric and dual task considerations. *Quarterly Journal of Experimental Psychology, 59*, 1070–1100.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2002). Causal conditional reasoning and semantic memory retrieval: A test of the semantic memory framework. *Memory & Cognition, 30*, 908-920.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2003a). Inference suppression and semantic memory retrieval: Every counterexample counts. *Memory & Cognition, 31*, 581-595.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2003b). Causal conditional reasoning and strength of association: The disabling condition case. *European Journal of Cognitive Psychology, 15*, 161-176.

Delgrande, J. P. (1987). A first-order conditional logic for prototypical properties. *Artificial Intelligence, 33*, 105-130.

Demeure, V., Bonnefon, J.-F., & Raufaste, E. (2009). Politeness and conditional reasoning: Interpersonal cues to the indirect suppression of deductive inferences. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 35*, 260-266.

Dennett, D. C. (1987). Cognitive wheels: The frame problem of AI. In Z. W. Pylyshyn (Ed.), *The robot's dilemma* (pp. 41-64).Norwood, NJ: Ablex Publishing Corporation.

Dewitz, S. K., Ryu, Y., & Lee, R. M. (1994). Defeasible reasoning in law. *Decision Support Systems, 11*, 133-155.

Dieussaert, K., De Neys, W., & Schaeken, W. (2005). Suppression and belief revision, two sides of the same coin? *Psychologica Belgica, 45*, 29-46.

Dieussaert, K., Schaeken, W., Schroyens, W., & d'Ydewalle, G. (2000): Strategies during complex conditional inferences. *Thinking & Reasoning, 6*, 125-160.

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

Doyle, E., Torres, G., & Judah, S. (2015, September 14). The Facebook vigilantes catching thieves - and punishing them. *BBC Trending.* Retrieved from http://www.bbc.com/news/blogs-trending-34224196

Dworkin, R. (1977). *Taking rights seriously.* Cambridge, Mass.: Harvard University Press.

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

Elqayam, S., & Evans, J. St. B. T. (2011). Subtracting "ought" from "is": Descriptivism versus normativism in the study of human thinking. *Behavioral and Brain Sciences, 34*, 233-290.

Elqayam, S., Thompson, V. A., Wilkinson, M. R., Evans, J. S., & Over, D. E. (2015). Deontic introduction: A theory of inference from is to ought. *Journal of Experimental Psychology: Learning, Memory and Cognition, 41*, 1516-1532.

Evans, J. St. B. T. (1977). Linguistic factors in reasoning. *Quarterly Journal of Experimental Psychology, 29*, 297-306.

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

Evans, J. St. B. T. (2012). Questions and challenges for the new psychology of reasoning. *Thinking & Reasoning, 18*, 5-31.

Evans, J. St. B. T., Clibbens, J., & Rood, B. (1995) Bias in conditional inference: Implications for mental models and mental logic, *The Quarterly Journal of Experimental Psychology Section A, 48*, 644-670.

Evans, J. St. B. T., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning, 11*, 382-389.

Evans, J. St. B. T., Handley, S. J., Neilens, H., & Over, D. E. (2010). The influence of cognitive ability and instructional set on causal conditional inference. *The Quarterly Journal of Experimental Psychology, 63*, 892-909.

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*, 321-355.

Evans, J. St. B. T., Neilens, H., Handley, S. J., & Over, D. E. (2008). When can we say 'if'? *Cognition, 108*, 100-116.

Evans, J. St. B. T., Newstead, S. E., & Byrne, R. (1993). *Human reasoning: The psychology of deduction*. Hove (UK): Lawrence Erlbaum Associates.

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

Evans, J. St. B. T, & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8*, 223-241.

Evans, J. S. B. T., Thompson, V., A., & Over, D. E. (2015). Uncertain deduction and conditional reasoning. *Frontiers in Psychology, 6*. doi: 10.3389/fpsyg.2015.00398

Evans, J. St. B. T, & Twyman-Musgrove, J. (1998). Conditional reasoning with inducements and advice. *Cognition, 69*, B11–B16.

Feather, N. T. (1964). Acceptance and rejection of arguments in relation to attitude strength, critical ability, and intolerance of inconsistency. *Journal of Abnormal and Social Psychology, 69*, 127-136.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415*, 137-140.

Fehr, R., Gelfand, M. J., & Nag, M. (2010). The road to forgiveness: A meta-analytic synthesis of its situational and dispositional correlates. *Psychological Bulletin, 136*, 894-914.

Ferguson, H. J., & Sanford, A. J. (2008). Anomalies in real and counterfactual worlds: An eye-movement investigation. *Journal of Memory and Language, 58*, 609–626.

Fernbach, P. M., & Erb, C. D. (2013). A quantitative causal model theory of conditional reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 1327–1343.

Figner, B., & Murphy, R. O. (2011). Using skin conductance in judgment and decision making research. In M. Schulte-Mecklenbeck, A. Kuehberger, & R. Ranyard (Eds.), *A handbook of process tracing methods for decision research* (pp. 163-184). New York, NY: Psychology Press.

Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. *Journal of Personality and Social Psychology, 38*, 889-906.

Fodor, J. A. (1987). Modules, frames, fridgeons, sleeping dogs, and the music of the spheres. In Z. W. Pylyshyn (Ed.), *The robot's dilemma* (pp. 139-149).Norwood, NJ: Ablex Publishing Corporation.

Gangemi, A., Mancini, F., & Johnson-Laird, P. N. (2014.). Emotion, reasoning, and psychopathology. In I. Blanchette (Ed.), *Emotion and reasoning* (pp. 44-64). London: Psychology Press.

Geiger, S. M., & Oberauer, K. (2007). Reasoning with conditionals: Does every counterexample count? It's frequency that counts. *Memory and Cognition, 35*, 2060-2074.

Geiger, S. M., & Oberauer, K. (2010). Towards a reconciliation of mental model theory and probabilistic theories of conditionals. In M. Oaksford & N. Chater (Eds.), *Cognition and conditionals: Probability and logic in human thinkin*g (pp. 289-307). Oxford: Oxford University Press.

Georg, C. (1995). The endorsement of the premises: Assumption-based or belief-based reasoning. *British Journal of Psychology, 86*, 93-111.

Georg, C. (1997). Reasoning from uncertain premises. *Thinking & Reasoning, 3*, 161-189.

Geteiltes Echo auf Kachelmann-Freispruch (2011, May 31). *Zeit Online/ dpa.* Retrieved from http://www.zeit.de/news-052011/31/iptc-bdt-20110531-204-30677846xml

Godden, D. M., & Walton, D. (2008). Defeasibility in judicial opinion: Logical or procedural? *Informal Logic, 28*, 6-19.

Goldstein, D. M. (2003). In our hands: Lynching, justice, and the law in Bolivia. *American Ethnologist, 30*, 22-43.

Goodenough, O. R. (2001). Mapping cortical areas associated with legal reasoning and moral intuition. *Jurimetrics, 41*, 429-442.

Grasmick, H. D., Kinsey, K., & Cochran, J. K. (1991). Denomination, religiosity and compliance with the law: A study of adults. *Journal for the Scientific Study of Religion, 30*, 99-107.

Green, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293*, 2105-2108.

Greenberg, J., Pyszczynski, T., & Solomon, S. (1986). The causes and consequences of a need for self-esteem: A terror management theory. In R. E Baumeister (Ed.), *Public self and private self* (pp. 189-212). New York: Springer-Verlag.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41-58). New York: Academic Press.

Groefsema, M (1995). Can, may, must and should: A relevance theoretic account. *Journal of Linguistics, 31*, 53-79.

Gromet, D. M., & Darley, J. M. (2009). Retributive and restorative justice: Importance of crime severity and shared identity in people's justice responses. *Australian Journal of Psychology, 61*, 50-57.

Haas, N. E., de Keijser, J. W., & Bruinsma, G. J. N. (2012). Public support for vigilantism: an experimental study. *Journal of Experimental Criminology, 8*, 387-413.

Haas, N. E., de Keijser, J. W., & Bruinsma, G. J. N. (2014). Public support for vigilantism, confidence in police and police responsiveness. *Policing and Society: An International Journal of Research and Policy, 24*, 224-241.

Habermas, J. (1992). *Faktizität und Geltung: Beiträge zur Diskurstheorie des Rechts und des demokratischen Rechtsstaat*. Frankfurt a. M: Suhrkamp.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108,* 814-834.

Haidt, J. (2007). The new synthesis in moral psychology. *Science, 316*, 998-1002.

Hans, V. P. (1986). An analysis of public attitudes towards the insanity defense. *Criminology, 24*, 393-414.

Hansen, D. E., Vandenberg, B., & Patterson, M. L. (1995). The effects of religious orientation on spontaneous and nonspontaneous helping behaviors. *Personality and Individual Differences, 19,* 101-104.

Hardy, S. A., & Carlo, G. (2005). Religiosity and prosocial behaviours in adolescence: The mediating role of prosocial values. *Journal of Moral Education, 34,* 231-249.

Harnischfeger, J. (2003). The Bakassi boys: Fighting crime in Nigeria. *Journal of Modern African Studies, 41,* 23-49.

Hart, H. L. A. (1949). The ascription of responsibility and rights. *Proceedings of the Aristotelian Society, New Series, 49*, 171-194.

Hilgendorf, E. (2001). Recht und Moral. *Aufklärung und Kritik, 1*, 72-90.

Hilton, D. J., Jaspars, J. M. F., Clarke, D. D. (1990). Pragmatic conditional reasoning: Context and content effects on the interpretation of causal assertion. *Journal of Pragmatics, 14,* 791-812.

Holyoak, K. J., & Cheng, P. W. (1995a). Pragmatic reasoning with a point of view. *Thinking & Reasoning, 1,* 289-313.

Holyoak, K. J., & Cheng, P. W. (1995b). Pragmatic reasoning with a point of view: A response. *Thinking & Reasoning, 1,* 373-388.

Horn, K. (2014). *Die Rolle von Ausnahmen und Emotionalität bei deduktiven Schlussfolgerungsprozessen im rechtlichen Kontext* (Unpublished Bachelor‘s Thesis). Justus-Liebig-Universität Gießen, Gießen.

Huggins, M. K. (Ed.). (1991). *Vigilantism and the state in modern Latin America: Essays on extralegal violence.* New York: Praeger Publishers.

Ikegami, T. (1993). Positive-negative asymmetry of priming effects on the impression Formation. *European Journal of Social Psychology, 23*, 1-16.

Ito, T. A., & Cacioppo, J. T. (2005). Variations on human universal: Individual differences in positivity offset and negativity bias. *Cognition and Emotion, 19*, 1-26.

Ito, T. A., Cacioppo, J. T., & Lang, P. J. (1998). Eliciting affect using the International Affective Picture System: Bivariate evaluation and ambivalence. *Personality and Social Psychology Bulletin, 24*, 855–879.

Janveau-Brennan, G., & Markovits, H. (1999). The development of reasoning with causal conditionals. *Developmental Psychology, 35*, 904-911.

Johnson-Laird, P. N. (1978). The meaning of modality. *Cognitive Science, 2*, 17-26.

Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in Cognitive Sciences, 5*, 434-442.

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

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A Theory of meaning, pragmatics, and inference. *Psychological Review, 109*, 636-678.

Johnson-Laird, P. N., Khelmani, S. S., & Goodwin, G. P. (2015). Logic, probability, and human reasoning. *Trends in Cognitive Science, 19*, 201-214.

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

Johnson-Laird, P.N., & Wason, P.C. (1970). A theoretical analysis of insight into a reasoning task. *Cognitive Psychology, 1*, 134-148.

Juan, I. R., & Haley, H. B. (1970). High and low levels of dogmatism in relation to personality, intellectual, and environmental characteristics of medial students. *Psychological Reports, 26,* 535-544.

Jung, N., Wranke, C., Hamburger, K., & Knauff, M. (2014). How emotions affect logical reasoning: Evidence from experiments with mood-manipulated participants, spider phobics, and people with exam anxiety. *Frontiers in Emotion Science, 5,* 1-12.

Jüttner, J. (2011, May 31). Urteil in Mannheim: Im Zweifel für Kachelmann. *Spiegel Online.* Retrieved from http://www.spiegel.de/panorama/justiz/urteil-in-mannheim-im-zweifel-fuer-kachelmann-a-765928.html

Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. In K. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 267-294). Cambridge, UK: Cambridge Univ. Press

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

Kahneman, D, & Tversky, A. (1983). *Choices, values, and frames.* American Psychologist, 39, 341-350.

Kanouse, D. E., & Hanson, L. R. (1972). Negativity in evaluations. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 47–62). Morristown, NJ: General Learning.

Karremans, J. C., Van Lange, P. A. M., & Holland, R. W. (2005). Forgiveness and its association with prosocial thinking, feeling, and doing beyond the relationship with the offender. *Personality and Social Psychology Bulletin, 31,* 1315-1326.

Keller, L. B., Oswald, M. E., Stucki, I., & Gollwitzer, M. (2010). A closer look at an eye for an eye: Laypersons‘ punishment decisions are primarily driven by retributive motives. *Social Justice Research, 23,* 99-116.

Kilpatrick, S. G, Manktelow, K. I., & Over, D. E. (2007): Power of source as a factor in deontic inference. *Thinking & Reasoning, 13,* 295-317.

Kilpatrick, D. G., Sutker, L. W., & Sutker, P. B. (1970). Dogmatism, religion and religiosity, a review and re-evaluation. *Psychological Reports, 26,* 15-22.

Kirchner, C. (2013). *Der Einfluss von Religion und Zusatzinformation auf das Denken im rechtlichen Kontext* (Unpublished Bachelor‘s Thesis). Justus-Liebig-Universität Gießen, Gießen.

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

Knauff, M. (2006). Deduktion und logisches Denken. In J. Funke (Ed.), *Enzyklopädie der Psychologie, Themenbereich C, Bd. 8, Denken und Problemlösen* (pp. 167-264). Göttingen: Hogrefe.

Koch, H.-J., & Rüßmann, H. (1982). Juristische Begründungslehre: *Eine Einführung in die Grundprobleme der Rechtswissenschaft*. München: Beck.

Kraus, S., Lehmann, D., & Magidor, M. (1990). Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence, 44,* 167-207.

Kravitz, D. A., Cutler, B. L., & Brock, P. (1993). Reliability and validity of the original and revised legal attitudes questionnaire. *Law and Human Behavior, 17,* 661-677.

Kruglanski, A. (2013). Only One? The default interventionist perspective as a unimodel – commentary on Evans & Stanovich. *Perspectives on Psychological Science, 8,* 242-247.

Laurin, K., Shariff, A. F., Henrich, J., & Kay, A. C. (2012). Outsourcing punishment to God: Beliefs in divine control reduce earthly punishment. *Proceedings of the Royal Society B, 279,* 3272-3281.

Leiner, D. J. (2014). *SoSci Survey (Version 2.4.00-i)* [Computer Software]. Available from https://www.soscisurvey.de

Lerner, M. J. (1970). The desire for justice and reactions to victims. In J. Macaulay & L. Berkowitz (Eds.), *Altruism and helping behavior.* New York: Academic Press.

Lerner, M. J., & Miller, D. T. (1978). Just world research and the attribution process: Looking back and ahead. *Psychological Bulletin, 85*, 1030-1051.

Lewicka, M., Czapinski, J., & Peeters, G. (1992). Positive-negative asymmetry or 'When the heart needs a reason'. *European Journal of Social Psychology, 22*, 425-434.

Liu, I.-m. (2003). Conditional reasoning and conditionalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 694-709.

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.

MacCormick, N. (1998). Legal reasoning and interpretation. In E. Craig (Ed.), *Routledge encyclopedia of philosophy* (Volume 5, pp. 525-531). London: Routledge.

Manktelow, K. I., & Fairley, N. (2000). Superordinate principles in reasoning with causal and deontic conditionals. *Thinking & Reasoning, 6*, 41-65.

Manktelow, K. I., Fairley, N., Kilpatrick, S. G., & Over, D. E. (2000). Pragmatics and strategies for practical reasoning. In W. Schaeken, G. De Vooght, A. Vandierendonck, & G. d'Ydewalle (Eds.), *Deductive reasoning and strategies* (pp. 111-130). New York: Routledge.

Manktelow, K. I., & Over, D. E. (1991). Social rules and utilities in reasoning with deontic conditionals. *Cognition, 39*, 85-105.

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

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

Markovits, H. (1988) Conditional reasoning, representation, and empirical evidence on a concrete task. *The Quarterly Journal of Experimental Psychology, 40A,* 483-495.

Markovits, M. (2000). A mental model analysis of young children's conditional reasoning with meaningful premises. *Thinking & Reasoning, 6,* 335-347.

Markovits, H., & Barrouillet, P. (2002). The development of conditional reasoning: A mental model account. *Developmental Review, 22,* 5-36.

Markovits, H., Brunet, M.-L., Thompson, V., & Brisson, J. (2013). Direct evidence for a dual process model of deductive inference. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 39,* 1213-1222.

Markovits, H., Fleury, M.-L., Quinn, S., & Venet, M. (1998). The development of conditional reasoning and the structure of semantic memory. *Child Development, 69,* 742-755.

Markovits, H., Forgues, H. L., & Brunet, M.-L. (2010). Conditional reasoning, frequency of counterexamples, and the effect of response modality. *Memory & Cognition, 38,* 485-492.

Markovits, H., & Potvin, F. (2001). Suppression of valid inferences and knowledge structures: The curious effect of producing alternative antecedents on reasoning with causal conditionals. *Memory & Cognition, 29,* 736-744.

Markovits, H., & Quinn, S. (2002). Efficiency of retrieval correlates with "logical" reasoning from causal conditional premises. *Memory & Cognition, 30,* 696-706.

Markovits, H., & Vachon, R. (1990). Conditional reasoning, representation, and level of abstraction. *Developmental Psychology, 26,* 942-951.

Mattis, J. S., Jagers, R. J., Hatcher, C. A., Lawhon, G. D., Murphy, E. J., & Murray, Y. F (2000). Religiosity, volunteerism, and community involvement among African

American men: An exploratory analysis. *Journal of community psychology, 28,* 391-406.

McCarthy, J., & Hayes, P. (1969). Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer and D. Michie (Eds.), *Machine Intelligence 4* (463-502). Edinburgh: Edinburgh University Press.

McCullough, M. E. (2001). Forgiveness: Who does it and how do they do it. *Current directions in Psychological Science, 10,* 194-197.

McCullough, M. E., Rachal, K. C., Sandage, S. J., Worthington, E. L., Jr, Brown, S. W., & Hight, T. L. (1998). Interpersonal forgiving in close relationships: II. Theoretical elaboration and measurement. *Journal of Personality and Social Psychology, 75,* 1586-603.

McCullough, M. E., & Worthington, E. L. (1999). Religion and the forgiving personality. *Journal of Personality, 67,* 1141-1164.

McNamara, P. (2014). Deontic Logic. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy.* Retrieved from http://plato.stanford.edu/archives/win2014/entries/logic-deontic

Mishler, W., & Rose, R. (2001). What are the origins of political trust?: Testing institutional and cultural theories in post-communist societies. *Comparative Political Studies, 34,* 30-62.

Murray, S. (2016). The Frame Problem. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy*. Retrieved from http://plato.stanford.edu/archives/spr2016/entries/frame-problem

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106,* 226-254.

Newstead, S. E. (1989). Interpretational errors in syllogistic reasoning. *Journal of Memory and Language, 28*, 78-91.

Newstead, S. E., Ellis, M. C., Evans, J. St. B. T., & Dennis, I. (1997): Conditional reasoning with realistic material. *Thinking & Reasoning, 3,* 49-76

Norris, C. J., Larsen, J. T., Crawford, L. E., & Cacioppo, J. T. (2011). Better (or worse) for some than others: Individual differences in the positivity offset and negativity bias. *Journal of Research in Personality, 45*, 100-111.

Nute, D. (1997). *Defeasible deontic logic*. Dordrecht: Springer-Science+Business

Oaksford, M., & Chater, N. (1991). Against logicist cognitive science. *Mind & Language, 6*, 1-38.

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

Oaksford, M., & Chater, N. (1995b). Information gain explains relevance which explains the selection task. *Cognition, 57*, 97-108.

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

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

Oaksford, M., & Chater, N. (2003b). Probabilities and pragmatics in conditional inference: Suppression and order effects. In D. Hardman & L. Macchi (Eds.), *Thinking: Psychological perspectives on reasoning, judgment and decision making* (pp. 95-122). Chichester, UK: John Wiley & Sons Ltd.

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

Oaksford, M., & Chater, N. (2009). Précis of Bayesian rationality: The probabilistic approach to human reasoning. *Behavioral and Brain Sciences, 32*, 69-84.

Oaksford, M., & Chater, N. (2013). Dynamic inference and everyday conditional reasoning in the new paradigm. *Thinking & Reasoning, 19*, 346-379.

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.

Oaksford, M., Morris, F., Grainger, B., & Williams, J. M. G. (1996). Mood, reasoning, and central executive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 476-492.

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.

Ohm, E., & Thompson, V. A. (2004). Everyday reasoning with inducements and advice. *Thinking & Reasoning, 10*, 241-272.

Ohm, E., & Thompson, V. A. (2006). Conditional probability and pragmatic conditionals: Dissociating truth and effectiveness. *Thinking & Reasoning, 12*, 257-280.

Onken, H. (2011). Lynching in Peru in the nineteenth and early twentieth centuries. In M. Berg and S. Wendt (Eds.), *Globalizing lynching history: Vigilantism and extralegal punishment from an international perspective (pp. 173-186).* New York: Palgrave Macmillian.

Osman, M. (2013). A case study: Dual-process theories of higher cognition – Commentary on Evans & Stanovich (2013). *Perspectives on Psychological Science, 8*, 248-252.

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

Over, D. E., Manktelow, K. I., & Hadjichristidis, C. (2004). Conditions for acceptance of deontic conditionals. *Canadian Journal of Experimental Psychology, 58*, 96-105.

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105*, 761-781.

Peeters, G., & Czapinski, J. (1990). Positive-negative asymmetry in evaluations: The distinction between affective and informational negativity effects. *European Review of Social Psychology, 1*, 33–60.

Pelletier, F. J., & Elio, R. (1997). What should default reasoning be, by default? *Computational Intelligence, 13*, 165–187.

Perham, N. R., & Oaksford, M. (2005). Deontic reasoning with emotional content: Evolutionary psychology or decision theory? *Cognitive Science, 29*, 681-718.

Peterson C. and Seligman M. E. P. (2004) *Character strengths and virtues: A handbook and classification.* Washington DC: American Psychological Association.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model Of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (vol. 19, pp. 123-205). New York: Academic Press.

Pfeifer, N. (2013). The new psychology of reasoning: A mental probability logical perspective. *Thinking & Reasoning, 19*, 329-345.

Pfeifer, N. & Douven, I. (2014). Formal epistemology and the new paradigm psychology of reasoning. *The Review of Philosophy and Psychology, 5*, 199-221.

Pfeifer, N., & G.D. Kleiter. 2005. Coherence and nonmonotonicity in human reasoning. *Synthese 146*, 93–109.

Pfeifer, N., and G.D. Kleiter 2010. The conditional in mental probability logic. In M. Oaksford, N. Chater (Eds.), *Cognition and conditionals: Probability and logic in human thought* (pp. 153–173). Oxford: Oxford University Press.

Politzer, G. (2007). Reasoning with conditionals. *Topoi, 26*, 79-95.

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

Pollock, J. L. (1987). Defeasible reasoning. *Cognitive Science, 11*, 481-518.

Popper, K. R. (1959). *The logic of scientific discovery.* London: Hutchinson.

Prakken, H.(1997). *Logical tools for modelling legal argument: A study of defeasible reasoning in law*. Dordrecht: Springer Science+Business Media B.V.

Prakken, H.(2001). Modelling defeasibility in law: Logic or procedure? *Fundamenta Informaticae, 48*, 253-271.

Prakken, H., & Sartor, G. (2004). The three faces of defeasibility in law. *Ratio Juris, 17*, 118-39.

Prehn, K., & van de Meer, E. (2014.). Pupil size reflects cognition emotion interactions in analogical reasoning. In I. Blanchette (Ed.), *Emotion and reasoning* (pp. 134-153). London: Psychology Press.

Quelhas, A. C., Byrne, R. M. J. (2003). Reasoning with deontic and counterfactual conditionals. *Thinking & Reasoning, 9*, 43–65.

Quinn, S., & Markovits, H. (1998). Conditional reasoning, causality and the structure of semantic memory: Strength of association as a predictive factor for content effects. *Cognition, 68*, B93-B101.

Quraishi, S., & Oaksford, M. (2014.). Emotion as an argument strategy: How induced mood affects the evaluation of neutral and inflammatory slippery slope arguments. In I. Blanchette (Ed.), *Emotion and reasoning* (pp. 95-118). London: Psychology Press.

Radbruch, G. (1946). Gesetzliches Unrecht und übergesetzliches Recht. *Süddeutsche Juristenzeitung, 1*, 105–108.

Ramsey, F. P. (1990). General propositions and causality (1929). In D. H. Mellor (Ed.), *Philosophical papers by F. P. Ramsey* (pp. 145–163). Cambridge: Cambridge University Press. (Original work published 1929)

Raz, J. (2009). *The authority of law: Essays on law and morality.* Oxford: Oxford University Press.

Richards, A., & French, C. C. (1992). An anxiety-related bias in semantic activation when processing threat/neutral homographs. *The Quarterly Journal of Experimental Psychology, 45A*, 503-525.

Richards, A., French, C. C., Keogh, E., & Carter, C. (2000). Test anxiety, inferential reasoning and working memory load. *Anxiety, Stress and Coping, 13*, 87-109.

Rips, L. J. (1994). *The psychology of proof*. Cambridge, MA: M.I.T. Press.

Robinson, P. H., & Darley, J. M. (2007). Intuitions of justice: Implications for criminal law and justice policy. *Southern California Law Review, 81*, 1-68.

Robinson, P. H., & Kurzban, R. (2007). Concordance and conflict in intuitions of justice. *Minnesota Law Review, 91*, 1829-1907.

Roper, M. (2015, September 22). Forced to stand on anthills, made to eat raw chili peppers and whipped with belts: Vigilante justice goes viral on Peru's streets as residents post Facebook videos of how they're taking matters into their own hands. *Dailymail – Mail online*. Retrieved from http://www.dailymail.co.uk/news/article-3241283/Forced-stand-anthills-eat-raw-chili-peppers-whipped-belts-Vigilante-justice-goes-viral-Peru-s-streets-residents-post-Facebook-videos-taking-matters-hands.html

Rosenblatt, A., Greenberg, J., Solomon, S., Pyszczynski, T., & Lyon, D. (1989). Evidence for terror management theory I: The effects of mortality salience on reactions to those who violate or uphold cultural values. *Journal of Personality and Social Psychology, 57*, 681-690.

Ross, C. F. J., Francis, L. J., & Craig, C. L. (2005). Dogmatism, religion, and the psychological type. *Pastoral Psychology, 53*, 483-497.

Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, and disgust) and three moral codes (community, autonomy, and divinity). *Journal of Personality and Social Psychology, 76*, 574-586.

Rumain, B., Connell, J., Braine, M. D. S. (1983). Conversational comprehension processes are responsible for reasoning fallacies in children as well as adults: If is not the biconditional. *Developmental Psycholology, 19*, 471–481.

Ryu, Y. U. (1995). Conditional deontic logic augmented with defeasible reasoning. *Data & Knowledge Engineering, 16*, 73-91.

Saroglou, V., Pichon, I., Trompete, L., Verschueren, M., & Dernelle, R. (2005). Prosocial behavior and religion: New evidence based on projective measures and peer ratings. *Journal for the Scientific Study of Religion, 44*, 323-348.

Sartor, G. (2009). *Defeasibility in Legal Reasoning* (EUI Law Working Paper No. 2009/2). Retrieved from http://hdl.handle.net/1814/10768

Schimmack, U. (2005). Attentional inference effects of emotional pictures: Threat, negativity, or arousal? *Emotion, 5*, 55-66.

Schleim, S., Spranger, T. M., Erk, S., & Walter, H. (2011). From moral to legal judgment: The influence of normative context in lawyers and other academics. *Scan, 6*, 48-57.

Schmeltzer, C., & Hilton; D. J. (2014). To do or not to do? A cognitive consistency model for drawing conclusions from conditional instructions and advice. *Thinking & Reasoning, 20*, 16-50.

Schwarz, N., & Clore, G. L. (2007). Feelings and phenomenal experiences. In A. Kruglanski & E. T. Higgins (Eds.), *Social psychology. Handbook of basic principles* (2nd ed., pp. 385-407). New York: Guilford.

Schweitzer, N. J., & Saks, M. J. (2007). The CSI effect: Popular fiction about forensic science affects the public's expectations about real forensic science. *Jurimetrics, 47*, 357-364.

Siegert, R. J., & Ward, T. (2002). Evolutionary Psychology: Origins and Criticisms. *Australian Psychologist, 37*, 20-29.

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

Singmann, H., Klauer, K. C., & Over, D. (2014). New normative standards of conditional reasoning and the dual-source model. *Frontiers in Psychology, 5*:316. doi: 10.3389/fpsyg.2014.00316

Skitka, L. J., & Houston, D. A. (2001). When due process is of no consequence: Moral mandates and presumed defandant guilt or innocence. *Social Justice Reasearch, 14*, 305-326.

Skowronski, J. J. and Carlston, D. E. (1987). Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases. *Journal of Personality and Social Psychology, 52*, 689-699.

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin, 105*, 131-142.

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

Smith, E. S. (1999). The effects of investments in the social capital of youth on political and civic behavior in young adulthood: A longitudinal Analysis. *Political Psychology, 20*, 553-580.

Sperber, D, & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Oxford, UK: Blackwell Publishers Ltd.

Spiecker, S. C., & Worthington, D. L. (2003). The influence of opening statement/ closing argument organizational strategy on juror verdict and damage awards. *Law and Human Behavior, 27*, 437-456.

Sporer, S. L., & Goodman-Delahunty, J. (2009). Disparities in sentencing decisions. In M. E. Oswald, S. Bieneck & J. Hupfeld-Hienemann (Eds.), *The social psychology of punishment of crime* (pp. 379-401). Chichester, UK: John Wiley & Sons, Ltd.

Sprecher, S., & Fehr, B. (2005). Compassionate love for close others and humanity. *Journal of Social and Personal Relationships, 22*, 629-651.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences, 23*, 645-726.

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology, 94*, 672–695.

Stenning, K., & van Lambalgen, M. (2004). A little logic goes a long way: Basing experiment on semantic theory in cognitive science of conditional reasoning. *Cognitive Science, 28*, 481-529.

Stenning, K., & van Lambalgen, M. (2005). Semantic interpretation as reasoning in nonmonotonic logic: The real meaning of the suppression task. *Cognitive Science, 29*, 919-960.

Stenning, K., & van Lambalgen, M. (2008). *Human reasoning and cognitive science.* Cambridge, MA: MIT Press.

Stenning, K., & van Lambalgen, M. (2010). The logical response to a noisy world. In M. Oaksford & N. Chater (Eds.), *Cognition and conditionals: Probability and logic in human thinking* (pp. 85-102). Oxford: Oxford University Press.

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

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

Störring, G. (1908). Experimentelle Untersuchungen über einfache Schlussprozesse. *Archiv für die gesamte Psychologie, 11*, 1-27.

Swann, W. B., Gómez, A., Buhrmester, M. D., López-Rodríguez, L., Jiménez, J., & Vázquez, A. (2014). Contemplating the ultimate sacrifice: Identity fusion channels pro-group affect, cognition, and moral decision making. *Journal of Personality and Social Psychology, 106*, 713-27.

Swash, R. (2009, January 17). Online piracy: 95% of music downloads are illegal. *The Guardian.* Retrieved from http://www.theguardian.com/music/2009/jan/17/music-piracy

Swindel, D. H., L'Abate, L. (1970). Dogmatism, and repression-sensitization. *Journal for the Scientific Study of Religion, 9*, 249-251.

Tankebe, J. (2009). Self-help, policing, and procedural justice: Ghanaian vigilantism and the rule of law. *Law & Society Review, 43*, 245-270.

Taplin, J. E. (1971). Reasoning with conditional sentences. *Journal of Verbal Learning and Verbal Behavior, 10*, 219-225.

Tetlock, P. E., Kristel, O. V., Elson, S., S., & Lerner, J. S. (2000) The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology, 78*, 853-870.

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

Thompson, V. A. (1995). Conditional reasoning: The necessary and sufficient conditions. *Canadian Journal of Experimental Psychology, 49*, 1-58.

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

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131.

Tyler, T. R. (2006). *Why people obey the law*. Princeton and Oxford: Princeton University Press.

Tyler, T. R., & Boeckmann, R. J. (1997). Three strikes and you are out, but why? The psychology of public support for punishing rule breakers. *Law & Society Review, 31*, 237-265.

Tyler, T. R., & Huo, Y, J. (2002). *Trust in the law: Encouraging public cooperation with the police and courts*. New York: Russell Sage Foundation.

Tyson, A. (2013). Vigilantism and violence in decentralized Indonesia: The case of Lombok. *Critical Asian Studies, 45*, 201-230.

United Nation Office on Drugs and Crime (2013). *Global study on homicide 2013: Trends, contexts, data.* Vienna: Author.

Vadeboncoeur, I., & Markovits, H. (1999). The effect of instructions and information retrieval on accepting the premises in a conditional reasoning task. *Thinking & Reasoning, 5*, 97-113

Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin, 134*, 383-403.

Varga, A. L., & Hamburger, K. (2014). Beyond type 1 vs. type 2 processing: The tri-dimensional way. *Frontiers in Psychology, 5, 993*. doi: 10.3389/fpsyg. 2014.00993

Verschueren, N, Schaeken, W., & d'Ydewalle, G. (2005). A dual-process specification of causal conditional reasoning. *Thinking &Reasoning, 11*, 239-278.

Vidmar, N., & Miller, D. T. (1980). Socialpsychological processes underlying attitudes towards legal punishment. *Law & Society Review, 14*, 565-602.

Walster , E. (1966). Assignment of responsibility to an accident. *Journal of Personality and Social Psychology, 3*, 73-79.

Wang, P.-H. (2004). *Defeasibility in der juristischen Begründung.* Baden-Baden: Nomos Verlagsgesellschaft.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology, 20*, 273-281.

Weidenfeld, A., Oberauer, K., & Hörnig, R. (2005). Causal and noncausal conditionals: An integrated model of interpretation and reasoning. *The Quarterly Journal of Experimental Psychology, 58A*, 1479-1513.

Wenzel, M., Okimoto, T. G., Feather, N. T., & Platow, M. J. (2008). Retributive and restorative justice. *Law and Human Behavior, 32*, 375-389.

Wolf, A. G., Rieger, S., & Knauff, M. (2012). The effects of source trustworthiness and inference type on human belief revision. *Thinking & Reasoning, 18*, 417-440.

World Values Survey Association. (2014). *World Values Survey Wave 6: 2010-2014* (data analysis tool). Retrieved from: http://www.worldvaluessurvey.org/WVSOnline.jsp

Wróblewski, J (1992). *The judicial application of law* (Ed. by Z. Bankowski & N. MacCormick). Dodrecht: Springer.

Wyer, R. S., & Hinkle, R. L. (1976). Informational factors underlying inferences about hypothetical persons. *Journal of Personality and Social Psychology, 34*, 481-495.

Young L., Cushman F., Hauser M., & Saxe R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *PNAS, 104*, 8235-8240.

Zacharakis, Z. (2013, September 25). Eltern müssen vor Beschneidung aufklären. *Zeit Online/ AFP*. Retrieved from http://www.zeit.de/politik/deutschland/2013-09/beschneidung-olg-beschluss-aufklaerung

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 guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis“ niedergelegt sind, eingehalten.“

Gießen, den 29.03.2016

_____

Lupita Estefania Gazzo Castañeda