Similarities of theoretical and practical reasoning processes – behavioral and brain imaging evidences.

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

Justus-Liebig-University Giessen
Faculty 06 – Psychology and Sports Science
Otto-Behaghel-Strasse 10F
35394 Giessen - Germany

Presented on 01/25/2009

by

Dipl.-Psych., M.A. Phil. Patrick S. Wiedenmann
born on February the 2nd, 1976
in Bamberg - Germany

First reviewer and advisor:

Prof. Dr. Markus Knauff (Experimental Psychology and Cognitive Science)

Second reviewer:

Prof. Dr. Rudolf Stark (Bender Institute of Neuroimaging)

To my parents Maria-Luise and Volker,

my sister Tonia,

and

to my love Karoline.

Acknowledgements

First of all, I thank Prof. Dr. Markus Knauff for the support and encouragement to pursue my research interests on human reasoning. Furthermore, I thank him for his contributions to the idea of utilizing the human rights articles as well as the application of the deductive inference tasks.

I also want to thank Prof. Dr. Frank Bremmer, Prof. Dr. Karl Gegenfurtner, and the graduate program ‘NeuroAct’ (DFG graduate program ‘Neural representation and action control 885/2) for insights into various research fields and topics as well as for the valuable discussions and financial support.

For support and helpful discussions concerning the fMRI experiment as well as the opportunity to conduct an fMRI study, I thank the Bender Institute of Neuroimaging and all its members. In particular, I thank Bertram Walter for his exceptional support.

I also thank my student assistant Anne Kunkel for her committed work and support, assisting in literature research and data collection. Further, I thank my diploma students Alexandra Schmoranzer (intelligence experiment), Michaela Kandl (training experiment), and Eva Schäfer (patient experiment, not reported within the current thesis, since it was only a single case study) as well as the students who wrote their research papers for their classes under my supervision. All of them showed extraordinary commitment conducting their respective experiments.

I would also like to thank Prof. Dr. Manfred Kaps and Prof. Dr. Horst Traupe as well as all the staff members from the Neurology/Neurosurgery and Neuroradiology departments of the University Clinic of Giessen for their aid and support in the patient study.

I also thank all participants and the patient taking part, thus rendering all experiments possible. For the support acquiring participants with superior intelligence, I particularly thank Martin Dressler from the “Mensa e.V.”, the members of the “Deutschen Schülerakademien”, and the members of the “Junior Akademien”.

Moreover, I would also like to thank Prof. Dr. Christof Schuster, Dr. Jochen Ranger, and Dr. Knut Drewing for their statistical advices and validation of methods and analysis used in this work.

For advice and suggestions pertaining to the English language, I thank Dr. Marcos Sanchez, Thomas Paasch, Lhai Egesdörfer, and Dr. Jennifer Leung.

Last but not least, I thank Dr. Kai Hamburger for accommodating discussions, comments, and his lingering support. Moreover, I also thank all members of the Experimental Psychology and Cognitive Science Unit, especially Christina Wranke and Dr. Leandra Bucher for their support on various topics.

My special thanks go to Marko Kaeming, who did almost impossible, extensive, and exhaustive work concerning the fMRI experiment, and who became a good friend in the meanwhile.

Finally, I thank my parents and grandparents who provided me with mental and financial support at all times, and who believe in me without any exception. I also thank my sister for the motivational impulses and refreshing periods in Austria. My final gratitude goes to my partner Karoline Pieritz, who never gave up on me and supported me all throughout, especially in the last months, even though she had enough of her own stress to deal with.

Zusammenfassung

Seit Jahrhunderten erforschen Philosophen, wie Menschen denken und welche Denkfähigkeiten sie besitzen. Außerdem versuchen sie herauszufinden, wie Menschen ihr Denken einsetzen können und sollten, um ein friedliches und fruchtbares Zusammenleben zu erreichen. Dabei entstanden die Disziplinen ‚theoretische' und ‚praktische' Philosophie, wobei sich die erste hauptsächlich mit Epistemologie, und die zweite mit Ethik und Moral beschäftigt. Philosophen wie Kant entwickelten ihre Theorien über praktisches Denken oft aufbauend auf ihren Theorien zu theoretischem Denken, und präsentierten so relativ ‚all-umfassende' Theorien. Im letzten Jahrhundert begannen auch die Naturwissenschaften praktische und theoretische Denkprozesse des Menschen zu erforschen. Im Unterschied zur philosophischen Forschung, in der oft ‚Gedankenexperimente' eingesetzt wurden, nutzten psychologische oder neurowissenschaftliche Forschung Verhaltensexperimente, und die Modellannahmen die postuliert wurden, beziehen sich entweder auf theoretisches oder praktisches Denken. Gegenwärtige Theorien menschlichen Denkens, die auf experimentellen Ergebnissen basieren, sind unter anderem ‚Zwei-Prozess' Modelle. Sie werden ‚Zwei-Prozess' Modelle genannt, da in ihnen ‚rationale' und ‚emotional/intuitive' kognitive Prozesse als Grundlage des Denkens angenommen werden. Solche Erklärungsansätze gibt es sowohl im Bereich der theoretischen als auch der praktischen Denkforschung. Allerdings wird zu beiden Denkbereichen meist getrennt geforscht und dementsprechend gelten die jeweiligen Modelle auch nur für theoretisches oder praktisches Denken. Daher war die Idee für die vorliegende Arbeit, diese beiden Forschungsbereiche gemeinsam mittels Verhaltensexperimenten und bildgebenden Verfahren zu untersuchen, um Evidenzen dafür zu bekommen, dass beide Denkbereiche auf ‚rationalen' und ‚emotional/intuitiven' Prozessen basieren. Sollte es Evidenzen dafür geben, wären wohl ‚Zwei-Prozess' Modelle der beste Erklärungsansatz für beide Denkdomänen. Um dieses Ziel zu erreichen wurden drei Vorstudien und fünf Experimente durchgeführt, die von Verhaltensexperimenten mit „normalen" Personen, über Extremgruppenvergleiche (z.B., hoch-intelligente vs. durchschnittlich-intelligente Personen) bis hin zu bildgebenden Verfahren reichten. Die Ergebnisse dieser Experimente liefern erste Evidenzen für die eben getroffenen Annahmen. Theoretisches und praktisches Denken scheinen auf ‚rationalen' und ‚emotional/intuitiven' kognitiven Prozessen zu beruhen, die wiederum mit Aktivierungen in fronto-temporo-parietalen Gehirnstrukturen assoziiert sind. Die Ergebnisse werden im Rahmen der aktuellen Forschung zu theoretischem und praktischem Denken interpretiert und diskutiert.

Abstract

For hundreds of years, philosophers explored how humans reason and what reasoning abilities humans possess. They were also interested at how these reasoning abilities could and should be applied to reach a peaceful and flourishing coexistence. The two sub-disciplines investigating these issues are theoretical and practical philosophy. Epistemology is the main research domain of the former, whereas morals and ethics represent the main research field of the latter. Philosophers like Kant often derived their theories on practical reasoning from their assumptions on theoretical reasoning, trying to propose an 'all-encompassing' theory. Since the last century, even the field of natural sciences investigated on practical and theoretical human reasoning processes. In contrast to philosophy, disciplines like neuroscience or psychology used behavioral experiments instead of 'thought experiments', and the models proposed often refer either to practical or to theoretical reasoning. Actual proposals derived from experimental findings to explain human reasoning often suggest 'dual-process' accounts. 'Dual-process' models contain 'rational' and 'emotional/intuitive' cognitive processes that are assumed to be involved in human reasoning. Such assumptions have been made in the areas of theoretical and practical reasoning research. However, these reasoning domains have mainly been investigated in isolation and the respective 'dual-process' accounts thus refer to either theoretical or practical reasoning. Therefore, the idea of the current thesis being presented here, is to investigate theoretical and practical reasoning combined, applying behavioral and brain imaging experiments to provide evidence that both reasoning domains are similarly based on 'rational' and 'emotional/intuitive' processes. If such evidence could be found, 'dual-process' models should account for both of these reasoning domains. To reach this aim, three pre-studies and five experiments were conducted ranging from behavioral experiments with "normal" participants over extreme groups (i.e., persons with superior intelligence as compared to persons with average intelligence) to brain imaging techniques. The results reveal first hand evidences supporting the above assumptions. Theoretical and practical reasoning seem to require 'rational' and 'emotional/intuitive' cognitive processes based on activations in a fronto-temporo-parietal network in the brain. These results and further findings will be interpreted and discussed within the context of the current research on theoretical and practical reasoning.

Contents

1. Introduction 1
1.1 Basic principles of logic 3
1.2 Deductive inferences 5
1.2.1 Conditionals 5
1.2.2 Syllogisms 8
1.3 Behavioral experiments on theoretical reasoning 9
1.4 Theories of theoretical reasoning 15
1.5 Behavioral experiments on practical reasoning 18
1.6 Brain imaging experiments on theoretical reasoning 23
1.7 Brain imaging experiments on practical reasoning 26
1.8 Working memory 32
1.9 Recognition 34
1.10 Reasoning, emotion, and the human brain 38
1.11 Excursus – The method of brain imaging 43
1.12 Integration part and derivation of experimental paradigm 46

2. Pre-Study I 52
2.1 Method 52
2.1.1 Participants 52
2.1.2 Material 52
2.1.3 Procedure 54
2.2 Results 54
2.3 Discussion 55

3. Experiment I 56
3.1 Method 56
3.1.1 Participants 56
3.1.2 Material 56
3.1.3 Procedure 57
3.2 Results 57
3.2.1 Error rates 57
3.2.2 Decision times 59
3.3 Discussion 60

4. Pre-Study II 63
4.1 Method 63
4.1.1 Participants 63
4.1.2 Material 63
4.1.3 Procedure 64
4.2 Results 64
4.3 Discussion 65

5. Experiment II 67
5.1 Method 67
5.1.1 Participants 67
5.1.2 Material 67
5.1.3 Procedure 67
5.2 Results 68
5.2.1 Error rates 68

5.2.2 Decision times 69
5.3 Discussion 70

6. Pre-Study III 74
6.1 Method 74
6.1.1 Participants 74
6.1.2 Material 74
6.1.3 Procedure 75
6.2 Results 75
6.2.1 Deductive reasoning problems 75
6.2.2 Recognition items 77
6.3 Discussion 78

7. Experiment III – Training 80
7.1 Excursus - Training in theoretical reasoning 80
7.2 Method 83
7.2.1 Participants 83
7.2.2 Material 83
7.2.2.1 Logic Training 83
7.2.2.2 Pseudo Training 84
7.2.2.3 Moral Judgment Test 85
7.2.2.4 Computer Experiment 86
7.2.3 Procedure 86
7.3 Results 87
7.3.1 Task by group analysis 88
7.3.2 Pre-post-test comparison and MJT 90
7.3.3 Deductive reasoning problems 90
7.3.3.1 Error rates 90
7.3.3.2 Decision times 94
7.3.4 Recognition items 96
7.3.4.1 Error rates 96
7.3.4.2 Decision times 97
7.4 Discussion 98

8. Experiment IV – Intelligence 101
8.1 Method 102
8.1.1 Participants 102
8.1.2 Material 102
8.1.3 Procedure 103
8.2 Results 103
8.2.1 Task by group analysis 104
8.2.2 MJT 106
8.2.3 Deductive reasoning problems 107
8.2.3.1 Error rates 107
8.2.3.2 Decision times 109
8.2.4 Recognition items 111
8.2.4.1 Error rates 111
8.2.4.2 Decision times 112
8.2.5 Correlation 113
8.3 Discussion 114

9. Experiment V – fMRI 118
9.1 Method 119
9.1.1 Participants 119
9.1.2 Material 120
9.1.3 Procedure 120
9.1.4 Analysis of behavioral data 120
9.1.5 Image acquisition and analysis 121
9.2 Results 122
9.2.1 Behavioral data 122
9.2.1.1 Error rates 122
9.2.1.2 Decision times 123
9.2.2 Imaging data 124
9.2.2.1 Results of the Region of Interest (ROI) analyses 124
9.2.2.2 Results of the explorative whole-brain analysis 126
9.3 Discussion 127
9.3.1 Behavioral Results 127
9.3.2 Imaging data 128
9.3.2.1 Abstract problems 128
9.3.2.2 Moral problems 128
9.3.2.3 Neutral problems 129
9.3.2.4 Conjunction analysis 129
9.2.3.5 Difference analysis 130

10. General Discussion 132
10.1 Pre-Studies 133
10.2 Experiments I + II 136
10.3 Experiment III – Training 139
10.4 Experiment IV – Intelligence 143
10.5 Experiment V – fMRI 145
10.6 Critics and limitations 147

11. Conclusions and outlook 152

Bibliography 154

Appendix 166

Statutatory declaration (“Eidesstattliche Erklärung”) 208

1. Introduction

For hundreds of years, philosophers have tried to unlock the basis of human 'eudaimonia' (from the Greek, often translated as happiness, felicitousness, bliss, and the like), the basis of a peaceful, flourishing life (e.g., Stoa, see Weinkauf, 2001; Plato, see Wolf, 1996; Aristotle, 2002). To accomplish such a "research" goal, two main sub-disciplines developed dealing with the important issues involved: theoretical and practical philosophy.

Theoretical philosophy deals with the understanding of how humans' reason, or at least, what reasoning abilities and capacities humans possess. The main sub-discipline of theoretical philosophy is 'epistemology' (i.e., the study of knowledge and justified belief). Thereby, a main question is how true knowledge and justified beliefs can be achieved, or how knowledge can be evaluated as true and beliefs as justified (Sandkühler, 1999).

Practical philosophy on the other hand, is concerned with the ways on how humans should reason to act accordingly or to evaluate the actions of others as appropriate or inappropriate. Practical philosophy therefore asks what kind of actions and behaviors are adequate and useful to be able to attain 'eudaimonia'. The main sub-discipline of practical philosophy is ethics dealing with morality and morals (Sandkühler, 1999).

Many philosophers contributed different theories in both theoretical and practical philosophy. The ideas concerning 'how we should live and act' (i.e., ethics/morality) were of particular interest, not only for philosophers but by many. The basis of the reasoning processes and mechanisms involved in practical reasoning however, were often derived from the theoretical assumptions developed before. Thus, theoretical and practical philosophy are often strongly intertwined.

Nowadays, theoretical and practical reasoning is not only investigated in the domain of philosophy, but also in the areas of experimental research like psychology or neuroscience. However, many of the tasks applied in experiments are derived from philosophy and often date back to Aristotle's time (cp., Manktelow, 2004). A main difference between the former philosophical theories and actual empirically based proposals is that theoretical and practical reasoning are now investigated separately and that the former reciprocal correlation between theoretical and practical reasoning seem to have been shelved. Thus, one of the purposes of this current project is to recombine theoretical and practical reasoning research within a single experimental design.

The methods used in theoretical philosophy, and consecutively also applied in practical philosophy, were often based on logic. Logic in turn was thought to correspond to rationality or as Kant (1781/1998) put it, 'pure reason'. However, for the purposes of clarity, Kant (1781/1998) understood logic as the science of the necessary laws of reason thus he denoted that logic tells us how to reason, not based on empirical (i.e., subjective or psychological) principles, but rather on objective (i.e., ‘a priori’) principles. Therefore, Kant’s term of logic and the use of logic referred to the mere form of the reasoning process, independent of any kind of content. The same holds for the sub-discipline of ‘formal logic’ (in the following referred to as ‘logic’). Logic delivers a formal system on how to reason correctly, often using axiomatic propositions. Today, logic is understood as the system proposing the principles of drawing valid inferences based on mere formal propositions, statements or judgments (Tugendhat & Wolf, 2004). Thereby, making a deductive inference means inferring a conclusion from a diversity of given premises (i.e., propositions/statements). The opposite method is inductive inference (Sandkühler, 1999). Deductions mean inferences drawn from the general to the particular, while inductive inferences mean general conclusions derived from statements about particular cases. However, an inductive inference is insecure and could be wrong. Deductive inferences (also often labeled deductive reasoning or logical reasoning) on the other hand, lead to valid statements/conclusions. A special relationship between the given premises and the conclusion therefore, exists: If the premises are true, then the conclusion is true (Hoyningen-Huene, 2006). Note that there is no assertion about the premises, which might be wrong. However, if the premises are true, then, the conclusion is also true, given that the deduction is valid. This refers to the fact that the validity of an inference is independent of the truth status of the premises or conclusion. Drawing a valid inference depends not on the content, but only on the logical form of the argument. An inference may therefore be valid, even though the conclusion is not true (i.e., false).

As mentioned above, philosophers also applied the method of logic in practical philosophy. Early approaches in ethics were, for example, the virtue theories by Plato (see Wolf, 1996) or Aristotle (2002), the founder of (formal) logic. They proposed living in accordance with our nature as a desirable goal, including development of virtues and prevention of vices. The idea of living only according to our “god”-given nature may seem antiquated; however, it provided ‘guidelines’ for human behavior. Other theories focused on emotions as the basis for practical reasoning and decision making processes (emotive ethics; e.g., Shaftesbury, 1711/1963; Hutcheson, 1755/1968; Hume, 1748/2002, 1751/2003; Smith, 1759/ 2004). These concepts were much closer to current research findings (e.g., Haidt, 2001). The same goes with exponents (utilitarianism/consequentialism; e.g., Bentham, 1789/2007; Mill 1863/1994) stressing the outcomes of the actions themselves instead of the agents involved (e.g., Greene, Morelli, Lowenberg, Nystrom, & Cohen, 2008). Finally, one of the best-known points has been made by Kant, who wanted to relieve us from our self-inflicted immaturity. He seemed to be the first to systematically explore the ‘pure’ (rational) reasoning abilities which humans possess, and based thereupon, how we should reason and consequently act (Kant, 1781/1998, 1785/2000, 1797/2007, 1788/2008).[1] His ‘categorical imperative’, as basis for a life according to duties, is the most famous statement aimed to come up with a guide to human behavior. It is important to note that Kant had derived his assumptions from ‘pure rationality’ which humans should apply in reasoning. This is called deontology or deontological ethics.

As was mentioned earlier in the history-related part of the introduction, main research interests in philosophy as well as in the natural sciences had tried to explore human reasoning, both in the theoretical and the practical domain. Thereby, a fundamental issue concerns the reasoning processes involved in either of these reasoning domains. The current thesis was inspired by the idea laid out in the original philosophical research to explore these reasoning domains in combination. However, it is not to deduce practical from theoretical reasoning, but to investigate the cognitive processes presumably involved in both, thus common to theoretical and practical reasoning, through the application of experimental methods. Therefore, the following introduction will report research findings on theoretical and practical reasoning, utilizing behavioral and brain imaging experiments. The findings then shall be integrated to derive the leading research idea and to present the developed experimental paradigm used for the experiments conducted within this thesis.

Since logic plays such an important role in philosophy and in experimental researches, it is only helpful to introduce some terms used, clarifications, and principles of ‘classical’ (formal) logic.

1.1 Basic principles of logic

The starting points in logic are simple statements/propositions that could be true or false. Since logic deals with symbols, abstracting from the content, one may use ‘A’ and ‘B’ for statements[2], and the abbreviations ‘T’ for true and ‘F’ for false. ‘T’ and ‘F’ are truth values in logic, indicating that a statement could be true or false, but nothing else (“tertium non datur”,

[1] Kant started with the ‘critique of pure reason’ investigating our ’pure’ logical reasoning abilities before introducing the ‘critique of practical reason’ and the ‘metaphysics of morals’ as basis of morality.

[2] Normally, lower case letters are used for propositional logic. However, to stay consistent with the following experiments, where upper case letters are used, these are accordingly applied here.

from Latin: there is no third, cp., Hoyningen-Huene, 2006).[3] Such statements do not need to stand alone (e.g., 'A'), but could be combined (e.g., 'A' and 'B'). Combining different statements could lead to extensional sentential connectives. Extensional sentential connectives are also called 'truth functional' connectives and combine two partial statements to an overall statement, whereby the truth value of the resulting overall statement is unambiguously defined by the truth values of the partial statements (cp., Hoyningen-Huene, 2006). Extensional in the domain of logic signifies to abstract from the meaning (cp., Hoyningen-Huene, 2006).

A part of logic, namely 'propositional logic' (also called 'junctor logic', or 'sentential logic', or 'statement logic'), deals with these extensional sentential connectives. Symbols (i.e., junctors) applied to connect statements/propositions are for example '∧' as truth functional 'and' (i.e., conjunction), '∨' as truth functional 'or' (i.e., disjunction), and '¬' as truth functional 'not' (i.e., negation). A simple method to infer the truth value of the overall statement from the partial statements with respect to their 'truth functional junctor are truth tables, which allow the determination of the truth value of a sentential connective statement. These truth tables show all potential truth values of the partial statements followed by the resulting truth values for the sentential connective as a whole. In Table 1 below, the potential truth values for 'A' and 'B', and the resulting truth values according to conjunction, disjunction, and negation are depicted.

Table 1: Truth table showing potential truth values for 'A' and 'B' as well as the resulting truth values of the conjunction ('∧') and the disjunction ('∨') of both, with 'T' as true and 'F' as false. The truth table for negation ('¬') is shown separately below, since this table only contains two columns.

A	**B**	**A ∧ B**	**A ∨ B**
T	T	T	T
T	F	F	T
F	T	F	T
F	F	F	F

A	**¬B**
T	F
F	T

Despite these "simple" combinations of statements, even more elaborate sentential connectives are used. These deal with two more junctors that both refer to statements connected via an "If…, then…" sentence. The first part of this sentence is called 'antecedent', the second one 'consequent', and the whole statement is referred to as 'conditional' (cp., Hoyningen-Huene, 2006). The first junctor connecting 'A' and 'B' is '→', called 'material

[3] Note that this applies to binary logic, which is important for the current work, but also other logic systems exist.

implication, the second junctor connecting ‘A’ and ‘B’ is ‘↔’, named ‘material equivalence’ or ‘biconditional’. The truth tables for both of these possible “If…, then…” connections are presented in Table 2 below. Note that both of these forms can be rewritten applying negation and conjunction, thus simplifying the evaluation of their respective truth values.

Table 2: Truth table with the antecedent ‘A’, the consequent ‘B’, and the potential truth values ‘T’ for true and ‘F’ for false; on the left side is the conditional (A → B) with the material implication (→); right side shows the biconditional with the material equivalence (↔); (‘¬’) as negation and (‘∧’) as conjunction.

A	**B**	**A → B or ¬ (A ∧ ¬ B)**	**A**	**B**	**A ↔ B or (A → B) ∧ (B →A)**
T	T	T	T	T	T
T	F	F	T	F	F
F	T	T	F	T	F
F	F	T	F	F	T

To avoid misunderstandings, logicians often use the terms ‘iff’, or the connective ‘if and only if’ to denote a ‘material equivalence’ instead of a ‘material implication’ (cp., Manktelow, 2004). As could be seen, the truths table of the ‘material implication’ shows that it is not the case that ‘A’ and simultaneously ‘not B’. Thus, the only false truth value results if ‘A’ is true but concurrently ‘B’ is false. In contrast, the ‘material equivalence’ denotes that ‘A’ and ‘B’ are equivalent interchangeably. ‘A’ implies ‘B’ and vice versa. Thus, either both statements are true, or both statements are false, but it cannot be the case that ‘A’ is false and ‘B’ is true or ‘B’ is false and ‘A’ is true.

Now, the basics of logic are set and the presentation will show how deductive inferences can be drawn from these ‘simple’ statements.

1.2 Deductive inferences

1.2.1 Conditionals

The last section reported the basics of logic, especially of single statements/propositions and their connections. It has been shown how truth values of single statements and sentential connectives could be determined. These basics of logic can now be applied to deductive reasoning. In philosophy, deductive inferences are also drawn by using the above mentioned principles as well as further deduction rules to draw valid conclusions from given premises. A form of deductive reasoning encompasses the conditional sentences reported above and is also known as conditional reasoning (cp., Manktelow, 2004). This form of inference consists of the conditional “If A, then B” sentence, then also labeled ‘major premise’, a second sentence, named ‘minor premise’, and a third sentence, the ‘conclusion’ (cp., Hoyningen-Huene, 2006).

The minor premise is an affirmation of the ‘A’ or the ‘B’ part of the conditional sentence, or a negation of the ‘A’ or ‘B’ part of the first sentence. The third sentence, the conclusion, is then derived from the first two sentences. Formally, such a deductive inference is written as follows.

If A, then B.	(first premise)
A.	(second premise)
B.	(conclusion

Above, the truth values of the two possible “If…, then…” sentences were described. If a second premise is now introduced, which could represent four possible assertions of the ’if’ or ‘then’ parts of the conditional (i.e., ‘A’, ‘¬A’, ‘B’, ‘¬B’), it seems as if eight possible conclusions could be drawn. The question is however, whether these inferences are valid or invalid. See Table 3 for an illustration of these possible assertions of the minor premise combined with the two conditional sentence connections.

Table 3: Conditional sentences with ‘material implication’ or with ‘material equivalence’ as major premise and the four possible assertions of the minor premise.

	material implication/conditional				material equivalence/biconditional			
major premise	$A \rightarrow B$				$A \leftrightarrow B$			
minor premise	A	¬A	B	¬B	A	¬A	B	¬B

As it turns out under the ‘material equivalence’ condition, four inferences can be drawn rendering the resulting conclusion always valid. In contrast, under the ‘material implication’ condition, only two inferences lead to valid conclusions, whereas the other two inferences lead to invalid conclusions. This is due to the fact that the statements ‘A’ and ‘B’ are interchangeably equivalent in the former case, but not in the latter. Interchangeably means that ‘If A, then B.’ can be rewritten as ‘If B, then A.’, since these terms are ‘equivalent’. In contrast, under ‘material implication’ ‘A’ implies ‘B’, but not vice versa. Valid inferences are labeled with Latin names, i.e. ‘Modus Ponens’ and ‘Modus Tollens’ (for illustration see Table 4).

Table 4: Conditional inferences and validity with major premise, minor premise, and conclusion for ‘material implication’ and ‘material equivalence’ (adapted from Manktelow, 2004).

		inference type	material implication	material equivalence
Major premise	If A, then B.			
Minor premise	A.	Modus Ponens (MP) (Affirmation of the antecedent)	B. / valid	B. / valid
	Not A.	Denial of the antecedent (DA)	Invalid	Not B. / valid
	B.	Affirmation of the consequent (AC)	Invalid	A. / valid
	Not B.	Modus Tollens (MT) (Denial of the consequent)	Not A. / valid	Not B.

Another way to illustrate the relationship between the ‘If’ part of a conditional sentence and the ‘Then’ part differentiated according to ‘material implication’ and ‘material equivalence’ are ‘Euler Circles’. Here, the position of the circles represents the relationship and the labeling of the circles indicates the ‘If’ and the ‘Then’ parts of the conditional sentence. Note that the size or shape of an Euler circle is not relevant. The significance of the circles depends on their potential overlap (Figure 1).

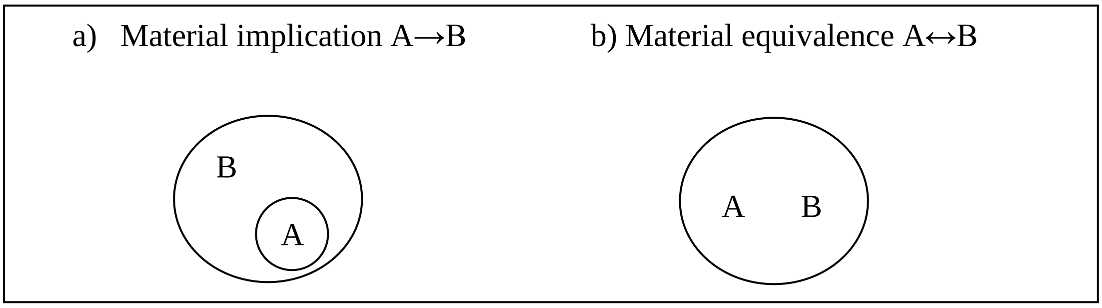


Figure 1: Euler circle representations for ‘material implication’ and ‘material equivalence’ of conditional sentences (adapted from Manktelow, 2004).

This graphical illustration representing the two different readings of conditional sentences seems to provide an easier way to infer correctly as to which inferences are valid or invalid, given a ‘material implication’ or a ‘material equivalence’. Problems might occur when someone does not know truth tables and/or Euler circles, and when it is not clear whether the presented “If…, then…” clause represents a ‘material implication’ or a ‘material equivalence’: In fact, concerning studies with conditionals state that “[…] there is still no complete universally accepted account of its [‘If’] use.” (Manktelow, 2004, p. 37). The following section on empirical studies of conditional reasoning will discuss this issue more elaborately. However, since many studies on human reasoning also applied syllogisms, a short overview on syllogisms shall be given beforehand.

1.2.2 Syllogisms

Besides conditional sentences, other forms of deductive inferences are called syllogisms. Originally based on the ‘syllogistic’ of Aristotle (Sandkühler, 1999), containing quantifiers like “all”, “some”, or “none” they differ from conditional deductions wherein a new proposition (i.e., conclusion) is drawn from two propositions (i.e., premises; cp. Sandkühler, 1999). In the current version of ‘categorical syllogistic’, the two premises encompass a middle term ‘B’, not occurring in the conclusion, and two end terms ‘A’ and ‘C’. An example for a valid syllogistic inference is shown below.

All A are B.	(first premise)
All B are C.	(second premise)
All A are C.	(conclusion)

Now, it can be seen more clearly that the conclusion results from the relation between the two end terms ‘A’ and ‘C’, whereas the middle term ‘B’ is eliminated. The different quantifiers referred above result in four basic expressions in the Aristotelian logic, which are shown in Table 5.

Table 5: The quantified statements applied in syllogisms of Aristotelian logic (adapted from Manktelow, 2004).

A	All A are B.	Universal	Affirmative
I	Some A are B.	Particular	Affirmative
E	No A are B.	Universal	Negative
O	Some A are not B.	Particular	Negative

These single instances of the four basic quantifiers are the starting basis of the syllogistic arguments, which consist of two premises and a conclusion, and are known as ‘moods’. As with the conditional sentences referred to above, there are different potential combinations if only the premises are taken into consideration. The relation or ‘figure’ as Johnson-Laird (1983) labeled it could vary according to the arrangement of the premises (Table 6).

Table 6: Syllogistic figures according to their relation of A, B, C in the first and second premise (adapted from Johnson-Laird, 1983).

First premise	A – B	B – A	A – B	B – A
Second premise	B – C	C – B	C – B	B – C

According to the explanations by Manktelow (2004), a total of 512 valid syllogisms can be constructed, whereas Aristotle only allowed 14 valid syllogisms. The reason for this huge difference is the admittance of 'existential presuppositions' and whether "weak" conclusions are accepted or not (cp. Manktelow, 2004). Since these issues are far beyond the scope of the current work, only an illustration for one statement of the first premise will be shown using Euler circles (Figure 2).

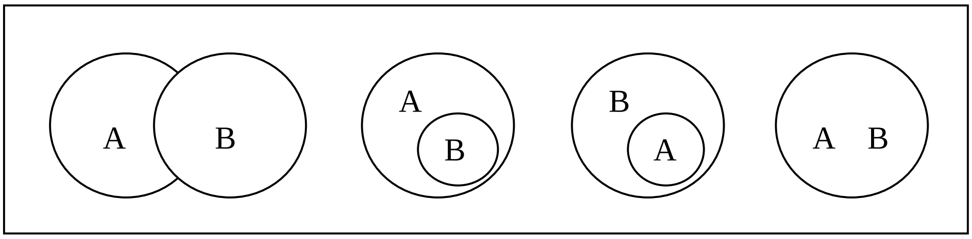


Figure 2: Euler circles for the first premise of a syllogistic statement 'I' "Some A are B" (adapted from Manktelow, 2004).

The examples given above with the Euler circles show potentially different readings of a "Some A are B." statement, indicating why various numbers of valid conclusions are possible with the four syllogistic figures.

Philosophers further analyzed, discussed, and developed the issues concerning 'logic' and 'logical reasoning', but an interesting question still remains: how and whether 'normal' people who are not trained in logic, are able to reason 'logically' or to deal with deductive reasoning problems. Since the late 1950s, the domain of 'cognitive psychology' began the investigation on human reasoning abilities and performances in numerous experiments applying various reasoning problems (cp., Manktelow, 2004). A few of the major findings will be reported in the succeeding part.

1.3 Behavioral experiments on theoretical reasoning

Experimental research is often concerned with the investigation of three different forms of deductive reasoning problems (e.g., Knauff, 2009a): conditional reasoning in the domain of propositional logic (using conditional sentence structures like 'If…, then...'), syllogistic reasoning in the domain of predicate logic (using quantifiers like 'all', 'some', or 'none'), and relational reasoning (using n-place relations like 'bigger/smaller' or 'to the right/left of'). Here, the focus will almost be exclusively on conditional reasoning problems since this kind of reasoning tasks were used in all of the experiments which shall be reported later on. However, a few experimental results reported in the following presentation also refer to syllogistic reasoning problems which were introduced earlier.

The different versions of deductive inferences utilizing conditional sentence structures have already been reported in their abstract version above. These deductive inference problems were used in many experiments in the investigation of human reasoning processes (Manktelow, 2004). The task of the participants is to judge the validity of a given conclusion, or to derive the conclusion by themselves. It seems important to denote again that the validity of these four inference schemas does not depend on the potential truth status of the premises, but depends on their formal logical relation embedded within the whole problem. This is important to keep in mind with studies about reasoning, taking into account that everyday reasoning is not content-free or is not dealing with abstract problems. Thus, research on human reasoning abilities addresses both the ability to deal with abstract (i.e., content-free) problems as well as with meaningful ones (cp., Knauff, 2009a). Compared to processes involved in reasoning about abstract "materials", it seems more relevant to understand the reasoning processes and capacities that humans access when handling content-laden "material", especially when it comes to contents that are relevant in everyday tasks. To illustrate the difference between abstract and meaningful material in detail, formal logical content-free (i.e., abstract) versions of the four deductive inference problems as well as content-filled examples are shown in Table 7 below according to the 'material implication'.

Table 7: The four inference schemas of conditional reasoning (adapted from Knauff, 2007).

Inference schema	Logical validity	Example with content
MP, Modus Ponens (affirmation of antecedent) If A, then B. A. B.	valid	 If it rains, the street is wet. It rains. The street is wet.
AC, affirmation of consequent If A, then B. B. A.	invalid	 If it rains, the street is wet. The street is wet. It rains.
DA, denial of antecedent If A, then B. ¬ A. ¬ B.	invalid	 If it rains, the street is wet. It does not rain. The street is not wet.
MT, Modus Ponens (denial of consequent) If A, then B. ¬ B. ¬ A.	valid	 If it rains, the street is wet. The street is not wet. It does not rain.

MP and MT are logically valid inferences; AC and DA are logical invalid inferences; ¬ = represents the logical symbol for negation and is equivalent to "not".

The validity/invalidity of the conclusions of these four inferences (see e.g., 'Modus Ponens' or 'Denial of the Antecedent' above) could be shown with the symbols and deduction rules of logic. Here, the deductive inference will be explained with the content-filled reasoning problems illustrating more “natural” reasoning. The first problem seems to be easy: ‘If it rains, the street is wet.’; ‘It rains.’ The antecedent proposes that raining leads to the consequence of a wet street. Since ‘rain’ is given as second premise, the consequence matches the precondition to be fulfilled, thus the conclusion of a wet street is valid. An explanation for the invalid conclusion of the second inference schema could be circumstances other than raining caused the street to be wet. Thus, the conclusion is false. For the third problem, one might assume that the street could be wet, although it did not rain. The main point is, the antecedent only says that ‘IF IT RAINS’, something special follows. The antecedent says nothing about a potential consequence if it does not rain at all. Vice versa, the antecedent proposes no consequence at all ‘if it DOES NOT rain’. In contrast, concerning the latter case (the MT) an explanation might be given as follows. The antecedent proposes the consequent, thus, what must happen if it rains. This means, the consequence necessarily follows the antecedent. However, the second premise states that the consequence is not fulfilled. Since there is a strong relation between antecedent and consequent, one could assume that no wet street means not raining. This means that it could not be that the consequence does not appear, but that the antecedent is given. This would violate the sense of an ‘If…, then…’ connection, also in a colloquial sense, or even better, in a general understanding of such a clause. To sum up, the main point about drawing a deductive inference in achieving a logically valid conclusion is that a conclusion is always valid if no counter examples exist.

The table above illustrated content-filled deductive inference problems used in reasoning experiments. One can imagine that persons not educated to be logicians might draw conclusions from the premises that are based on semantics rather than logic. That seems to imply that these persons might fail to draw logically correct conclusions.[4] Another point might be that persons do not interpret a conditional sentence as ‘material implication’, but as ‘material equivalence’, thus evaluating all presented conclusions as valid. Therefore, the question is how participants in experiments on deductive reasoning deal with reasoning problems and whether there are differences in the performances between the different inference schemas (i.e., MP, MT, DA, and AC). Indeed, many experiments revealed that persons have difficulties in handling deductive inference problems (cp., Evans, Newstead, & Byrne, 1993). Evans et al. (1993) summarize the results of diverse studies in the investigation of abstract reasoning problems indicating that participants show differences in performance

[4] Note that a conclusion follows valid (i.e., necessary) or not from the given premises, but participants dealing in experiments with deductive reasoning problems can give correct or incorrect answers.

and difficulties faced in these tasks. Participants had nearly no problems with the MP (89 – 100 % correct), more problems with the MT (41 – 81 % correct), and much more difficulties with the invalid versions of AC (23 – 75 % correct) and DA (17 – 73 % correct) (Evans et al., 1993). Manktelow (2004) refers to this review as pointing to the wide variation in the acceptance of the different inferences, but also presents weighted average frequencies with 96.6 % correct answers for the MP, 60.2 % for the MT, 39.6 % for the DA, and 39.5 % for the AC (cp., Manktelow, 2004). Also, Manktelow denotes the fact that we ‘normally’ do not reason about such content-free problems. Thus, the interesting question is, whether or not content-based reasoning problems are solved better. Furthermore, it is of great interest as to which factors cause errors during the reasoning process, which cognitive mechanisms guide reasoning, and additionally, not only whether or not content and background knowledge influence reasoning but also what kind of content plays a critical role in these processes (Knauff, 2007). Therefore, many experiments applying different tasks including deductive reasoning problems have been conducted in order to explore these factors (cp., Manktelow, 2004).

One of the best-known experimental paradigms in investigating deductive reasoning is the Wason Selection Task (WST, Wason, 1966). It was one of the first tests used to explore not only the abstract version (Figure 3), but also content effects in deductive reasoning. In the original abstract version, participants are shown a conditional sentence ‘‘If there is a vowel on one side of the card, then there is an even number on the other side’’ and four cards subsequently (Figure 3). The cards have a visible number or letter on the front side which corresponds to the four possibilities of the second premise (i.e., ‘A’, ‘not A’, ‘B’, ‘not B’) explained above. The back side of the cards also contains a number or a letter which is not visible to the participant. Participants are instructed that they are allowed to turn over the cards to test whether the rule (i.e., the conditional sentence) is true or false. This kind of reasoning task corresponds to the four inference schemas reported above and a correct solution would consist of drawing the valid inferences MP and MT. According to the cards presented, the cards with the “E” and the “3” had to be turned over.

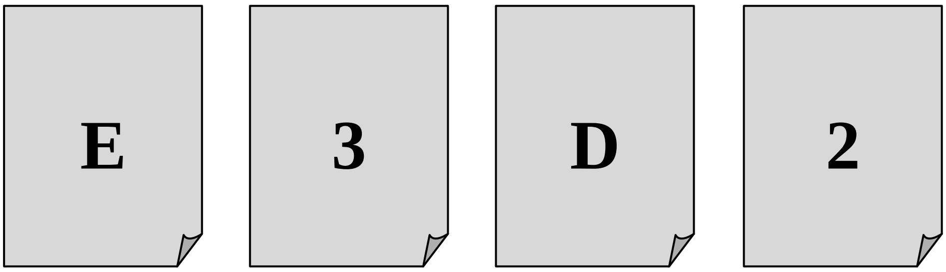


Figure 3: The Wason Selection Task (WST). The participants have to verify the rule ‘‘If there is a vowel on one side of the card, then there is an even number on the other side’’. The visible letters and numbers on the card correspond to the four possible propositions A, ¬A, B, and ¬B. Participants were allowed to turn the cards and falsify the rule, thus evaluating the validity of the given conditional rule.

The WST has been repeated many times in this abstract version, but only a few participants were able to solve it correctly (ca. 10 %) and choose the right cards (cp., Evans et al., 1993; Wason & Johnson-Laird, 1972). Most often, participants tend to verify the rule and choose the cards “E” and “2”, or just select the card “E” alone. However, if the WST is filled with content, performance improves dramatically. Wason and Shapiro (1971) as well as Johnson-Laird, Legrenzi, and Legrenzi (1972) could show for example that the content embedded in the task and in contrast to the pure abstract version facilitated solving the WST correctly. The content Wason and Shapiro (1971) applied consisted of the statement “Every time I go to Manchester I travel by car” and subsequently presented cards, which had either the name of a city (e.g., Manchester, Leeds) or the travel mode “car” and “train” on their front side. Participants were allowed to inspect the cards before the experiment, to ensure that all had names of cities or travel modes also on the back side. In this version of the WST, 10 out of 16 participants solved it correctly, whereas only 2 out of 16 of the same participants solved the abstract version correctly. Thus, comparing the 62.5 % participants who answered correctly with the usual rate of only 10 % or less in solving the WST correctly shows an indeed enormous performance increment. Johnson-Laird et al. (1972) conducted a further study revealing even more evidence of a facilitating content effect. They presented the conditional sentence “If a letter is sealed, then it has a 50 lire stamp on it” and subsequently closed or unclosed envelopes or envelopes from the front side with 50 or 40 lire stamps on it. This time, participants were explicitly advised to imagine that they are post officers. They had to check which of the envelopes they have to turn over to find out whether or not the envelope violates the rule. Results showed the above mentioned facilitating content effect since 21 out of 24 participants solved the content-filled WST correctly, whereas only 2 out of 24 managed the abstract version, also applied in this study. The reported facilitating content effects of realistic material involved in deductive reasoning tasks lead to the assumption that persons have only difficulties when dealing with the abstract version derived from ‘classical’ logic.

However, many experiments were conducted using diverse contents that showed no or even perturbing content effects. For example, Manktelow and Evans (1979) or Evans, Barston, and Pollard (1983) as well as Evans, Newstead, Allen, and Pollard (1994) could not find any facilitating content effects in their experiments.

Cox and Griggs (1982) proposed an idea regarding as to why and when content facilitates the reasoning process. They varied the content systematically and found that only familiar content enhances the reasoning performance. Cox and Griggs (1982; Griggs & Cox, 1982) argue that past experiences connected with the content enhances the reasoning performance, whereas others (Canessa, Gorini, Cappa, Piatelli-Palmarini, Danna, Fazio, & Perani, 2005; Cheng & Holyoak, 1985; Cosmides, 1989; Gigerenzer & Hug, 1992; Manktelow & Over, 1991) claim it is the kind of content, i.e., social content, which facilitates reasoning.

Socially relevant or familiar content refers to prior knowledge gained through experience. Therefore, other researchers (e.g., Evans et al., 1983) conducted experiments testing reasoning with prior knowledge and beliefs. The 'belief bias' effect they found is a well-known phenomenon today (c.p., Evans, 2008; Knauff, 2007) which states the existence of "[…] a tendency to judge the argument of a syllogism as valid or invalid on the basis of whether or not it is *a priori* believable." (Evans et al., 1994, p 266).[5] So, Evans et al. (1983) could show that participants accepted far more valid than invalid and believable conclusions than unbelievable ones. In addition, the 'belief bias' effect was stronger for invalid problems. Later, Evans, Over, and Manktelow (1993) presented findings that participants accept conclusions as being rather logically valid if they are consistent with prior beliefs than conclusions that contradict existing beliefs.

Besides socially relevant content and prior beliefs/knowledge, the emotional load of the material also seems to have an impact on deductive reasoning. Blanchette and Richards (2004), and Blanchette (2006) investigated reasoning with inference problems containing emotional content and could show that positive as well as negative emotional content also disturbs the reasoning process. However, Gangemi, Mancini, and Johnson-Laird (2006) and Blanchette and Campbell (2005) conducted experiments yielding opposite results: emotions could support the reasoning process. In a further study, Blanchette, Richards, Melnyk, and Lavda (2007) combined emotional material with socially relevant content (due to prior experience), testing neutral, emotional, and terrorism-related syllogisms. The syllogisms were

[5] Despite this quotation, which refers to syllogistic reasoning problems, the belief bias effect has also been found in studies using conditional inference problems, although most often syllogisms were applied (cp., Evans & Feeney, 2004).

either believable or unbelievable, and valid or invalid. The problems were congruent (believable/valid and unbelievable/invalid) or incongruent (believable/invalid and unbelievable/valid). It is worth taking note that the participants experienced the terrorist attacks in London in 2005. As control groups, people from Manchester (England) and Canada were tested. Blanchette et al. (2007) could investigate far more realistic emotional and experience-based influences than normally possible in the laboratory. In total, their results showed that all participants performed better with the neutral problems as compared to the emotional or terrorism-related ones, but the London and Manchester participants were additionally slightly better in the incongruent-terrorism versions. A follow up experiment, six months after the attacks, yielded that only the directly affected London group fared still better in the incongruent-terrorism version compared to the other two groups. Blanchette et al. (2007) concluded that the London group was the only one that was indeed directly affected by the attacks and this could account for their better reasoning performance on the incongruent terrorism-related problems.

Although, there is still an ongoing debate whether reasoning is based on formal and logical processes or based more on heuristic, intuitive or knowledge-related ones (Chater & Oaksford, 2001; Evans, 2003; Stanovich & West, 2000), Blanchette et al. (2007) suppose that both types of processes are required.

To sum up, theoretical reasoning processes seem to be influenced by socially relevant and emotional content as well as by prior beliefs and knowledge or additional task-specific criteria (i.e., time pressure; e.g., Evans & Curtis-Holmes, 2005; Evans, Handley, & Bacon, 2009). It is of interest whether or not there are theories or models available, which might explain the different findings and provide further suggestions on how to continue in order to explore the capacities of human reasoning. Therefore, a short overview on different reasoning theories will be given in the next chapter.

1.4 Theories of theoretical reasoning

There are two main theoretical strands explaining deductive reasoning and why errors might occur: the semantic and the syntactic theories (cp., Knauff, 2009a). The first strand encompasses syntactic theories like the theory of mental logic or mental rules theory (Braine & O’Brien, 1998; Braine, O'Brien, Noveck, Samuels, Lea, Fisch, & Yang, 1995; Rips, 1994). They propose that humans have a universal and internalized repertory of abstract inference rules that is used to derive or prove conclusions. Those theories are closely related to (formal) logic and divide the reasoning process into three steps. First, the logical form of the premises, that is the syntactic form to which inference rules apply, must be detected. That means that the following steps of the reasoning process only deal with an abstract form of the premises and ignore the content. Second, the available inference rule set is used to either generate or verify/falsify a conclusion. This process includes the execution of reasoning strategies for combining inference rules to chains of derivation, defining intermediate goals and incorporating newly derived premises into the chain of derivation. Third, the 'abstract conclusion' is translated back into the content of the premises. Errors can occur in all three steps, whereby the possibility of the occurrence of errors increases according to the problems' complexity, and when indirect rules need to be applied. The latter additionally requires conscious efforts, in contrast to when direct rules of inference are applicable. Application of direct rules is assumed to be made with less effort and with more accuracy (Braine & O'Brien, 1998). Extra exertion of efforts due to indirect rules or the complexity of problems is believed to require increased working memory (WM) capacity. Since WM capacity is limited (cp., Miller, 1956), WM overload can additionally increase errors while processing information.

The second branch on deductive reasoning comprises the mental model theory (Johnson-Laird, 1983, 2006; Johnson-Laird & Byrne, 1991; Johnson-Laird, Byrne, & Schaeken, 1992). It is seen as a semantic theory as reasoning is based on a semantic interpretation of the premises. This theory also provides three steps in reasoning and again, in all of these steps errors may occur. The first step includes building a mental model based on the meaning defined by the premises given and based on relevant general knowledge related to the premises. The second step contains the derivation of a conclusion from the model whereby the conclusion includes information exceeding that of the premises. Therefore, the semantic information of the premises has to be maintained. Finally, the conclusion of the model is expected to be regarded as valid if no counter-examples can be found. If at least one counter-example is found, one must return to the second step (Johnson-Laird et al., 1992). Task complexity according to this theory increases if several models must be constructed, such a case occurring for MT inferences for example, but not with MP deductions. Again, a relation to WM capacity is postulated, which might be exceeded if too many models must be constructed due to complex problems (Johnson-Laird, 2006). Furthermore, a relation between reasoning performance and intelligence is proposed such that participants with higher intelligence deal better with deductive reasoning problems (Johnson-Laird, 2006). This display of better reasoning performance in turn, is related to superior working memory capacities (Johnson-Laird, 2006).

Other approaches in deductive reasoning, namely domain-specific accounts, focus more on the content of the reasoning process and thus postulate qualitative differences. The theory of pragmatic reasoning schemas (Cheng & Holyoak, 1985) for example proposes that deduction is based on abstract knowledge structures (pragmatic reasoning schemas) induced from typical life experiences. A similar approach is the logic of social exchange (Cosmides, 1989) which also proposes schemas but differs with regards to the acquisition of the schemas. Cosmides (1989) represents an evolutionary point of view and claims that humans developed domain-specific algorithms with inference rules for solving adaptive problems through natural selection. Social exchange, or the "[…] cooperation between two or more individuals for mutual benefit […]" (Cosmides, 1989, p. 187) is an adaptive problem and an essential part for human survival. Therefore, humans have developed schemas which are utilized particularly for reasoning in dealing with social exchange. Again, domain-specific accounts of reasoning are also prone to errors.

Finally, 'dual-process' accounts/models try to combine different streams of deductive reasoning research. Evans (2003, 2008) claims that all of these theories postulate two different systems (also named processes or types) and may agree that 'System 1' processes are unconscious, rapid, automatic, and of high capacity, whereas 'System 2' processes are conscious, slow, and deliberate. In his view, the often ignored emotions should be included in System 1. Support for 'dual-process' accounts comes from diverse approaches showing that both systems can work together or compete with each other and therefore facilitate or even impair reasoning (Evans, 2003). Further evidences have been presented by Reber (1993) or Stanovich (1999). Evans (2008) subsumes that they proposed "[…] a link between System 2 processing and general intelligence, with the corollary System 1 processes are independent of general intelligence." (p. 262), indicating individual differences in reasoning abilities. Colom, Rebollo, Palacios, Juan-Espinosa, and Kyllonen (2004) showed that general intelligence and working memory capacities are highly correlated. Evans (2008) follows from these findings, "[…] one of the stronger bases for dual-systems theory is the evidence that "controlled" cognitive processing correlates with individual differences in general intelligence and working memory capacity, whereas "automatic" processing does not." (p. 262). Additionally, Evans (2008) believes that the mental logic theory (e.g., Braine & O'Brien, 1998) as well as the mental model theory (Johnson-Laird, 1983) implicitly contain dual-process components. In his view both theories include assumptions that could account for pragmatic influences on reasoning and for the relation of accuracy in reasoning to individual differences based on cognitive capacity. Thus, he concludes that research with logical conditionals and syllogisms has yielded evidences for ‘dual-process’ accounts and has proven that people can deal with logic, but are influenced by prior knowledge, instructions, time pressure, working memory load (mental effort), age, and general intelligence as can be seen in errors such as matching errors or ‘belief bias’.

This overview indicates that many theories try to explain deductive reasoning as well as the errors people make in deductive reasoning. Aside from the different assumptions, all theories explicitly postulate influences of and/or relations to prior knowledge, working memory capacity, and intelligence. The unsolved question is which of these theories explains reasoning and related errors best, and which of the approaches present proposals to avoid errors or to improve reasoning abilities. This is also one of the objectives of the current experiments. Before reporting the current experiments done within the framework of this thesis however, the findings of practical reasoning research shall first be reported.

1.5 Behavioral experiments on practical reasoning

Practical reasoning has mainly been investigated in the domain of morals, particularly with regard to children’s development of moral judgment abilities in the last century (e.g., Piaget, 1932). Due to the fact that morality research has been based on ideas of Piaget (1932), and later on Kohlberg (1969) who consider moral judgment levels as being developed during the process of growing-up and as being dependent on intelligence, it is not surprising that moral reasoning was considered to be based on rationality. Following this idea, the level of the ability to think rationally corresponds to the level of moral competence. Neither Piaget nor Kohlberg ever thought of emotions as part of the cognitive processes which contribute to moral judgments. Although it was Damasio (2006) who proposed in 1994 that emotions are encompassed in the cognitive process, it took until 2001 (Haidt, 2001) to take them into consideration again.[6]

However, since 2001 and along with Haidt’s article “The emotional dog and its rational tail: a social intuitionist approach to moral judgment” (Haidt, 2001), there had been a revival of the debate on the basis of moral reasoning. In practical reasoning research -as it is the case in theoretical reasoning research- the debate on the processes that might be involved persists as well. Different researchers postulate different processes accounting for moral reasoning and decision making.

[6] As mentioned earlier, in philosophy, different approaches exist postulating that emotions are relevant for practical reasoning and moral judgment making (e.g., Smith, 1759/ 2004) However, in the upcoming experimental research, these proposals seem to have been neglected since Haidt’s article in 2001.

Haidt (2001), for example, denotes the importance of intuitions and emotions in his theory as well as cultural influences for moral reasoning and judgment making. He especially doubts that (rational) reasoning causes moral judgments or guides moral actions. Haidt explicitly denotes four arguments for his assumptions:

> "[…] (a) There are two cognitive processes at work –reasoning and intuition– and the reasoning process has been overemphasized; (b) reasoning is often motivated; (c) the reasoning process constructs post hoc justifications, yet we experience the illusion of objective reasoning; and (d) moral action covaries with moral emotion more than with moral reasoning." (Haidt, 2001, p. 815).

Later in his article, he also states that the "Rationalist models focus on Links 5 and 6. In the social intuitionist model, in contrast, moral judgment consists primarily of Links 1-4 […]." (Haidt, 2001, p. 819). These links explaining moral reasoning and judgment making refer to different processes: links 1-4 refer to rather social-intuitive processes whereas the last two links refer to rather rational processes. One might assume then that Haidt's position as claiming emotional/intuitive processes to be involved in moral reasoning and judgment making rather than denoting rational processes. More importantly, Haidt's assumptions were mainly based on theoretical evidences at that time and thus required further empirical support.

The experimental evidence for his theory came in the same year from Greene and colleagues (Greene, Sommerville, Nystrom, Darley, & Cohen, 2001), who showed that emotions seem to play a critical role in moral reasoning and judgment making. Greene et al. (2001) used moral dilemmas according to the "classical" trolley and footbridge dilemmas in a functional magnetic resonance imaging (fMRI) study. Since these two "classical" dilemmas paved the way for numerous experiments that followed and were as well target of critics[7], they will be quoted in full:

> "One such dilemma is the trolley dilemma: A runaway trolley is headed for five people who will be killed if it proceeds on its present course. The only way to save them is to hit a switch that will turn the trolley onto an alternate set of tracks where it will kill one person instead of five. Ought you to turn the trolley in order to save five people at the expense of one? Most people say yes. Now consider a similar problem, the footbridge dilemma. As before, a trolley threatens to kill five people. You are standing next to a large stranger on a footbridge that spans the tracks, in between the oncoming trolley and the five people. In this scenario, the only way to save the five people is to push this stranger off the bridge, onto the tracks below. He will die if you

[7] In the section on brain imaging studies, these critics will be further described.

> do this, but his body will stop the trolley from reaching the others. Ought you to save the five others by pushing this stranger to his death? Most people say no." (Greene et al., 2001).

The dilemmas in Greene et al.'s (2001) experiment were derived from these two classical versions. Greene et al. (2001) applied personal and impersonal moral dilemmas as well as non-moral neutral stories in the scanner. They declared the personal moral dilemmas as personal instead of impersonal because these dilemmas produce serious bodily harm to others, were emotionally more salient, and deflected an existing threat actively onto a different person/group of people. In the experiment, their participants had to judge the action described in each dilemma as morally appropriate or inappropriate. Greene et al. (2001) found that participants needed more time to judge a 'footbridge dilemma version' as appropriate compared to a 'trolley dilemma version'. Greene et al. (2001) concluded that we normally tend to judge a 'footbridge version' as inappropriate because of an intensive emotional response attached to this type of dilemma. This emotional response has first to be overridden before we can make an opposite judgment. As to their imaging results, Greene and colleagues (Greene et al., 2001) found increased brain activity for the moral personal condition as compared to both of the other conditions in the medial frontal gyrus (BA 9/10), the posterior cingulate (BA 31), and the bilateral angular gyrus (BA 39). In contrast, the right middle frontal gyrus (BA 46) and the bilateral parietal lobe (BA 7/40) showed less activity in the personal moral conditions as compared to both of the others. Greene et al. (2001) denominate the former areas as associated with emotional processes whereas the later ones are related to working memory processes.

After these first evidences in experimental research showing that emotional/intuitive processes seem to play an important role in moral reasoning and judgment making, further proposals were made thus igniting diverse researchers to argue about varying positions assuming different processes involved in moral reasoning and judgment making.

Rational processes are stressed by representatives of one group (e.g., Kohlberg, 1969; Rest, 1979; Schaich Borg, Hynes, Van Horn, Grafton, & Sinnott-Armstrong, 2006), emotional/intuitive processes by representatives of another group (e.g., Haidt, 2001; Pizarro, Uhlmann, & Bloom, 2003). A completely different approach focuses on genetic and inborn facets. Representatives of this approach claim that emotions or rationality only come into play after a moral judgment had been made (e.g., Hauser, 2006, 2007; Mikhail, 2007). Apart from these rather extreme positions, the latest attempts tried to reconcile emotional and rational proportions by proposing 'dual-process' models or even 'threefold-process' accounts (e.g., Bucciarelli, Khemlani, & Johnson-Laird, 2008; Greene, 2007; Greene et al., 2008; Monin, Pizarro, & Beer, 2007; Moll, Zahn, De Oliveira-Souza, Krueger, & Grafman, 2005; Nichols, & Mallon, 2006). In addition to these different models, several mediating or other influential factors were discussed, among them are gender differences (Björklund, 2003; Gilligan, 1982; Skoe, Eisenberg, & Cumberland, 2002), time pressure (Björklund, 2003), personality traits (Stojiljković, 1998), and impacts of audience (Finger, Marsh, Kamel, Mitchell, & Blair, 2006).

Aside from the idea of an innate moral grammar (Hauser, 2006, 2007), there exist many proposals such as morality is learned and thus dependents on knowledge and experience (Bore, Munro, Kerridge, & Powis; 2005; Bucciarelli et al., 2008; Churchland, 1998; Goodenough & Prehn, 2004; Hare, 1981; Luo, Nakic, Wheatley, Richell, Martin, & Blair, 2006; Moll et al., 2005; Pasupathi & Staudinger, 2001). This is in line with other models suggesting domain-specific knowledge, acquired through learning and experience as being crucially involved in reasoning (Evans, 2003; Knauff, 2007), and thus matches ideas that can be found in the domain of theoretical reasoning research as well.

Some researchers had also shifted in their position due to new findings in further experiments they conducted (e.g., Greene, 2007; Greene, Nystrom, Engell, Darley, & Cohen, 2004). Greene (2007) for example now proposes a ‘dual-process’ account of moral reasoning with an interaction of emotions and cognition (i.e., rationality) where different proportions might support or even interfere with each other in the judgment making process. Another ‘dual-process’ account was proposed by Goodenough and Prehn (2004) who claimed that moral emotions and knowledge-based rationality are part of the reasoning processes for moral decisions. Pizarro, Uhlmann, and Salovey (2003) also provided a ‘dual-process’ account, which emphasizes the importance of emotions but also stresses rational processes: “[…] although moral intuitions seem to be de facto guides when arriving at judgments of moral responsibility […] a more deliberative mindset can “undo” the effects of moral intuitions.” (p. 658). Others even propose three component models of moral reasoning, like Nichols and Mallon (2006), where “[…] cost/benefit analysis, checking for rule violations, and emotional activations.” (p. 540) are involved in moral reasoning. Casebeer (2003a) also proposes a ‘threefold-process’ model of moral judgment making with moral emotions, abstract moral reasoning and theory of mind (ToM), and assumes that moral decisions are a result of “[…] multiple cognitive subprocesses […]” (p. 841). The shortcut from ToM to Piaget (1932), who assumed that perspective taking is important for moral reasoning, seems evident. The third ‘threefold-process’ account has been introduced by Moll and colleagues (Moll et al., 2005) and is called ‘EFECs’ model (i.e., event-feature-emotion complexes). This model contains “[…] structured event knowledge, which corresponds to context-dependent representations of events and event sequences in the PFC; social perceptual and functional features, represented as context-independent knowledge in the anterior and posterior temporal cortex; and central motive and emotional states, which correspond to context-independent activation in limbic and paralimbic structures.” (Moll et al., 2005, p. 804). Evidently, their model is mainly based on brain research results and assumes complex interactional processes between emotional and rational proportions and is clearly dependent on learning, thus knowledge and experience-based.

Currently, Bucciarelli et al. (2008) suggested a theory of moral reasoning based on “[…] an account of inferences in general about permissible situations (Bucciarelli & Johnson-Laird, 2005), on a theory of emotions (Oatley & Johnson-Laird, 1996), and on an account of intuitions (Johnson-Laird, 2006).” (p. 121). They provide a few clarifications on different terms and point to the important difference between their concept of intuition and that of Haidt (2001) for example. In Haidt’s approach, intuitions and emotions are treated as almost identical whereas Bucciarelli et al. (2008) differentiate the two. For them, intuitions represent unconscious premises that yield to conscious conclusions in contrast to conscious reasoning where both processes are conscious, even though “[…] the process of reasoning is itself largely unconscious.” (Bucciarelli et al., 2008, p. 123). Conscious reasoning is the only process that extensively uses working memory and thus may overwrite intuitions. In their concept, emotions are made up of cognitive evaluations which can either be conscious or unconscious. For their theory it is also crucial that moral reasoning is based on moral propositions and these are in turn deontic propositions stating what is allowed or forbidden and how one should act. It is also important that deontic propositions are not always moral propositions; they can also concern rules in games or manners, and the like. Following that, they deliver four principles for moral reasoning and corresponding empirical evidence: 1. Indefinability of moral propositions; 2. Independent systems; 3. Deontic reasoning; 4. Moral inconsistency (for details see Bucciarelli et al., 2008). In their four experiments, Bucciarelli et al. (2008) yielded evidences for these four principles and some arguments against proposals of an ‘innate universal moral grammar’ (e.g., Hauser, 2006, 2007) or accounts postulating utilitarian approaches of moral reasoning (e.g., Greene et al., 2008).

Integrating these findings of practical reasoning research shows that there are a few relations and similarities to the domain of theoretical reasoning research. In both domains, different theories and models are available explaining the reasoning processes involved.

Meaning for both theoretical reasoning and practical reasoning research, current models propose the involvement of rational and emotional/intuitive processes along with influences and mediations of intelligence and working memory capacities as well as time pressure. The ‘dual-process’ accounts therefore, seem to explain the processes involved in practical reasoning best. Further evidences for these assumptions of theoretical and practical reasoning research have been shown through the use of brain imaging experiments, which the following section shall report.

1.6 Brain imaging experiments on theoretical reasoning

In the domain of theoretical reasoning, many brain imaging experiments have been conducted, exploring reasoning with and without content (e.g., Goel & Dolan, 2003a, 2003b). Goel and colleagues (Goel, Büchel, Frith, & Dolan, 2000), for example, conducted an fMRI study investigating deductive reasoning. They addressed the question on whether deductive reasoning is based on sentential linguistic processes or requires spatial manipulation and search processes. To test this, Goel et al. (2000) applied syllogistic reasoning sentences with and without semantic content. All sentences had to be judged according to their validity and were either valid or invalid, but the content-laden sentences also contained believable and unbelievable statements, thus their truth values were consistent with the validity status (congruent) or not (incongruent). They found a main effect of reasoning for both conditions with brain activations in a wide-spread network encompassing the bilateral cerebellum, the bilateral fusiform gyrus, the left superior parietal lobe, the left middle temporal gyrus, the bilateral inferior frontal gyrus, the bilateral basal ganglia, and the brain stem. The content-laden syllogisms alone activated the left middle/superior temporal lobe (BA 21/22), the left inferior prefrontal cortex (BA 44/45), the right cerebellum, and the bilateral basal ganglia. The content-free sentences in contrast recruited the bilateral fusiform gyrus (BA 18), the left superior parietal lobe (BA 7), the left inferior frontal lobe (BA 44/45), the right inferior frontal lobe (BA 45), the bilateral cerebellum, the bilateral basal ganglia, and the brain stem. The authors additionally conducted a conjunction analysis to obtain the common activations of content-based and content-free reasoning and found activations in the left inferior prefrontal cortex (BA 44), the left fusiform gyrus (BA 18), the right fusiform gyrus (BA 37), the bilateral basal ganglia, and the right cerebellum. Furthermore, Goel et al. (2000) determined the distinct activations of content-based reasoning in contrast to content-free reasoning. The comparison of content-based minus content-free sentences revealed activations in the left middle/superior temporal lobe (BA 21/22), the left inferior frontal lobe

(BA 47), the bilateral striate cortex (BA 17), and the bilateral lingual gyri (BA 18). The inverse contrast yielded activations in the bilateral occipital cortex (BA 18/19), the bilateral superior and inferior parietal lobes (BA 7/40), the bilateral precentral gyrus (BA 6), and the bilateral middle frontal gyrus (BA 6). From these findings the authors arrived to this conclusion: “Syllogistic reasoning implicates a widespread network involving occipital, temporal, and parietal lobes, prefrontal cortex, and, surprisingly, cerebellum and basal ganglia nuclei.” (Goel et al., 2000, p. 509). Further, they infer: “However, closer analysis reveals two dissociable anatomical networks for reasoning, with shared common circuits in basal ganglia nuclei, cerebellum, fusiform gyri, and left prefrontal cortical regions.” (Goel et al., 2000, p. 509). However, due to the distinct findings, Goel et al. (2000) also assume that the required left hemispheric temporal system for content-based reasoning reveals evidence for sentential linguistic processes, whereas the activated parietal areas support ideas of spatial reasoning systems. Therefore, the authors finally state “[…] that syllogistic reasoning is implemented in two distinct systems whose engagement is primarily a function of the presence or absence of semantic content.” (Goel et al., 2000, p. 504). This study revealed evidence that reasoning per se requires a network shared with ‘pure’ reasoning as well as content-based reasoning. The content in this study however, appears to be not controlled for other factors (e.g., emotional load, prior knowledge), thus constraining the conclusions that could be made regarding content influences and potential differences of the reasoning processes involved associated with different contents. Another study by Goel and Dolan (2003a) was explicitly concerned with such a question.

In this fMRI experiment, Goel and Dolan (2003a) used syllogisms with emotional or neutral content. The items had either a valid or invalid conclusion and were again true (believable) or untrue (unbelievable) resulting in congruent and incongruent trials. Participants had to judge the validity of the sentences. Goel and Dolan (2003a) also implemented a baseline condition in which the third sentence was switched around so that no conclusion could be drawn. For the reasoning tasks as compared to the baseline, they found bilateral activations in striate cortex (BA 17), lingual gyri (BA 18), left cerebellum, left temporal lobe (BA 21/22), left temporal pole (BA 21/38), right basal ganglia nuclei, medial frontal gyrus (BA 6), left middle frontal gyrus (BA 6), and left inferior frontal gyrus (BA 44). Syllogistic reasoning with neutral content activated the left dorsolateral prefrontal cortex (DLPFC, BA 44/8), whereas syllogistic reasoning with emotional content activated the ventromedial prefrontal cortex (VMPFC) and the right fusiform gyrus (BA 37). Interestingly, the activation of the DLPFC suppressed the VMPFC activation and vice versa. The authors interpret this as a dynamic neural system for reasoning depending on emotional saliency. In another study by Goel and Dolan (2003b), deductive reasoning problems were also applied in the scanner. This time however, they were interested in the neural correlates of the ‘belief bias’ in deductive reasoning. The ‘belief bias’ effect (cp., Evans et al., 1983) refers to the fact that peoples’ deductive reasoning performance (e.g., validity judgments) is influenced by prior knowledge and beliefs. Thus, Goel and Dolan (2003b) applied syllogisms again, but this time with belief-laden contents. These syllogisms had valid or invalid conclusions and true (believable) or false (unbelievable) truth values. Of special interest was, whether the prepotent response due to beliefs or deductive reasoning dominates the participant’s response when there is a mismatch (i.e., incongruence) between believability and validity. As control condition, Goel and Dolan (2003b) applied belief neutral syllogisms, and the baseline consisted again of syllogisms wherein the third sentence was switched around so that no conclusion could be drawn. Participants showed more errors in the reasoning than in the baseline condition and were especially more prone to errors in the inhibitory/incongruent reasoning condition indicating a ‘belief-bias’ effect. The brain imaging data yielded a left temporal lobe system activated for belief-based reasoning and bilateral parietal activations for belief-neutral reasoning. When participants solved the deductive problems correctly, even though there was a conflicting belief, the right lateral PFC was activated. The authors interpret this activation as cognitive monitoring. When the participants failed to complete the deductive problems correctly, due to a ‘belief bias’, the VMPFC was activated. Goel and Dolan (2003b) assume that ‘belief bias’ effects in reasoning are mediated through effects of emotional processes on reasoning.

Canessa and colleagues (Canessa et al., 2005) conducted a similar fMRI study, but were interested in the effects of social content on deductive reasoning instead of the influence of beliefs. They applied the Wason selection task (WST) filled with different contents. One version described an arbitrary relation between two actions; the other version was a social-exchange relation between two persons. The analysis of both versions compared to a baseline revealed activations in the medial frontal cortex, the left DLPFC, and parietal regions. The authors conclude that these regions provide evidence for the major role of the left hemisphere in deductive reasoning. Comparing both of the reasoning versions yielded additional activations for the social-exchange condition in right fronto-parietal regions. According to Canessa et al. (2005), these right hemispheric activations seem to be required for reasoning given social contents.

The reported studies so far indicate that theoretical reasoning is based on a wide-spread network of brain areas. Moreover, investigating theoretical reasoning yielded evidences that reasoning might be based on more than just rational processes, but is also influenced by emotions (e.g., Goel & Dolan, 2003a), prior knowledge (e.g., Goel & Dolan, 2003b), and the imageability of the reasoning problems' content, for example (e.g., Knauff, 2009b; Knauff, Fangmeier, Ruff, & Johnson-Laird, 2003; Ruff, Knauff, Fangmeier, & Spreer, 2003).

A few articles present overviews about the results of theoretical reasoning brain imaging research –as for practical reasoning research– (e.g., Evans, 2008; Goel, 2007; Knauff, 2007, 2009a). Knauff (2007), for example, describes a fronto-parieto-temporo-occipital network with hemispheric specialization, depending on stimulus material/content, task demands, etc., for deductive reasoning. Lieberman (2003, cited in Evans, 2008) proposes a reflexive (System 1) and a reflective (System 2) system for cognitive processes, which correspond to neurological X- and C-systems. The X-system includes the basal ganglia, the amygdala, and the lateral temporal cortex, whereas the C-system encompasses the ACC, the prefrontal cortex, and the medial temporal cortex (including hippocampus). Evans (2008), in turn, assumes that these two systems yield evidence for a 'dual-process' account of deductive reasoning. A summary of the same diversity –as suggested in moral reasoning– for deductive reasoning is presented by Goel (2007): "[…] human reasoning is underwritten by a fractioned system that is dynamically configured in response to specific task and environmental cues." (p. 439). He also proposes that "The main effect of reasoning activates a large bilateral network including occipital, parietal, temporal, and frontal lobes, basal ganglia, and cerebellar regions." (Goel, 2007, p. 438). As with the practical reasoning research domain, Goel (2007) also suggests a 'dual-process/-system' account for deductive reasoning since the brain regions involved are associated with emotional/intuitive and rational processes.

1.7 Brain imaging experiments on practical reasoning

One of the first brain imaging studies investigating practical reasoning, especially moral reasoning and judgment making, has already been reported above (Greene et al., 2001). Since then, many experiments on moral reasoning and decision making have been conducted. As mentioned earlier, many of these studies utilized moral dilemmas, but pictures, sentences, or stories were also used. Heekeren and colleagues (Heekeren, Wartenburger, Schmidt, Schwintowski, & Villringer, 2003) for example applied simple ethical sentences that were unambiguous and not highly emotional laden describing morally appropriate or inappropriate actions according to the norms of our culture. As control condition they used neutral sentences that were semantically correct or incorrect. They found activations for the moral tasks compared to the semantical tasks in the left posterior superior temporal sulcus (STS, BA 39), the left middle temporal gyrus (BA 21), bilaterally temporal poles (BA 38), the lateral PFC (BA 47), the bilateral VMPFC (BA 9/10), and the right cuneus (BA 17/18). The semantic task activated the bilateral DLPFC (BA 46), the right precentral gyrus, and the left caudate gray. The authors conclude that especially the posterior STS and the VMPFC are involved in moral decision making, independent of complexity or emotional aspects. Further, Heekeren et al. (2003) associate the temporal pole activations with the fact that their participants had to judge the actions of others (i.e., whether they are good or not). The authors therefore suggest that these decision processes include reasoning about intentions which is a feature of the Theory of Mind (ToM). ToM is known as "[…] the ability to think about mental states, such as thoughts and beliefs, in oneself and others […]" (Premack & Woodruff, 1978, cited in Carrington, & Bailey, 2009). Heekeren et al. (2003) also associate the STS with detecting intentionality and ToM since it has been proposed that the STS is involved in the representation of the actions of others (cp., Frith & Frith, 1999).

Another study by Heekeren and colleagues (Heekeren, Wartenburger, Schmidt, Prehn, Schwintowski, & Villringer, 2005) investigated the influence of bodily harm/highly arousing emotions on moral decision making. This time, they used semantic correct/incorrect and morally appropriate/inappropriate sentences with and without bodily harm in the scanner. Materials with moral decisions and the sentences with bodily harm were processed faster than the semantic sentences or the sentences devoid of bodily harm. Moral decision making in contrast to semantic decision making produced greater activations in the VMPFC (bilateral, BA 10/11), the right VMPFC (BA 9), the posterior STS (bilateral, BA 39), the right posterior cingulated cortex (PCC, BA 31), and the temporal pole (bilateral, BA 21). Comparing semantic decision making and moral decision making yielded increased activations in the DLPFC (middle frontal gyrus, BA 46), the insula (BA 13), and the supramarginal gyrus (BA 40). Bodily harm compared to no harm in both conditions showed reduced activation in the bilateral temporal poles (middle temporal gyrus, BA 21). Heekeren et al. (2005) assume that the reduced activation in the temporal poles coupled with the faster reaction times might be a sign of processing advantage and reduced processing depth for items containing bodily harm. The authors then propose: "Weaker activity in the temporal poles might be a correlate of limited generation of the semantic and emotional context during the evaluation of actions of another agent related to bodily harm that is made with respect to the norms and values guiding our behavior in a community." (Heekeren et al., 2005, p. 895).

Moll, Eslinger, and De Oliveira-Souza (2001) also investigated participants in the scanner using moral sentences and factual/neutral control sentences. However, they were not only interested in potential emotional influences on moral reasoning, but also in differences concerning the judgment complexity of the various sentences applied. The participants had to decide silently whether the content of the sentences was right (i.e., factual correct or morally appropriate) or wrong (i.e., factual false or morally inappropriate). After scanning, the participants had to rate the moral content, the emotional load, and the judgment difficulty of each sentence. To control the emotional proportions of the judgments in the brain activations, the emotional ratings served as additional regressor in the fMRI data analysis. The moral sentences showed the highest degree of moral content and emotional load. In the moral judgment condition the authors found the frontopolar cortex (FPC), the medial frontal gyrus, the right anterior temporal cortex, the lenticular nucleus, and the cerebellum more activated compared to the control condition. Activation of the FPC and the medial frontal gyrus (BA 10/46 and 9) was still present after controlling for emotional load. The authors therefore conclude that the frontal and temporal regions especially contribute to complex judgment processes that encompass moral decision making. A further Study by Moll and colleagues (Moll, De Oliveira-Souza, Bramati, & Grafman, 2002) presenting short statements with unpleasant emotional, unpleasant moral, and neutral content yielded similar brain activations for the moral sentences. The moral items activated the medial OFC, the temporal pole, and the left STS. In another experiment, Moll and colleagues (Moll, De Oliveira-Souza, Eslinger, Bramati, Mourão-Miranda, Andreiuolo, & Pessoa, 2002) did not apply sentences or statements, but used pictures of the International Affective Picture System (IAPS, Lang, Bradley, & Cuthbert, 1995) with moral (unpleasant), negative emotional, and neutral content. The results revealed a basic emotional network consisting of amygdala, thalamus, and upper midbrain, which was activated in the moral and in the emotional condition as compared to the neutral one. Additionally, the authors found the orbital and medial PFC as well as the STS only activated for the moral pictures compared to the emotional or neutral ones (notably, these structures were still activated after controlling for emotional valence and visual arousal). Since the activated brain structures somewhat differed from the Moll et al. study in 2001 (where participants had to make active judgments), the authors postulate that the OFC might be more involved in automatic processing of moral emotions, while explicit moral reasoning requires additional prefrontal regions. These slightly diverging findings and interpretations of their two experiments seem to indicate that task demands per se might also influence the brain activations found in moral reasoning and judgment making experiments.

Therefore, Luo and colleagues (Luo et al., 2006) chose a totally different approach investigating morals. They criticize existing studies about moral reasoning and decision making (e.g., Greene et al., 2001). Their main point is that former studies required explicit decisions about moral questions from the participants. “However, such measures are susceptible to voluntary control and allow a participant the ability to conceal their genuine attitudes.” (Luo et al., 2006, p. 1450). Therefore, the authors adopted the Implicit Association Test (IAT, Greenwald, McGhee, & Schwartz, 1998) which assesses automatic and implicit attitudes, and filled it with moral content. Furthermore, they used pictures of the IAPS (Lang et al., 1995; Lang & Greenwald, 1985) depicting legal or illegal scenes with either highly arousing or low arousing character. They also applied pictures with animals, eliciting negative or positive emotions. Due to the fact that they presented pictures with illegal scenes and positively associated animals as well as pictures with legal scenes and negatively associated animals (and vice versa), their stimulus material contained congruent and incongruent conditions. Luo et al. (2006) found that participants responded slower in the incongruent trials as compared to the congruent ones. Participants were also slower for the low arousing stimuli as compared to the high arousing ones, and were finally slower for the legal stimuli as compared to the illegal ones. They suggest that this IAT effect (i.e., incongruence effect) “[…] is consistent with our suggestion that healthy individuals have an “automatic” association between, in particular, illegal/immoral behaviors and negative valence.” (Luo et al., 2006, p. 1452). Concerning the brain imaging results, they obtained increased activity for the high arousing stimuli in the ventromedial orbitofrontal cortex and the right amygdala. The incongruent trials yielded increased activity in the right ventrolateral prefrontal cortex (BA 47), left subgenual cingulate gyrus (BA 25), bilateral premotor cortex (BA 6), and the left caudate. Comparing illegal with legal trials revealed activations in the right medial frontal gyrus (BA 6), left precentral gyrus (BA 4/6), the postcentral gyrus (BA 3), and the superior temporal gyrus (bilateral, BA 22/41). The authors interpret their results proposing “[…] that an individual’s automatic moral attitude to an event involves an integrated neural response involving both the amygdala and medial orbitofrontal cortex that is proportional to the emotive strength (due to previous learning) of the stimulus.” (Luo et al., 2006, p. 1454). They believe that this neural circuit provides the evaluation of an event or stimulus as good or bad, but does not provide the final moral judgment: “[…] a full judgment of an action’s immorality is more than an automatic moral attitude of the action’s ‘badness’.” (Luo et al., 2006, p.

1455). As in the other studies referred above, Luo et al. (2006) associate the STS activations with processing intentionality and therefore, ToM.

An interesting attempt investigating practical reasoning may be seen in the brain imaging study done by Fiddick and colleagues (Fiddick, Spampinato, & Grafman, 2005). This seemed to be the first study which combined theoretical reasoning tasks (i.e., conditionals) with practical reasoning. However, they did not investigate theoretical reasoning directly, but only used conditionals as a basis for investigating practical reasoning. They contrasted social contract rules with non-social precaution rules, both based on conditional sentence structures. The social contract rules activated the dorsomedial PFC (BA 6/8), the bilateral ventrolateral PFC (BA 47), the left angular gyrus (BA 39), and the left orbitofrontal cortex (BA 10). In contrast, the precaution trials used activated the bilateral insula, the left lentiform nucleus, the posterior cingulate (BA 29/31), the anterior cingulate (BA 24), and the right postcentral gyrus (BA 3). The final conjunction analysis which the authors conducted to find common activations for reasoning about prescriptive rules yielded activations in the dorsomedial PFC (BA 6/8). Since Fiddick et al. (2005) found different brain structures that were activated for social and non-social rules, they conclude "[…] that human reasoning is not a unified phenomenon, but is content-sensitive." (Fiddick et al., 2005, p. 778). Due to their findings, they do not believe that reasoning could be explained by 'dual-process' accounts with its intuitive, context-sensitive processes on the one hand, and the more rational, decontextualized processes on the other hand. However, Fiddick et al. (2005) also admit that their results cannot exclude the possibility of rational processes involved in reasoning, but they reject proposals that "[…] nonlogical reasoning is collectively assigned to the realm of intuition." (p. 785). As a critic of this study it could be commented that they did not investigate theoretical reasoning but only practical reasoning, thus drawing such far reaching conclusions seems inadequate. Further, it seems as if they misleadingly intersperse the terms 'intuition', 'context', and 'content'. Finally, the multiple content-dependent reasoning systems they propose appear to refer only to the well-known fact that diverse contents can influence reasoning differently (e.g., Evans, 2008). In particular, when this content is emotionally laden, a fact (c.p., Goel & Dolan, 2003a) Fiddick et al. (2005) did not control in their study.

At this point, the studies reported seem to indicate that moral reasoning and judgment making is based on a large network of brain areas, which are associated with diverse processes and concepts. Several review articles describe and merge different findings of even more studies and come up with similar assumptions (e.g., Allison, 2001; Casebeer, 2003a; Greene & Haidt, 2002; Moll, De Oliveira-Souza, & Eslinger, 2003; Moll et al., 2005; Raine &

Yang, 2006; Young & Koenigs, 2007). Allison (2001), for example, summarizes that the superior temporal sulcus (STS), the orbitofrontal cortex (OFC), the amygdala, the prefrontal cortex (PFC), the motor cortex, and the basal ganglia are involved in moral reasoning and judgment making. Others, like Greene and Haidt (2002), claim that the medial frontal gyrus (MFG, BA 9/10), the posterior cingulate gyrus, the precuneus, the retrospenial cortex (BA 31/7), the STS, the inferior parietal lobe (IPL, BA 39), the orbitofrontal and ventromedial frontal cortex (VMPFC, BA 10/11), the temporal pole (BA 38), the amygdala, the dorsolateral prefrontal cortex (DLPFC, BA 9/10/46), and the parietal lobe (BA 7/40) represent the structures involved.

Despite some variety between studies, the following brain structures are repeatedly and consistently discussed as being involved in moral reasoning[8]: DLPFC, mPFC, VMPFC, OFC, cingulate cortex (anterior and posterior), STS-region, and temporal pole (quite often also parietal and occipital regions). This list demonstrates some unity or overlap of brain structures involved in moral reasoning, yet it seems astounding that so many studies using different tasks and different stimuli produce such an extreme overlap (cp., Goodenough & Prehn, 2004; Moll et al., 2005).

A further critique (e.g., Wiedenmann & Knauff, 2008) on many studies is that they deal with items (mainly dilemmas), not validated and not evaluated according to an external (standardized) criterion (e.g., Ciaramelli, Muccioli, Làdavas, & Di Pellegrino, 2007). Moreover, these items are far from ‘everyday morality’, are complicated, highly negative emotional laden (Casebeer & Churchland, 2003; Harenski & Hamann, 2006; Moll et al., 2005; Waldmann & Dieterich, 2007), and thus not really adequate for exploring moral reasoning (we rarely get to choose between killing one person instead of five as required in the trolley-dilemma or footbridge-dilemma).

Nevertheless, these findings again seem to support ‘dual-process’ accounts as well as some overlap between the brain areas involved in practical and theoretical reasoning. However, before integrating the different findings and explaining the subsequent paradigm applied in the current work as well as the leading questions derived from the former findings, three excurses need to be done. Since mediating influences of working memory (WM) on reasoning have been discussed, a short explanation of WM processes and models seems appropriate. Further, two of the following experiments also implemented a recognition task, thus, a short overview on recognition memory research shall be presented. Finally, there are a

[8] Different concepts or labels are used quite often for similar brain areas, leading to irritations about which brain structures are involved. Especially for the frontal brain areas and the STS-region, no commonly accepted labels with clear borders seem available.

few researches that have already postulated a common basis for theoretical and practical reasoning, or else, a "unique" reasoning system per se. Therefore, these models will also be briefly referred.

1.8 Working memory

"Working memory is a limited capacity system required for the ability to maintain and manipulate information over short periods of time (e.g., a few seconds) in the service of other cognitive tasks (e.g., problem solving) […]" (Kensinger & Corkin, 2003, p. 378). Thus, working memory provides the necessary information at a specific time and a specific point that is required to execute a specific problem (Johnson-Laird, 2006). However, as mentioned above, emotionally laden material can disturb reasoning performance on deductive reasoning problems, for example (Blanchette & Richards, 2004). Other tasks may also impede working memory performance due to influences of emotional salient material (Kensinger & Corkin, 2003).

One of the best known working memory models has been proposed by Baddeley and Hitch in 1974. The model was modified and developed further, but its current version seems to resemble the one suggested over 30 years ago (Baddeley, 2007). It contains three components which serve different purposes: the phonological loop, the visuo-spatial sketchpad, and the central executive (cp., Baddeley, 2003).The phonological loop stores and manipulates auditory and verbal information, whereas the visuo-spatial sketchpad is concerned with visual and spatial information. Both systems have only limited capacity and information gets lost, if it is not held actively in memory, e.g. via sub-vocal rehearsal. The central executive coordinates and monitors these systems and guides attention to focus on task demands and relevant details as well as to inhibit processing of task irrelevant information (e.g., Baddeley, Emslie, Kolodny, & Duncan, 1998). Thereby, the attentional functions of the central executive seem to correspond to the 'Supervisory Attentional System' proposed by Norman and Shallice (1986). Involved working memory brain areas, especially for central executive-related processes are the frontal lobes and parietal lobes (cp., Baddeley, 2003).

A totally different working memory model has been proposed by Postle (2006), in which "Working memory functions arise through the coordinated recruitment, via attention, of brain systems that have evolved to accomplish sensory-, representation-, and action-related functions." (p. 23). Postle (2006) argues with his 'emergent-property-model' against the assumptions suggested by Baddeley and colleagues, who stated that working memory functions require specific and specialized sub-systems. Instead, he assumes that working memory functions are based on the same systems that serve these functions while processing information independent of working memory processes (e.g., perception).

Apart from the chosen working memory model, many experiments have been conducted to investigate working memory processes and their influences on deductive reasoning as well as the relationship between working memory and intelligence. Thus, working memory and deductive reasoning tasks are often presented in combination to test which kind of working memory load interferes most with the reasoning processes, and is therefore involved in reasoning. The influence of competing working memory tasks on syllogistic reasoning has been investigated for example by Gilhooly, Logie, Wetherick, and Wynn (1993). They found performance decrements in syllogistic reasoning when participants had to solve a random number task in parallel. The generation of random numbers in turn is associated with central executive WM functions. In contrast, enhancing the cognitive load of the phonological loop via repeating a series of numbers did not influence the reasoning performance. An additional burden on the visuo-spatial sketchpad also yielded no interfering effects. Vice versa, Gilhooly et al. (1993) showed that syllogistic reasoning impaired the working memory functions and thus performance of the phonological loop and the central executive. Gilhooly and colleagues (Gilhooly, Logie, & Wynn, 2002) conducted a further experiment, but modified the procedure. They presented the premises of the syllogisms successively, which resulted in even worse reasoning performance. This time, the performance of all competing working memory tasks was affected. The authors conclude that successive premise presentation increases working memory load in syllogistic reasoning. Another study by Toms, Morris, and Ward (1993) applying a ‘dual-task’ paradigm with conditionals provided further support that the central executive is involved in deductive reasoning. A further study conducted by Klauer, Stegmaier, and Meiser (1997) points in the same direction.

Other studies could show that participants with a higher working memory capacity solve syllogistic reasoning problems as well as conditional reasoning problems better than persons with lower working memory capacity (e.g., Barrouillet & Lecas., 1999; Capon, Handley, & Dennis, 2003; Copeland & Radvansky, 2004; Markovits, Doyon, & Simoneau, 2002; Verschueren, Schaeken, & D'Ydewalle, 2005a, 2005b). Verschueren et al. (2005a, 2005b) explain these findings by relating that persons with superior WM capacity have an easier access to deliberate analytical reasoning processes, whereas those with lower WM capacity rely more on intuitive and heuristic processes. Working memory capacity in turn is often associated with reasoning abilities in general (e.g., Bara, Bucciarelli, & Johnson-Laird, 1995; Barrouillet & Leças, 1999), which again are associated with intelligence (e.g., Stanovich, 1999; Stanovich & West, 1998; Stanovich & West 2000). Stanovich (1999), for example, cites numerous studies revealing a strong correlation between higher intelligence levels and better reasoning performance as well as higher working memory capacities. He concludes that persons with a superior IQ (i.e., intelligence quotient) have either an easier access recruiting rational processes or are able to use these rational processes more efficiently.

To sum it up, working memory processes seem to be involved in reasoning, especially the central executive of Baddeley's (2007) WM model, and reasoning performance appears to depend on WM capacity which is strongly related to intelligence level. Besides WM influences on theoretical and practical reasoning, the bearings of prior knowledge and beliefs are often discussed. Therefore, two of the experiments reported in the following section additionally applied a recognition test implementing prior knowledge about morals. So, as was mentioned earlier, it is only helpful to refer to a few findings on recognition research.

1.9 Recognition

Recognition is "[…] the ability to judge a recently encountered item as having been presented previously." (Squire, Wixted, & Clark, 2007, p. 872). Within recognition memory, two components are usually differentiated: recollection and familiarity (Mandler, 1980). "Recollection involves remembering specific contextual details about a prior learning episode; familiarity involves simply knowing that an item was presented, without having available any additional information about the learning episode." (Squire et al., 2007, p. 872). Tulving (1985, 2002) introduced similar terms (i.e., remember/know) to differentiate between episodic and semantic memory processes. These terms are also used for experimental paradigms (remember/know procedure) exploring the influences of familiarity and recollection on recognition performance (Baddeley, Eysenck, & Anderson, 2009). Recollection is assumed to be a controlled, slow, effortful, conscious, and attention demanding process, whereas familiarity judgments are fast, unconscious, automatic, and require less effort (Yonelinas, 2002). Apart from the concrete memory model, which actually seems to be the case, all models propose two processes involved recognition memory thus representing 'dual-process' accounts. In the following, a few experimental paradigms and methods investigating recognition memory as well as the results of diverse studies will be reported.

A typical word recognition test requires participants to learn a list of words which should be recognized later. After a delay period, participants are presented with a list of words again. However, this time, not only the learned words (i.e., targets) are included, but also new, unlearned ones (i.e., distracters/foils). The participants' task then is to judge whether they have seen/learned the word before or not. To evaluate participants' recognition performance, it is important to consider not only correct (i.e., hits) or incorrect recognized words (i.e., false alarms), but also missed words that were on the list before (i.e., misses) as well as correctly rejected words (i.e., rejections) that were not earlier on the list. Calculating the ratio of the participants' answers according to these different categories allows evaluating the recognition memory performance. This could be done with the 'signal-detection' theory for example, illustrated in a 'receiver operating characteristic' (ROC) graph (Squire et al., 2007). The "[…] signal-detection theory holds that each old item (or target) or new item (foil) is associated with a particular memory strength, which reflects the degree of certainty that an item did or did not appear on a recently presented list." (Squire et al., 2007, p. 874). Plotting the results of such an experiment on a graph illustrates potential performance differences of discriminating targets from foils since "[…] the mean and variance of the target distribution are greater than those of the foil distribution." (Squire et al., 2007, p. 874). The ROC curve shows the false alarms and hit rates of the participants together with a criterion value. Targets above the criterion value are correctly recognized, whereas foils above the criterion represent false alarms.

Other recognition tests are the 'remember/know procedure' by Tulving (1985) or the 'process dissociation procedure' by Jacoby (1991). These methods aim to explore the differential influences of recollection and familiarity on recognition performance. The remember/know procedure explicitly asks participants in the test phase whether they can remember the word together with additional details of the learning period, or whether they only have a feeling of familiarity with the item at hand. It is assumed that remember responses indicate recollection processes, whereas know responses measure familiarity (Yonelinas, 2002). The process dissociation procedure allows an even more reliable measure. Participants are split in two groups. Both groups learn two lists of words, presented once visually and once auditory. Then, in the test phase, one group should answer "Yes" to all words it recognizes (i.e., inclusion condition), the other group to only words it recognizes from the auditory learning condition (i.e., exclusion condition). If participants of the second group answer "Yes" to an old item that was on the visually presented learning list, these judgments are counted as familiarity-based. The ratio of the errors as compared to the overall correct recognized words of the inclusion group is then taken as evidence for the involved recollection performance (cp., Baddeley et al., 2009).

Jacoby and colleagues (Jacoby, Woloshyn, & Kelley, 1989) investigated the influence of divided attention on recollection and familiarity-based recognition processes. Divided attention impaired only recollection performance, whereas familiarity-based judgments were unaffected. In an overview of diverse experiments, Jacoby (1991) takes these and further results as evidence that recognition depends on two separate processes, thereby concluding that recollection is an intentional, attention demanding process, and familiarity is unconscious and automatic. Yonelinas (2002) also mentions a few evidences in his review that support 'dual-process' models explaining recognition performance as based on recollection and familiarity processes. First, recollection processes seem to be easier to disturb by distracters, both at the learning phase and in the test phase. A possible explanation might be that attention is divided and thus easily distracted that encoding of the learning situation fails, but a feeling of familiarity with the items could be built. Further, recollection judgments normally require much more time than familiarity judgments. Moreover, recognition performance decreases especially in older adults or patients with prefrontal cortex damages, when attentional resources are impaired. Yonelinas (2002) follows from these findings that recognition retrieval is based on two qualitatively distinct processes. These effects of divided attention tasks on recollection might also be linked to dual task effects on working memory. From experiments on working memory performance, it is known that a secondary task impairs performance (e.g., De Neys, Schaeken, & D`Ydewalle, 2005). Furthermore, working memory, especially the central executive in Baddeley's (cp., Baddeley, 2003) working memory model, has been associated with attentional processes. Thus, the impairing effects on recollection of the divided attention tasks in recognition could account for reduced working memory capacities which also explain impaired performance in dual task experiments.

Another explanation of the two different processes involved in recognition might be that, in general, different learning processes occur. Some materials are processed more deeply (i.e., 'depth of processing), thus are remembered better, whereas other materials are processed more cursorily (i.e., shallow processing), activating only a feeling of familiarity (Craik & Lockhart, 1972; Craik & Tulving, 1975).

In addition, the content of the learning material seems to play an important role in recognition performance. Ohira, Winton, and Oyama (1998) for example, investigated the recognition performance with negative and positive as well as neutral words, and found a better recollection-based recognition performance for the negative emotional stimuli as compared to both of the other conditions. These results were replicated in a study by Inaba, Nomura, and Ohira, (2005), with the negative emotional words being recognized best, followed by the positive emotional ones, and finally the neutral stimuli. They had however, reported a side effect where positive emotional words resulted in higher false alarm rates. In contrast to these experiments, Phelps, LaBar, and Spencer (1997) found a slightly better retrieval performance of positive compared to negative emotional stimuli, even though all affective laden words were recognized better than the neutral controls. An interesting study investigating attentional as well as emotional influences on recognition performance was conducted by Ferré (2003). He found that emotionally laden materials are always recognized better than neutral materials, even if participants did not focus on the affective valence of the items during the learning phase. To sum up, these studies show that emotionally laden materials are better recognized than neutral stimuli. An explanation might be that emotional material 'catches' attention, and is thus processed deeper (Kensinger & Corking, 2003). However, the last study particularly showed that attention cannot be the only factor accounting for these effects. The side effect found in the study by Inaba et al. (2005) might also be a hint for disturbing effects of emotional material on recognition performance.

Another factor that might influence recognition is prior knowledge. Anderson (1974) conducted a study where participants had to learn sentences containing particular locations or persons. Some of the persons and locations appeared in two or three sentences thus could be associated with different sentences, whereas other persons and locations were only presented once. Later, in the test condition, the learned sentences and new sentences were presented and participants had to judge whether a sentence was from the study list or not. Participants needed more time to recognize sentences with multiple associations than sentences with persons and locations that only appeared once. This associative interference effect has become known as the 'fan effect' (Anderson, 1999), since the more associations a memory record has, the longer the serial search through all stored records takes till the correct one is found. In turn, interference increases the more associations with one and the same cue have to be learned. A study by Pirolli and Anderson (1985) revealed evidence that also highly over-learned memories could still be interfered by multiple associations, although in total practice had positive effects on recognition performance. Peterson and Potts (1982) also investigated the impact of prior knowledge on recognition performance. However, they used true and false new facts as well as facts participants already knew before, thus implementing a far more realistic study design than the experiments before. They also found that newly learned information interfered with the retrieval of previously known facts. In contrast, the newly learned facts consistent with prior knowledge revealed a facilitating effect on recognition performance.

The referred studies so far seem to indicate the recognition memory and thus recognition performance is related to and influenced by prior knowledge, emotional load of the material applied, attentional processes, and working memory capacities. Furthermore, ‘dual-process’ models of recognition assuming differential influences of recollection and familiarity might account best for the processes involved. Thus, there is evidence that recognition is concerned with similar processes and influencing factors as theoretical and practical reasoning. Now, three proposals will be reported that tried to combine some of the findings reported so far in “all-encompassing” models.

1.10 Reasoning, emotion, and the human brain

A few researchers already postulated some kind of common basis for practical and theoretical reasoning based on interactions of emotional/intuitive and rational processes in relation to diverse influencing factors (e.g., Adolphs, 1999, 2006; Damasio, 2006; Dolan, 2002; Le Doux, 2006; Thagard, 2006). They did not however, explicitly investigate theoretical and practical reasoning processes in combination, but instead tried to explore and to describe reasoning per se starting at a “more emotional perspective” since emotions were almost condemned for a while, only known as impairing factor of humans’ higher reasoning abilities (cp., Damasio, 2006). The resulting models range from Damasio’s (2006) ‘somatic marker hypothesis’ mainly based on patient studies, over Thagard’s (2006) computer-based assumptions about ‘hot thoughts’, to LeDoux’s (2006) ‘high and low road’ proposal in the emotional brain.

Damasio (2006) proposed his ‘somatic marker hypothesis’ in 1994, stating that emotions are in the loop of reason, and “[…] that emotion could assist the reasoning process rather than necessarily disturb it, as was commonly assumed.” (preface, p. xvii). His suggestions are based on the idea that emotions are the source of the development of the human reasoning system via phylogeny. “[…] Emotions *marked* certain aspects of a situation, or certain outcomes of possible actions. Emotions achieved this marking quite overtly, as in a “gut feeling”, or covertly, via signals occurring below the radar of our awareness […]” (Damasio, 2006, preface, p. xviii). Further, he suggests that emotions play a role in intuition. Intuition is accumulated knowledge which comes unconsciously into the mind guiding or influencing decisions. These intuitions are therefore part of all reasoning processes that deal with prior knowledge. Often, they provide fast decisions, based on an automatic and unconscious process, and are emotionally ‘marked’. However, “the quality of one’s intuition depends on how well we have reasoned in the past […]” (Damasio, 2006, preface, p. xix). In opposition to many misinterpretations of his proposals however, Damasio (2006) explicitly denotes that reasoning processes are based on emotions, rationality, and intuitions (i.e., over learned or over practiced prior knowledge), thus making emotions provide "[…] cognitive information, directly and via feelings." (preface, p. xix). This means that the reasoning processes can be unconscious and automatic via intuitions, or deliberative and conscious, and both of these processes can be influenced by emotions, either directly or via feelings. Hence, Damasio (2006) subsumes emotional/intuitive and rational processes under the term cognition and states "[...] that the brain systems that are jointly engaged in emotion and decision-making are generally involved in the management of social cognition and behavior." (preface, p. xix). He obtained evidences for his hypothesis and for the connection of intuitions and emotions as well as for the influence of emotionally laden knowledge on ethical decision making due to numerous patient studies. Comparing late-onset (i.e., in adulthood) and early-onset (i.e., in childhood) patients with damage to the ventromedial prefrontal cortex (VMPFC) provided insights in the development of human morality and the important function of emotions and intuitions:

> "[…] the early-onset patient appeared not to have learned the social conventions and ethical rules that should have governed their behavior. Whereas the adult-onset patients knew the rules but failed to act according to them, the early-onset case had never learned the rules to begin with. In other words, while the adult-onset cases told us that emotions were required for the deployment of proper social behavior; the early-onset cases showed that emotions were also needed for mastering the know-how behind proper social behavior." (Damasio, 2006, preface, p. xx).

So far, the general assumptions of the 'somatic marker hypothesis' (Damasio, 2006) have been outlined. Now, the basic ideas behind the terms 'somatic' and 'marker' will be explained more explicitly. If we are confronted, for example, with a decision between two choices, a rationalist view requires reasoning about the different possibilities accurately, perhaps conducting a cost/benefit analysis, weighing losses and gains, and applying formal logic to arrive at the best solution available. Unfortunately, attention and working memory are of limited capacity, thus, even simple decisions would lead to a "cognitive overload". This scenario provides the arguments for Damasio's proposal of 'somatic markers' influencing and therefore enabling us to even achieve a decision at all. Damasio (2006) claims that before we start to do a cost/benefit analysis we imagine different possible outcomes of a decision. Good outcomes are experienced as prospective 'good feelings', bad outcomes as prospective 'bad feelings'. These feelings are about the body and Damasio (2006) "[…] gave the phenomenon the technical term *somatic* state ("soma" is Greek for body); and because it "marks" an image, [he] called it a *marker*." (p. 173). So, 'somatic markers' reduce the number of possible alternatives via emotional feelings and prior knowledge, but do not exclude subsequent rational reasoning processes. "Somatic markers may not be sufficient for normal human decision-making since a subsequent process of reasoning and final selection will still take place in many though not all instances. Somatic markers probably increase the accuracy and efficiency of the decision process. Their absence reduces them." (Damasio, 2006, p. 173). In essence, these feelings based on emotions always activate the whole body while a person imagining various potential outcomes of possible decisions seems to be the main point for critics. However, the overall assumption that decisions are influenced by prior knowledge and emotions seems to be a position currently shared by many researchers.

An example of another researcher assuming an involvement of emotional and rational processes in reasoning is LeDoux (2006) and his suggestions about the emotional brain. The main difference to Damasio's proposal is that LeDoux derived his suggestions from animal research and focuses therefore on one special emotion, namely fear. LeDoux however, extends his assumptions to humans and embeds his theory in the broader framework of cognition. He starts with the proposal that emotions represent a biological function of the nervous system which appears to be rather automatic and unintended than controlled and planned. LeDoux (2006) thinks that this is due to the fact that the connections of the emotional brain systems to the more rational ones are stronger than the reverse connections. He also states that basic emotions like fear are inborn and that we humans share them with many animals. Further, he assumes that emotions and cognitions (i.e., rational processes in LeDoux's proposal) operate unconsciously and only the results of these processes are conscious. He also claims that emotions and cognitions are separated but interacting mental functions based on separated but interacting brain systems. Then, he defines the fear system as a system to detect dangers and produce defense reactions to increase the probability of survival. Therefore, this system has been developed before any conscious feelings of fear during phylogeny. LeDoux follows from this assumption that feelings of fear are only a byproduct of the evolution of two neural systems, one producing defense reactions and the other causing consciousness. Due to animal experiments investigating the amygdala, LeDoux found diverse evidences for his suggestions. He could show that animals react on threatening stimuli for example by releasing stress hormones, increasing blood pressure, and a startle response or by freezing. These reactions and somatic responses depend on diverse brain areas as the central grey, the lateral hypothalamus, the paraventricular hypothalamus, and the reticulopontine caudate to which the central core of the amygdala projects. However, if one or several of these output pathways of the amygdala to the respective areas are lesioned, the specific reactions and responses are missing. The interesting fact about this system is that the amygdala itself receives two kinds of input. First, it receives input directly from the thalamus, and then later on it receives input from cortical areas like the visual or auditory system. Humans possess a so-called 'high-road' (i.e., cortical pathway) and a 'low-road' (i.e., sub-cortical pathway) as pathways in the brain. The low-road is a fast and imprecise way preparing us to react to a threatening stimulus as fast as possible, even if we did not have the total information about the stimulus. It might be that the stimulus is not threatening at all. For example, while walking through the forest, one detects something looking like a snake. The low-road activates the fear system so that we could react with a flight response. If the high-road then provides us with the information that the potential snake is only a stock of a tree, the flight response may be deemed inadequate. Therefore, the high-road represents the analyzing system, telling us what kind of stimulus we are really confronted with, enabling us therefore to suppress a potential flight-or-fight response. Before the amygdala received the input from the sensory cortices, it does not know whether the occurring stimulus is really a threatening one or not; thus one might say the amygdala is blind and deaf. Another important structure contributing to the fear system is the hippocampus formation. The hippocampus and circumjacent brain regions provide us with the necessary information to analyze more complex situations and to store memories of emotional experiences. However, if the Hippocampus is damaged we are still able to perceive threatening situations and create an implicit emotional memory of them. Interestingly, implicit emotional memories and explicit memories of emotional experiences are connected via working memory which is responsible for our immediate conscious experiences. The interplay of all these cortical and sub-cortical areas enables us to evaluate stimuli as threatening or not, and provides us with the adequate reactions, the emotional feelings, and the memories required in adequately handling similar situations in the future.

Although LeDoux (2006) focuses almost exclusively on fear, the preceding explanations revealed evidences that emotion and cognition (i.e., rationality) are closely related and interdependent. Thus, it seems as if emotion and rationality belong together and are based on each other, or as LeDoux (2006) puts it: 'also the neurologist Antonio Damasio underlines in his book *Descartes' Error* the rationality of emotion' (p. 41, loose translation from the German version).

Similar to Damasio's and LeDoux's proposals, Thagard (2006) stresses the importance of interacting emotions, intuitions, and rationality during decision making, a procedure he calls 'informed intuition'. He provides diverse examples as to why intuitions and emotions are necessary for coherent and efficient reasoning processes and proposes several computer-based models to illustrate his assumptions. Thagard (2006) especially tries to show that classical philosophical theories as well as research based on logic-related tasks (e.g., syllogistic reasoning) are insufficient to explain or to investigate human reasoning. He states for example:

> "Practical inference is not simply produced by practical syllogisms or cost-benefit calculations, but requires assessment of the coherence of positively and negatively interconnected goals and actions. This assessment is an unconscious process based in part on emotional valences attached to the various goals to be taken into consideration, and yields a conscious judgment that is not just a belief about what is the best action to perform but also a positive emotional attitude toward that action. Reason and emotion need not be in conflict with each other […]." (Thagard, 2006, p. 24 f).

Thagard (2006) also implements Damasio's (2006) 'somatic marker hypothesis' with its neural foundations of integrative cognitive-affective processes into a neurocomputational theory of decision making, named GAGE. The brain structures involved in his model are the ventromedial prefrontal cortex (VMPFC), the nucleus accumbens (NAcc), the amygdala, the hippocampus, and the ventral tegmental area. The hippocampus provides the contextual information, and associated with the amygdala and the VMPFC, and, depending on memories, an emotional evaluation is provided making predictions about future outcomes. The NAcc receives dopaminergic VTA projections and represents a gateway for somatic markers, whereby "[…] the hippocampus determines what passes through this gateway by limiting throughput to those responses that are consistent with the current environment." (Thagard, 2006, p. 95). The emotionally laden predicted outcome of these circuits is then passed to brain areas associated with higher-level cognitive processes. Although Thagard (2006) tried to model the brain processes of 'informed intuition' with a computer-based approach, he admits that computer-based models will run into difficulties and may never resemble human reasoning and problem solving abilities due to the fact that they lack emotions. When open and fuzzy problems appear or non-routine directions are necessary, computer models will fail. Thagard (2006) regards this insufficiency of computer models and artificial intelligence as further evidence for the necessary involvement of emotional cognition in human reasoning and problem solving.

The only one of the three cited researches explicitly referring to theoretical and practical reasoning is Thagard (2006) in his chapter about ‘critique of emotional reason’. He claims that “Emotion is relevant to theoretical reason, which concerns what to believe, and to practical reason, which concerns what to do.” (p. 251). However, since the chapter is mainly concerned with abductive reasoning and the generation and evaluation of scientific research hypotheses no further details will be reported. Instead, the following outlet tries to integrate and subsume the different research areas and findings reported to derive appropriate assumptions for the following experiments and propose the leading research idea. However, a final methodological excursus on brain imaging seems necessary beforehand since one of the following experiments applied this sophisticated technique.

1.11 Excursus – The method of brain imaging

Brain imaging techniques, especially functional magnetic resonance imaging (fMRI) represent a sophisticated, non-invasive method for cognitive neuroscience to explore the human brain and its functions while executing cognitive processes (cp. Raichle, 1994). Thereby, “understanding the brain depends on conceptual, anatomical, statistical, and causal models that link ideas about how it works to observations and experimental data.” (Friston, 2005, p. 58). Besides relating anatomical correlates to cognitive processes, research also aims at investigating the functional correlates of cognitive processes, exploring causal dependencies, and finally, predicting human behavior through brain activity (cp. Friston, 2005). The basic idea behind the MR-technology is that neurons in the brain show similar properties like neurons in the rest of the body. This means, they need energy (i.e., oxygen and glucose) delivered via blood flow to “[…] sustain their cellular integrity and to perform their specialized functions.” (Gazzaniga, Ivry, & Mangun, 2009, p. 152). fMRI measures these metabolic changes which are assumed to correlate with neural activity (Jäncke, 2005). If parts of the brain become more active than others, they need more energy. This leads to a hemodynamic response, which increases the regional cerebral blood flow (rCBF) in these specific parts of the brain to provide the increased energy needs (Jäncke, 2005). The required oxygen is transported by hemoglobin which has specific magnetic properties. The different magnetic properties of the hemoglobin as well as those of the encompassing brain tissue rely on the nuclear spins of the hydrogen atoms involved. When oxygenated, diamagnetic hemoglobin reaches the activated cells, the oxygen is absorbed and the hemoglobin becomes deoxygenated, and thus paramagnetic. Importantly, the rCBF delivers much more oxygen as could be absorbed by the active cells, therefore results in a surplus of oxygenated hemoglobin.

A magnetic resonance tomograph (MRT) measures the ratio between oxygenated and deoxygenated hemoglobin, which is referred to as the blood oxygenation level-dependent (BOLD) effect (Ogawa, Lee, Nayak, & Glynn, 1990). The measured BOLD signal intensities increase if the concentration of oxygenated hemoglobin rises, since the blood magnetic susceptibility of oxygenated hemoglobin changes. It then closely matches the magnetic susceptibility of the circumjacent biological tissue (Kwong, Belliveau, Chesler, Goldberg, Weisskoff, Poncelet, Kennedy, Hoppel, Cohen, Turner, Cheng, Brady, & Rosen, 1992). Deoxygenated hemoglobin in contrast leads to susceptibility artifacts and signal decrements. To detect the small ratio differences in the brain, an MRT-scanner produces a powerful magnetic field (e.g., 1.5, or 3, or 7 Tesla). The magnetic moments of the protons of the water molecules (i.e., hydrogen atoms) in a brain align to this magnetic field. Then, a short electromagnetic radio frequency pulse is given to disturb this alignment for a few of these protons. When this pulse is turned off, the protons turn back to their original magnetization alignment (Jäncke, 2005). These alignment changes can be measured. However, the measured ratio or BOLD effect is extremely small, reflecting only one to five percent signal change. This small signal change relies on the fact that the whole brain is permanently active, so that an additional task requiring specific cognitive processes and associated brain areas only leads to a minimal change of activity. Thus, repeated measures of the same cognitive processes, and therefore activated brain areas, are necessary to obtain a better signal-to-noise ratio via averaging of trials.

Up to now, it has been stated that fMRI measures increased blood flow correlated with increased brain activity of specific neurons executing specific cognitive processes. However, the increased blood flow with its oxygenated hemoglobin takes a few seconds to be directed to a specific brain area. Furthermore, this blood flow reaches its maximum not before 5 to 6 seconds. In contrast, the neural firing of active neurons only requires milliseconds (Jäncke, 2005). One might assume therefore, that there is some kind of critical shortage of the brain cells involved before new energy is delivered by the blood stream. This should be reflected in an initial dip of the oxygen-deoxygen-hemoglobin ratio. In fact, scanners with higher magnetic fields (e.g., 7 Tesla) are able to detect this undershoot or initial dip (Jäncke, 2005). This initial dip in turn seems to represent neural firing better than the “classical” BOLD signal. Nevertheless, this initial dip is part of the whole BOLD effect and therefore the BOLD signal is useful to measure brain activity (Jäncke, 2005). Applying high magnetic fields (e.g., 7 Tesla) to humans appears to still be critical concerning duration, sensations of vertigo, and the like (cp., Theysohn, Maderwald, Kraff, Moenninghoff, Ladd, & Ladd, 2008).

Apart from the latest findings "[…] the exact relationship between the measured fMRI signal [i.e., the 'BOLD' contrast] and the underlying neural activity is [still] unclear." (Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001, p. 150). According to Logothetis and colleagues for example, it reflects the "[…] input and intracortical processing of a given area rather than its spiking output." (Logothetis et al., 2001, p. 150). That is, the obtained fMRI signal reflects rather synaptic activity than action potentials (i.e., spikes). Others then showed that an increment of rCBF can be independent from spike activity, but always correlates with local field potentials (Thomsen, Offenhauser, & Lauritzen, 2004). Local field potentials are "[…] electrical fields recorded from microelectrodes in the brain thought to reflect the weighted average of input signals on the dendrites and cell bodies of neurons in the vicinity of the electrode." (Raichle & Mintun, 2006, p. 452). Viswanathan and Freeman (2007) then provided evidences that also the oxygen concentration can be independent from spiking activity, but again is correlated to LFPs. Actually, there is evidence that the brain activity measured by fMRI reflects rather synaptic input processes to neurons than cellular firing (Raichle & Mintun, 2006). Nevertheless, this seems to not prohibit applying this technique in measuring brain activity and inferring the associated cognitive processes from these activations since other researchers showed that LFPs and spiking activity could be used interchangeably to predict the measured imaging signals reliably (Mukamel, Gelbard, Arieli, Hasson, Fried, & Malach, 2005).

As was mentioned above, applying fMRI in research on cognitive processes requires the presentation of several stimuli of the same category to improve the signal-to-noise ratio. More importantly, due to the permanent background activity in the whole brain, fMRI data analyses not only deals with averaging images of one participant, but also calculates contrasts between a condition and the average of this background activity. Moreover, research also aims at exploring the specific functional areas executing specific cognitive processes. Thus, the subtraction method is also used to identify functionally specific brain areas by calculating regionally specific activity differences associated with the execution of different tasks.

Experimental designs for an fMRI experiment can be block-designs or event-related designs. Block-designs are used to present several stimuli of one category in succession. This often leads to an increased BOLD signal, since the same brain regions require more and more energy (Jäncke, 2005). However, effects of habituation or fatigue of the participants might then occur, and the BOLD signal might not increase, instead, it may reach a steady state or even decrease (be reminded that fMRI measures oxygen-deoxygen-hemoglobin differences, thus the BOLD signal changes with the time course of the blood flow). Furthermore, the single events cannot be analyzed in separation. In contrast, an fMRI experiment can also apply an event-related design resulting in a series of images that can be analyzed time-locked to specific single events. Most often, effects of habituation, fatigue, or preparedness can be excluded by implementing this procedure. Since most experimental studies are conducted to draw inferences about effects in populations, fMRI data are also analyzed averaging over several persons. This further improves the signal-to-noise ratio.

The current section described the most important technical aspects of one brain imaging method (i.e., fMRI) since this technique was also applied in the last experiment reported in the following. The advantage and reason for implementing such an experimental procedure depends on the fact that fMRI allows to measure brain processes and associate these brain processes to cognitive processes presumably involved in the execution of a specific task. Since numerous brain imaging studies in the domains of theoretical and practical reasoning are available, there are diverse evidences which brain areas are associated with specific cognitive processes involved in theoretical and/or practical reasoning. Thus, using fMRI seemed an appropriate way to explore the brain areas involved in the execution of the experimental task applied here in relation to the known findings, and subsequently, to infer the associated cognitive processes. This would provide further evidences for the findings and assumptions based on the preceding behavioral experiments within this thesis. Now, the findings referred so far will be integrated, and the leading research ideas as well as the applied research paradigm for the following experiments will be presented. Also, a few term clarifications will be proposed. The concrete hypotheses for each single experiment could be found in the respective experimental section.

1.12 Integration part and derivation of experimental paradigm

The preceding introduction reported diverse findings of theoretical and practical reasoning research based on behavioral and brain imaging experiments. However, as mentioned already, these studies were either on theoretical reasoning or on practical reasoning, but few investigated these reasoning domains together. An exception, at least partly, is the brain imaging study by Fiddick et al. (2005) reported above. Although theoretical and practical reasoning were mainly investigated in isolation, the referred studies and findings seem to provide some overlap and similarities between these two reasoning domains. Numerous studies of both domains yielded evidences that reasoning, both theoretical and practical, is based on rational *and* emotional/intuitive processes. The ‘dual-process’ models seem to be currently the most promising approaches in accounting for the reasoning processes involved in both reasoning domains. Theoretical and practical reasoning furthermore, can be influenced by and depend on working memory capacity and intelligence level as well as prior knowledge/beliefs. The content and the emotional load of a reasoning problem as well as the context in which it is embedded and therefore the relationship to the kind and level of prior knowledge seem to be critical for humans' reasoning performance. Interestingly, nearly all of these influencing and mediating factors can facilitate or impair reasoning processes. Conflicts in reasoning especially appear if the task at hand contains incongruities due to its structure or offers two or more possible decisions which contradict each other, especially, when participants are forced to come to an immediate decision. A 'classical' example of such a conflict is the Stroop task (Stroop, 1935), where over practiced/learned knowledge (i.e., reading, comprehending, and naming of color words, thus language processing/language skills) conflicts with the task demand of explicitly naming the ink color of the presented color words. Thus, automatic, perhaps unconscious, fast reasoning processes conflict with deliberatively more demanding, slow, and conscious ones. This is one of the oldest examples of an incongruence embedded in a task, resulting in heightened decision times and/or error rates. This incongruity refers to conflicting reasoning processes caused by the task demands. In deductive reasoning, similar effects might occur if the content of the problem conflicts with its logical form, representing an interaction of content and logic. In fact, Wilkins (1928) showed that persons' reasoning abilities are affected due to an interaction of the content of a reasoning task and its logical structure/form. A more current example for such an interaction has been reported above (i.e., 'belief bias', Evans et al., 1983). Other examples outside the theoretical reasoning research domain are dilemmas, where either decision leads to a negative outcome (e.g., Greene et al., 2001). However, only the contents related to the possible decisions are in conflict, and not content and logical structure of the task.

There are many more examples, but most of them seem to coerce persons to stop or to override upcoming intuitions (i.e., prior knowledge, beliefs, attitudes) and/or emotions, starting to reason more extensively, thus deliberative and rational about the actual problem. One might say such problems require an individual to switch from the low-road to the high-road (cp., LeDoux, 2006), or to monitor and to guide upcoming 'somatic markers' in the 'right' reasoning direction (cp., Damasio, 2006), or else, to control 'hot thoughts/informed intuitions' and turn them into 'cold reasoning processes' (cp., Thagard, 2006). Whatever explanation and model is chosen, errors or increasing decision times occurring in the reasoning process solving problems/dilemmas/tasks like those mentioned above seem to denote that diverse (sub)-processes are involved in human reasoning. Therefore, 'dual-process' accounts or even 'threefold-process' models appear to explain human reasoning best. It seems to be more important that rational and emotional/intuitive processes are involved in human reasoning, but the proportions of these processes involved in an actual problem are varied and weighted by the reasoner (e.g., WM capacity, prior knowledge, intelligence) as well as by the reasoning problem at hand (e.g., problem complexity, emotional load of the problem, problem content). Additionally, the brain imaging results of diverse studies of both reasoning domains yielded evidences that there are specialized brain areas for specific cognitive processes. Reasoning per se however, seems to activate almost always a kind of "basic" brain network independent from the actual task, which has been shown in the application of conjunction analysis.

Taken all these findings on human reasoning, both theoretical and practical, into account leads to several assumptions about human reasoning:

- Theoretical and practical reasoning are partly based on a common neural network in the brain.
- Theoretical and practical reasoning are similarly based on rational *and* emotional/intuitive cognitive processes.
- The brain areas and associated cognitive processes involved in a specific reasoning problem/task vary with the specific cognitive abilities of the reasoner (e.g., working memory capacity, intelligence level, prior knowledge/beliefs/attitudes) and the specific problem characteristics (e.g., problem complexity, emotional load of the problem, problem structure, decision alternatives). Thus, the contributions of specific brain areas and associated cognitive processes involved in solving a reasoning problem vary. However, a kind of 'basic' or common neural network, and seemingly similar associated processes are always involved. This 'basic' network and its associated cognitive processes are similar for theoretical and practical reasoning, whereby the involved processes represent rational and emotional/intuitive proportions.
- Theoretical and practical reasoning should and can be investigated together to provide evidence for these assumptions.

A few term clarifications seem necessary for the following work in order to avoid misunderstandings. 'Intuitive' means processes based on (over) learned prior knowledge, thus automatic and fast processes. 'Emotional' means emotional load/evaluations of these 'intuitive' processes, presumably activating emotional feelings again, when a specific 'emotionally' laden prior knowledge is retrieved. Therefore, a very sketchy connection is assumed between the terms ‘somatic markers’ (Damasio, 2006) and ‘emotional/intuitive’, representing almost the same. ‘Rational’ means processes related to the broad term of ‘pure’ reason, thus slow and deliberative processes. There is no assumption about conscious or unconscious processes or whether the assumed ‘somatic markers’ really activate somatic bodily states, since this is far beyond the scope of the current work. Furthermore, the suggestion of ‘emotional/intuitive’ and ‘rational’ processes denotes both as ‘cognitive processes’. Finally, both processes are derived from the proposal of ‘dual-process’ accounts explaining human reasoning by Evans (2008). In accordance with Evans, ‘emotional/intuitive’ processes refer to ‘System 1/Type 1’ processes, and ‘rational’ processes to ‘System 2/Type 2’ processes. The same constraints, Evans (2008) proposes himself, account for these assumptions since different terms and ‘dual-process’ accounts are available, although it seems that they share a common basis.

To investigate these assumptions it seemed necessary to create stimulus material combining theoretical and practical reasoning. A sub-goal of the following experiments was to avoid the weaknesses of the practical reasoning items, such as dilemmas, as mentioned above. To fulfill the aspect of a task concerning theoretical reasoning, deductive inference problems (in the following often short: ‘problems’) were chosen. To obtain practical proportions within these reasoning problems, they were partly filled with moral-related (i.e., moral, unmoral) content. This moral-related content was derived from the articles of the ‘Universal Declaration of Human Rights’ which seemed to be the only existing external criterion for morals available (see also methods).

The reasoning problems contained moral, unmoral, abstract, and neutral content for Pre-Study I and Experiment I. Pre-Study II and Experiments II + V applied additional positive and negative emotional control problems. Pre-Study III as well as the Experiments III + IV dealt with moral and unmoral reasoning problems and newly developed neutral content problems as well as an additional recognition task. Recognition items contained also moral, unmoral, and neutral content-laden sentences. The second sentence however, was rearranged so that drawing a deductive inference was made impossible. Participants’ task in all experiments was to judge the validity of the given conclusion of the problems as valid or invalid.[9] Concerning the recognition items in Experiments III + IV, participants had to judge whether the third sentence matched a part of the first one literally or not. The experiments contained the within-subject factors content and validity, whereby Experiments III + IV had

[9] Note that for the current purposes of the following experiments a valid deductive inference was one, whose conclusion is true in every case in which all its premises are true (cp., Jeffrey, 1981).

one extra task (within-subject factors content and ‘matching’) and three or two groups, respectively, encompassing the between-subject factor group and representing a mixed factorial design (task by group). Dependent variables were error rates and decision times for all experiments (i.e., reasoning performance), but Experiment III and IV also measured participants’ moral judgment competence level by applying an additional test (i.e., MJT by Lind, 1978, 2008, see corresponding methods sections) as well as participants’ recognition performance. Experiment V also recorded participants’ brain activity while solving the reasoning problems. Participants, different groups, or treatments of different groups represented the independent variables.

All moral-related content (i.e., moral, unmoral) problems were derived from the articles of the ‘Declaration of Human Rights’. Moral problems were according to these articles, but unmoral ones contradicted them. Abstract problems represented ‘pure’ deductive problems according to formal logic, as reported above. Neutral reasoning problems contained statements about daily tasks (e.g., visit a coiffeur) and emotional ones about highly emotional arousing situations (e.g., friendships, winning the jackpot, loneliness). The latter problems served only as controls. All of these problems were deductive inference problems borrowed from the Modus Ponens version of ‘classical’ conditional reasoning (see above). Since participants should judge the conclusion according to its validity, they were asked to reason deductively, independent of any content. However, because half of the moral-related problems were constructed with a kind of incongruity, participants should also show some reasoning and answering according to the moral-related content indicated by an incongruence effect.

The incongruity of the moral-related problems is related to the moral-invalid and unmoral-valid reasoning problems (congruent problems were moral-valid and unmoral-invalid) where validity status and morality status of the conclusion contradicted each other. For example, the conclusion “The person must be killed.” contradicts common moral beliefs (and the ‘Human Rights Articles’), but would follow necessarily from the premises if this conclusion is valid. This incongruence of the moral-related problems was assumed to be related to prior knowledge of morals. Knowledge of morals seems to be learned, especially concerning the contents of the human rights articles, and might be represented as beliefs or attitudes about moral issues. Further, it could be suggested that morals are learned through parents, school, religion, and the like, and represents highly emotional laden content due to its relation to obedience, guilt, and punishment. Thus, morals as emotionally laden prior knowledge could also reflect ‘somatic markers’ or strong personal attitudes. The incongruence should therefore lead to a reasoning process about morals. If participants would answer in the incongruent conditions according to morals, and ignore the deductive task demands of the reasoning problems, this would indicate that they did practical reasoning instead of ‘pure’ theoretical reasoning. This would be reflected in errors on these incongruent reasoning problems. In turn, if participants would answer correctly in the incongruent cases, this might serve as a hint for suppressed upcoming emotional/intuitive processes due to rational ones, perhaps mediated via working memory processes.

Concerning the neutral reasoning problems, a potential content effect was expected. If this neutral content improves reasoning performance compared to the abstract problems, this could also represent influences of prior knowledge since the neutral problems contained contents about daily and familiar situations.

The abstract problems in turn were assumed to represent ‘pure’ deductive, formal logical reasoning processes, and therefore ‘pure’ rational ones. There seems no possibility of emotional or intuitive influences, since they encompassed no content at all (although this assumption could be questioned; cp., Manktelow, 2004). If participants would have great difficulties with these problems, this could indicate that they are not able to reason in a ‘pure’ rational way, according to logic.

These were the basic ideas behind the experimental paradigm applied in the following experiments. For further details concerning the applied reasoning problems and recognition items as well as specific settings, groups, and measuring techniques, please see the particular introduction and methods section of each experiment. The same accounts for the concrete hypotheses for each single experiment. In the following chapter, the development and evaluation of the reasoning problems, and thus the first pre-study, will be reported.

2. Pre-Study I

The first pre-study was conducted to evaluate the moral, unmoral, abstract, and neutral deductive reasoning problems concerning their intended applicability. It was of interest as to how people perceive these problems according to their inherent logical structure (all problems were deductive inference problems), according to their content (differences between moral-related content, neutral content, and "no content" problems), and according to their emotional load.

2.1 Method

2.1.1 Participants

A total of 20 participants, 3 males with a mean age of 24.67 years (S.D. = standard deviation = ±1.53) and 17 females with a mean age of 24.41 years (S.D. ±6.15), participated in the pre-study. Participants came from local universities and gave written informed consent.

2.1.2 Material

The stimulus material for all following experiments was based on logical reasoning problems: deductive inference problems borrowed from the Modus Ponens version of 'classical' conditional reasoning problems. These deductive inference problems consisted of three sentences (major premise, minor premise, and conclusion). The first premise, the conditional ("If A, then B."), is followed by a second premise, the categorical ("A."), and finally a conclusion ("B."). This conclusion follows necessarily from the two premises, irrelevant of content. To obtain a variation in the problems and create a task that could be executed by the participants, all problems were presented once with a valid ("B.") and in the other instance with an invalid ("Not B.") conclusion. Thus, half of the problems contained a conclusion that followed necessarily from the premises whereas in the other half of the cases the conclusion did not follow from the premises.

Since letters or numbers were assumed to be 'the content-free version' of such a deductive inference problem, they seem to represent 'pure' deductive reasoning best (i.e., theoretical reasoning according to logic/rationality, see also introduction on logic above). Therefore, the version with varying letters was chosen to make up the 'pure' deductive (i.e., abstract) problem condition (original German abstract problems of Pre-Study I and Experiments I + II + V could be found in the Appendix A, p. 168).

The problems concerning practical reasoning (i.e., moral reasoning) were derived from the 'Universal Declaration of Human Rights', adopted and proclaimed by the General

Assembly of United Nations in 1948 (http://un.org/Overview/rights.html, 09/27/2008, 10.28 a.m.). It was chosen since it seems that there is no generally accepted basis for morality thus its articles were assumed to represent the only existing external criterion for an (almost) worldwide shared moral standard. The deductive inference problems for the morality condition were filled with moral (according to common moral beliefs) or unmoral (contradicting common moral beliefs) content. Thus, moral-related contents of the major premises were either in accordance with the articles of the 'Declaration of Human Rights' (e.g., "If a person is a prisoner of war, he must not be enslaved.") or contradicting these articles (e.g., "If a person is a prisoner of war, he must be enslaved."). The moral status of a problem was determined by the conclusion which was either moral or unmoral (e.g., "The person must not be enslaved" vs. "The person must be enslaved."). This last sentence was also varied by validity. The deductive inferences therefore, resulted in moral-valid, moral-invalid, unmoral-valid, and unmoral-invalid problems.[10]

The neutral content problems served as content-filled, theoretical reasoning related controls and contained statements about daily tasks (e.g., visit a coiffeur), again with a valid or invalid conclusion (original German neutral problems applied in Pre-Study I and Experiments I + II + V could be found in Appendix A, p.168).

Four deductive inference problems per condition were presented. Since each problem was presented twice, once with a valid conclusion and in the other instance with an invalid conclusion, a total of 32 reasoning problems resulted, representing a 4 (content) × 2 (validity) within-subject factorial design (Table 8).

Table 8: Design of Pre-Study I and Experiment I.

		content factor			
		moral	unmoral	abstract	neutral
validity factor	valid	4 problems	4 problems	4 problems	4 problems
	invalid	4 problems	4 problems	4 problems	4 problems

For the pre-study, Questionnaires were constructed with one problem per page (all three sentences). Participants were then asked how much this proposition is in their opinion related to logical reasoning, morals, and emotions, respectively (see Appendix C for an

[10] Note that problems were presented in German language, hence, were according to the German version of the 'Declaration of Human Rights'. Original German moral-related problems for all Experiments and Pre-Studies could be found in the Appendix A (p. 167).

example, p. 173). To avoid order-effects problems were randomly arranged. Participants could answer on a 4-point scale ranging from 1 (nothing) up to 4 (very much) points.

2.1.3 Procedure

Participants received a short instruction with an exemplary problem on page one. Then they were requested to fill out the questionnaire without a time limit. The questionnaire also asked for demographic data.

2.2 Results

Analyses were done with SPSS© 17.0 (SPSS Inc., Chicago, Illinois, USA 1989 - 2009) and significance level for all analyses was set to $p = .05$. Non-parametric Friedman and Wilcoxon signed rank tests were used for the analysis of all pre-studies.

Calculating the effects of the ratings for all problems differentiated by content and validity yielded significant effects in the logic dimension (χ^2(df = degrees of freedom = 7) = 40.281, $p < .001$), the morality dimension ($\chi^2(7) = 113.564$, $p < .001$), and the emotionality dimension ($\chi^2(7) = 92.381$, $p < .001$). The single comparisons according to Wilcoxon signed rank tests are reported in Appendix D (p. 174).[11]

Regarding only the effects according to the content without validity revealed significant differences in the logic dimension ($\chi^2(3) = 15.278$, $p = .002$), the morality dimension ($\chi^2(3) = 50.455$, $p < .001$), and the emotionality dimension ($\chi^2(3) = 40.901$, $p < .001$). The Wilcoxon signed rank tests (Figure 4) showed significant differences in the logic dimension for the unmoral problems as compared to all others (vs. moral: $z = -2.829$, $p = .005$; vs. neutral: $z = -2.399$, $p = .016$; vs. abstract: $z = -3.183$, $p = .001$), and for the abstract problems compared to the neutral ones ($z = -2.462$, $p = .014$). The other contrasts were insignificant (moral vs. neutral: $z = -.143$, $p = .886$; moral vs. abstract: $z = -1.677$, $p = .094$). In the morality dimension, the moral-related problems (moral and unmoral) yielded higher values than the neutral and abstract ones (moral vs. neutral: $z = -3.828$, $p < .001$; moral vs. abstract: $z = -3.828$, $p < .001$; unmoral vs. neutral: $z = -3.728$, $p < .001$; unmoral vs. abstract: $z = -3.730$, $p < .001$). The moral-related problems did not differ (moral vs. unmoral: $z = -1.747$, $p = .081$), and so did the abstract and neutral ones ($z = -1.000$, $p = .317$). Additionally, the moral-related problems were rated higher in the emotionality dimension than all other conditions (moral vs. neutral: $z = -3.578$, $p < .001$; moral vs. abstract: $z = -3.520$, $p < .001$;

[11] Note that for all experiments, particularly the pre-studies, many contrasts were calculated due to the explorative character of this work. These contrasts however, might not have been orthogonal.

unmoral vs. neutral: z = -3.578, p < .001; unmoral vs. abstract: z = -3.623, p < .001). Finally, the neutral problems differed in the emotionality dimension from the abstract ones (z = -3.189, p = .001), whereas the moral and unmoral ones showed no differences (z = -.106, p = .916).

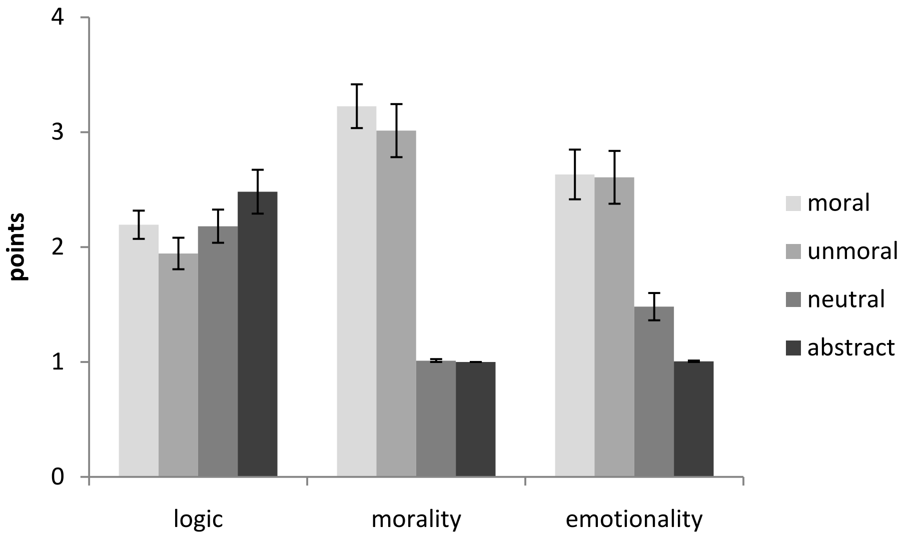


Figure 4: Mean ratings and standard errors for moral, unmoral, neutral, and abstract problems in the dimensions logic, morality, and emotionality.

2.3 Discussion

The results of the pre-study seemed to confirm the applicability of the reasoning problems. They were rated almost equally in the logic dimension, even though the invalid problems received lower ratings. This reflects the fact that participants did not only consider a problem as based on a logical structure, but also tried to draw an inference. With respect to the invalid problems making a valid inference is impossible, so that these ratings are appropriate. In the morality dimension, only the moral and unmoral problems were rated as containing moral-related content, thus differing from the neutral and the abstract problems. Finally, the moral-related problems appeared to be perceived as more emotionally laden than the other problem categories, providing first evidence for a 'learned morality' representing emotionally laden prior knowledge. In the following, the first experiment is reported, testing the stimulus material to a new group of participants.

3. Experiment I

The first experiment used the deductive reasoning problems evaluated in the first pre-study, to obtain answers to questions like: “Could participants deal with ‘pure’ deductive reasoning tasks?”; “Are participants influenced by the moral-related content, and do they show the intended incongruence effect, thus practical reasoning?”. Therefore, the concrete hypotheses are:

- Participants are faster and less prone to errors when dealing with the abstract reasoning problems as compared to all others, indicating therefore that they are able to reason ‘pure’ rationally.
- Participants are slower and show increased error rates when handling the moral-related reasoning problems as compared to all others, indicating the use of practical reasoning processes within a theoretical reasoning task.
- Participants produce higher decision times and error rates when given the incongruent moral-related reasoning problems as compared to the congruent moral-related ones, indicating therefore an incongruence effect.

3.1 Method

3.1.1 Participants

In this first experiment, 21 participants were tested. The sample consisted of 10 male participants with a mean age of 23.30 years (S.D. ±3.53) and 11 female participants with a mean age of 21.82 years (S.D. ±0.98). All were native German speakers, naïve with respect to deductive reasoning tasks, and right-handed according to Salmaso and Longoni’s Edinburgh Handedness Inventory (EHI, 1985, modified version of Oldfield, 1971). Participants came from local universities and gave informed written consent according to the Declaration of Helsinki (1964/2008). Participants were financially compensated or received course credits for an approximate of 30 minutes experimental session.

3.1.2 Material

The stimulus material for this first experiment consisted of the above described deductive inference problems with valid and invalid conclusion and moral, unmoral, neutral, or abstract content, thus representing a 4 (content) by 2 (validity) factorial within-subject design. The decision times and error rates (dependent variables) measured and recorded during the computer experiment represented the reasoning performance of the participants (independent variable).

Additionally, the "Freiburger personality inventory" (FPI-R.; Fahrenberg, Hampel, & Selg, 2001) was applied to control for influences or dysfunctions of personality traits.

3.1.3 Procedure

The 32 problems (4 problems per category) were randomly presented on a standard Dell personal computer (©Dell Inc., 1999–2009), using Super-Lab 4.0 software (©Cedrus Corporation, 1991–2006). Reasoning problems were shown twice, once with a valid conclusion and in the other instance with an invalid conclusion. Sentences were presented successively with self-paced reading (see Figure 5). The third sentence/conclusion (in red) had to be judged as valid or invalid according to the two premises/sentences before.

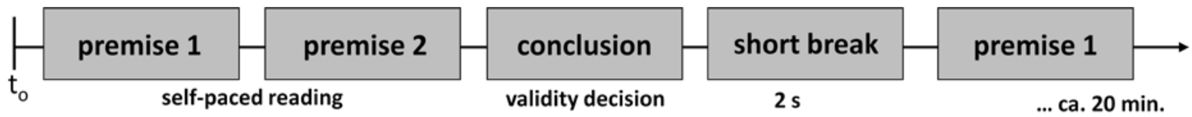


Figure 5: Procedure of the computer experiment.

Validity decisions were counted using a response-box (©Cedrus Corporation, 1991–2006), whereby response buttons were counterbalanced with hand preference between participants. Between each problem, a fixation cross was presented for 2 seconds. After completing the computer experiment, participants filled out the paper and pencil FPI-R questionnaire.

3.2 Results

Analyses were done with SPSS© 17.0 (SPSS Inc., Chicago, Illinois, USA 1989 - 2009) and significance level for all analyses was set to $p = .05$. Analyses include Greenhouse-Geisser corrected F-values of the general linear model for repeated measures, followed by post hoc t-tests. Decision times are solely made up of correct answers. Since the personality inventory showed no anomalies, no further analyses are reported.

3.2.1 Error rates

In total, the moral-related deductive inference problems produced higher error rates than the abstract and neutral ones. On a descriptive level, it could be seen that moral problems revealed 14.88% errors and unmoral ones showed 10.12%. Neutral problems resulted in 1.79% errors and abstract ones in 4.17%.[12]

[12] Descriptive values for this and all following experiments represent mean values, whereby standard errors for error rates and decision times could be found in the respective figures.

Analysis of error rates showed main effects (ME) for validity and content (ME validity: F(1, 20) = 9.474, p = .006; ME content: F(2.385, 47.696) = 6.278, p = .002) with an additional interaction effect (IE: F(1.310, 26.206) = 8.946, p = .003).

The analysis of the content main effect revealed that moral problems produced more errors than neutral (t(20) = 3.532, p = .002) and abstract ones (t(20) = 2.631, p = .016), but did not differ from the unmoral problems (t(20) = 1.793, p = .088). The unmoral problems yielded higher error rates than the neutral ones (t(20) = 3.005, p = .007), whereas the contrast with the abstract condition was insignificant (t(20) = 1.693, p = .106). Comparing the abstract and neutral conditions showed no differences (t(20) = -.777, p = .446). See Figure 6 below.

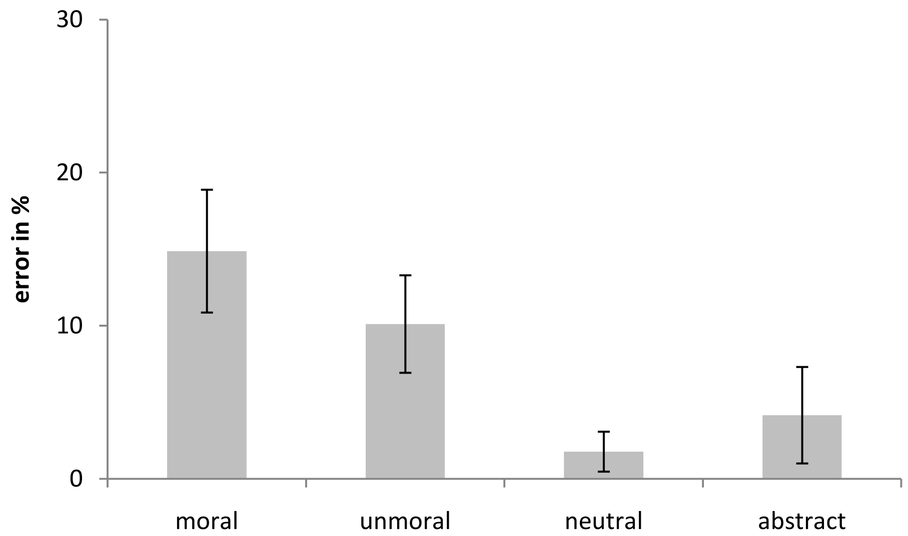


Figure 6: Mean error rates in percent (%) with standard errors for moral, unmoral, neutral and abstract problems.

Analyzing the validity main effect indicated that invalid problems were more prone to errors than valid ones (t(20) = -3.078, p = .006). Results of the post hoc t-tests for the significant interaction effect are reported in Appendix E (p. 175).

Finally, only the moral-related inference problems (moral and unmoral) concerning their inherent congruence-incongruence dichotomy, where logical validity and moral status of the conclusion conflict with each other, were analyzed with an additional General Linear Model. This GLM for repeated measures revealed that incongruent problems (moral-invalid/unmoral-valid) were more prone to errors than congruent (moral-valid/unmoral-invalid) ones (F(1, 20) = 9.931, p = .005; t(20) = -3.151, p = .005; Figure 7).

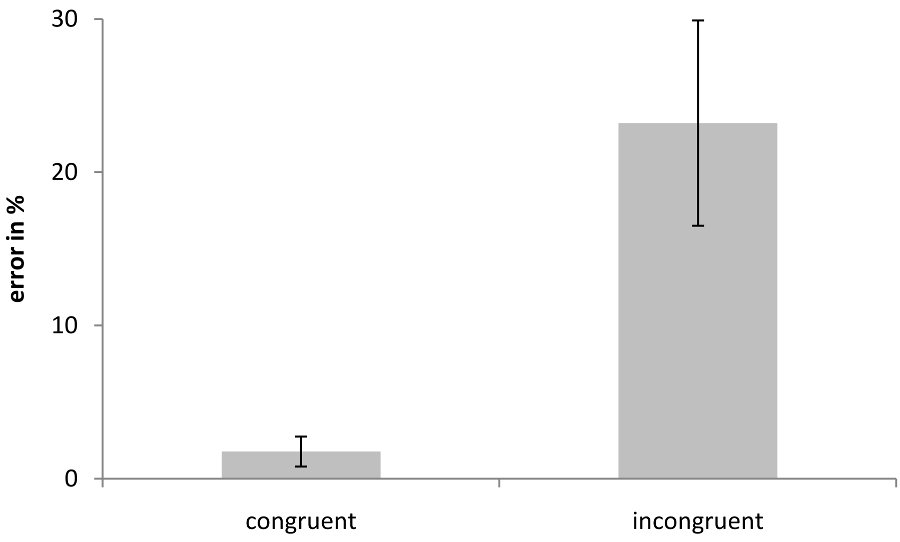


Figure 7: Mean error rates in percent (%) with standard errors for congruent (moral-valid + unmoral-invalid) and incongruent (moral-invalid + unmoral-valid) problems.

3.2.2 Decision times

Decision times, as error rates, showed higher values, namely longer decision times, for the moral-related inference problems (moral: 3162.84 ms, unmoral: 3868.40 ms) compared to neutral and abstract ones (neutral: 2310.25 ms, abstract: 1703.02 ms).

The analysis of the decision times yielded similar results for validity (ME: $F(1,20) = 4.635$, $p = .044$) and content (ME: $F(1.841, 36.813) = 25.329$, $p < .001$), but no interaction effect (IE: $F(2.299, 45.984) = 2.421$, $p = .093$).

Post hoc t-tests for validity revealed significantly faster decisions in the valid compared to the invalid problems ($t(20) = -2.153$, $p = .044$).

The content main effect analysis showed that unmoral problems took longest as compared to moral ($t(20) = -2.467$, $p = .023$), neutral ($t(20) = 5.038$, $p < .001$), and abstract ones ($t(20) = 5.930$, $p < .001$), followed by moral problems (moral vs. neutral: $t(20) = 3.863$, $p = .001$; moral vs. abstract: $t(20) = 6.785$, $p < .001$). Finally, the abstract condition was answered faster than the neutral ones ($t(20) = 4.072$, $p = .001$). See Figure 8 below.

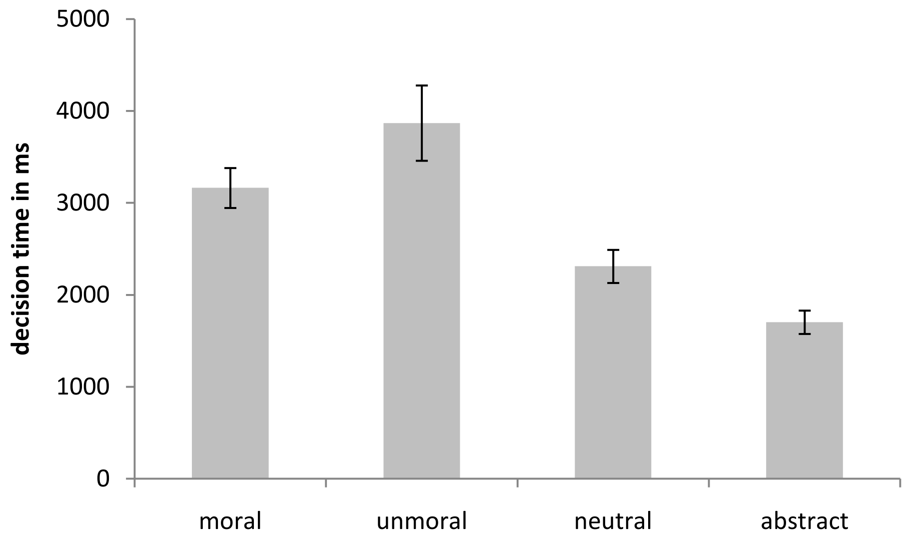


Figure 8: Mean decision times in milliseconds (ms) with standard errors for moral, unmoral, neutral and abstract problems.

3.3 Discussion

Experiment I showed that deductive inference problems with moral-related content are more prone to errors and require longer decision times than neutral or abstract ones. Participants encountered difficulties with the incongruent condition, showing that the intended incongruence effect when logical form and content of the task interact leads to conflicts in the reasoning process. This incongruity effect might be labeled ‘belief bias’, although the current response pattern does not perfectly match the one found in experiments yielding the ‘classical belief bias’ effects (Evans et al., 1983). As explained in the introduction, ‘belief biases’ occur if a person’s beliefs/prior knowledge related to the problem content conflict or contradict with the logical form of the arguments. According to the original ‘belief bias’ effect, the participants should have shown the highest error rate for the unmoral-invalid problems, since ‘belief bias’ effects are stronger for invalid and unbelievable problems. This was clearly not the case. In contrast, the incongruent conditions of moral-invalid and unmoral-valid problems revealed the highest error rates. Assuming intuitive/emotional reasoning processes getting in conflict with rational ones as responsible for these performance results might indeed be labeled as a kind of (new) ‘belief bias’. In the moral-invalid cases, participants might have thought and answered according to “Yes, this is moral”, although they should have thought and answered according to “No, this is invalid”. Vice versa, in the unmoral-valid condition they might have thought and answered according to “No, this is unmoral”, although they should have thought and answered according to “Yes, this is valid”. Apart from this possible interpretation, the occurring incongruence effect might be interpreted as participants indeed reasoned about and according to the moral-related content, therefore showing practical reasoning.

Another potential explanation of this incongruity effect might refer to moral attitudes (cp., Luo et al., 2006). These moral attitudes, in place of beliefs, could have played the major role in the participants' decisions. As reported above, Luo et al. (2006) criticized former studies on moral reasoning that they required explicit decisions from the participants and were therefore not able to measure the "real" underlying moral attitudes these persons might have. However, the current experiments did not even require from the participants to make a moral decision or to reason about morals. Instead, the embedded incongruence should ensure that automatic processes lead to moral reasoning. These processes might indeed represent moral attitudes. Furthermore, Luo et al. (2006) stated that a moral decision requires more than just upcoming automatic moral attitudes. In turn, these attitudes could also be based on prior knowledge stressing the special emotional load of morals, since also attitudes need to have been built. This means that also attitudes have to be developed via learning and experience. Moreover, the cognitive processes representing moral attitudes might indeed reflect emotional/intuitive processes, therefore emotionally laden prior knowledge represented as fast and automatic processes in moral reasoning and decision making. More importantly, these incongruence effects indicate emotional/intuitive processes involved in practical reasoning, but do not allow inferences to potential influences of rational processes in practical reasoning.

Apart from these issues, the results reveal that participants were unable to ignore the moral-related content and to just focus on validity. The errors participants made in the moral-related cases imply that they reasoned and answered according to the moral-related content. However, they had almost no problems correctly evaluating the conclusions of the moral-valid or unmoral-invalid cases, indicating that moral-related content per se did not perturb reasoning and decision making processes. This seems not astonishing since in the congruent conditions theoretical reasoning (i.e., 'pure' deductive reasoning) and practical reasoning (i.e., moral reasoning) should lead to the same outcome. It remains unclear therefore, whether participants reasoned morally or deductively when faced with congruent problems. Invalid problems resulted in more errors and higher decision times than valid ones indicating that the conflicting moral-invalid problems were affected more than the unmoral-valid ones, an effect that has been reported and described earlier (e.g., Evans et al., 1983).

Concerning the abstract problems, the results indicate that participants are able to deal with 'pure' deductive inferences, hence to reason rational. This assumption is limited however, by the fact that only 'simple' inference problems were chosen (cp., Knauff, 2007).

Shorter decision times of the abstract problems might just reflect the fact that abstract sentence length was much shorter than that of all other problems. Therefore, reading time alone had an effect contributing to the validity judgment.[13] Apart from the decision times and the inherent problem of sentence length of the abstract problems compared to all others, the neutral-valid problems produced the lowest error rates. This indicates a facilitative content effect (e.g., Johnson-Laird et al., 1972). Participants therefore, are able to reasoning according to logic, especially if only 'simple' reasoning problems are presented, but are even better if they could use prior knowledge.

The results lead to a few preliminary conclusions. Persons are able to reason 'pure' rationally, but performance is even better if prior knowledge could be used. Thus, theoretical reasoning processes seem to involve not only rational but also intuitive processes. Moral-related content, when constructed with an incongruity, reveals evidence that people tend to reason according to this content, i.e. to show practical reasoning. This might be taken as a hint that prior knowledge, perhaps emotionally laden, influences theoretical reasoning, and thus, in this special experimental paradigm, results in practical reasoning. Since the congruent moral-related problems were not affected this way, one might further assume that also practical reasoning involves rational processes aside from the emotional/intuitive ones. However, this assumption remains very speculative since there is also a possibility that no practical reasoning occurred in the congruent cases. Nevertheless, the strong content influences of the moral-related reasoning problems yield evidence of a learned morality according to the human rights articles. Another possible explanation might be that only the moral-related problems were highly emotional laden, and thus resulted in these strong content effects due to their emotional load. Whether participants showed practical reasoning related to prior knowledge or whether these content effects just reflected strong moral attitudes seems unclear. Nonetheless, moral attitudes might be based on over learned knowledge, particularly emotionally laden. In total, the preceding experiment and its results seem to support the assumption that the human rights articles represent an adequate means for investigating practical reasoning in combination with theoretical reasoning.

To further explore the potential influences and proportions of emotions/emotional load on and in reasoning, the second experiment implemented additional emotional control problems, which were firstly evaluated in the second pre-study.

[13] Note that abstract problems were necessarily presented in this form to represent 'pure' deductive reasoning. There was no control therefore, of sentence length for the other problems, since this seemed to be not valuable.

4. Pre-Study II

In the second pre-study, the newly developed positive and negative emotional control problems should be evaluated together with the moral-related ones. The new evaluation of the moral-related problems seemed necessary due to this division in positive and negative emotional controls. The reason for this segregation is that it is unknown as to whether morals are associated with positive emotions over negative ones, or vice versa. One might assume for example that morality has been learned via commandments, prohibitions, obedience, punishment, and the like during childhood and is therefore associated with negative emotions. It could also that morality is associated more with positive emotions due to its helpful function in regulating a peaceful and fruitful coexistence of human beings. Hence, the following pre-study was also a first step in exploring these different possible connotations of morals. Presenting moral-related problems together with the new emotional ones also ensured to avoid influences of task length or moral content per se. Abstract and neutral conditions were excluded since they had shown no affinities to morality or emotionality, respectively.

4.1 Method

4.1.1 Participants

In this second pre-study, 10 males with a mean age of 26.1 years (S.D. ± 6.54) and 10 females with a mean age of 21 years (S.D. ± 2.98), not involved in one of the former studies, took part. Participants came from local universities and gave written informed consent.

4.1.2 Material

Equivalent to the deductive inference problems implemented in the first experiment, control problems with emotional contents (almost matched for sentence length with the moral-related and neutral problems) were developed. These emotional problems contained either positive or negative emotional contents. They also consisted of a first premise (e.g., “If a person is successful, then he is lucky.”, or “If a person has no friends, then he is lonely.”), followed by a second premise (e.g., “A person is successful.”, or “A person has no friends.”), and finally a conclusion (e.g., “The person is lucky.”, or “The person is lonely.”). Each problem appeared twice, once with a valid conclusion (see example), and in the other instance with an invalid conclusion (all problems could be found in Appendix A, p. 169). Hence, Pre-Study II and Experiment II represented a 6 (content) by 2 (validity) factorial within-subject design applying a total of 48 reasoning problems (see Table 9).

Table 9: Design of Pre-Study II and Experiment II.

		content factor					
		moral	unmoral	abstract	neutral	positive emotional	negative emotional
validity factor	valid	4 problems	4 problems	4 problems	4 problems	4 problems	4 problems
	invalid	4 problems	4 problems	4 problems	4 problems	4 problems	4 problems

4.1.3 Procedure

The procedure for this second pre-study was equivalent to that of the first pre-study.

4.2 Results

Analyses were equivalent to that used in Pre-Study I as presented earlier.

The Friedman tests for all problem categories separated into content and validity yielded significant differences in the logic dimension ($\chi^2(7) = 36.810$, $p < .001$), the morality dimension ($\chi^2(7) = 69.080$, $p < .001$), and the positive emotionality dimension ($\chi^2(7) = 95.800$, $p < .001$) as well as the negative emotionality one ($\chi^2(7) = 89.502$, $p < .001$). The single comparisons with Wilcoxon signed rank tests are reported in Appendix F (p. 176).

Concerning only the analysis of the ratings according to content revealed significant effects for the logic dimension ($\chi^2(3) = 11.158$, $p = .011$), the morality dimension ($\chi^2(3) = 21.785$, $p < .001$), the positive emotionality ($\chi^2(3) = 44.299$, $p < .001$), and the negative emotionality dimensions ($\chi^2(3) = 43.015$, $p < .001$).

The post hoc Wilcoxon signed rank tests of the logic dimension showed that unmoral problems received lower ratings than moral (z = -2.493, p = .013), positive emotional (z = -2.851, p = .004), and negative emotional ones (z = -2.364, p = .018). The emotional problems differed slightly from each other (z = -2.651, p = .008), whereas the moral ones yielded no differences at all as compared to the positive (z = -.175, p = .861) and the negative emotional problems (z = -1.226, p = .220).

In the morality dimension the moral problems contained the maximum moral content (moral vs. unmoral: z = -2.010, p = .044; moral vs. positive-emotional: z = -3.737, p < .001; moral vs. negative-emotional: z = -3.699, p < .001), followed by the unmoral ones (unmoral vs. positive-emotional: z = -3.044, p = .002; unmoral vs. negative-emotional: z = -2.962, p = .003). The emotional problems revealed no differences in the morality dimension (z = -1.262, p = .297).

In the positive emotionality dimension the positive emotional controls obtained the highest ratings (positive emotional vs. moral: z = -2.529, p = .011; positive emotional vs. unmoral: z = -3.924, p < .001; positive emotional vs. negative-emotional: z = -3.923, p <

.001), followed by the moral problems (moral vs. unmoral: z = -3.849, p < .001; moral vs. negative-emotional: z = -3.329, p = .001), and the negative-emotional ones before the unmoral problems (z = -2.022, p = .043).

Finally, the negative emotionality dimension showed that the unmoral (unmoral vs. moral: z = -3.925, p < .001; unmoral vs. positive-emotional: z = -3.851, p < .001) and negative emotional problems (negative emotional vs. moral: z = -3.932, p < .001; negative emotional vs. positive emotional: z = -3.767, p < .001) were rated as the most negative ones. Positive emotional and moral problems (z = -1.291, p = .197) as well as unmoral and negative emotional ones (z = -1.535, p = .125) revealed no differences. For an overview see Figure 9 below.

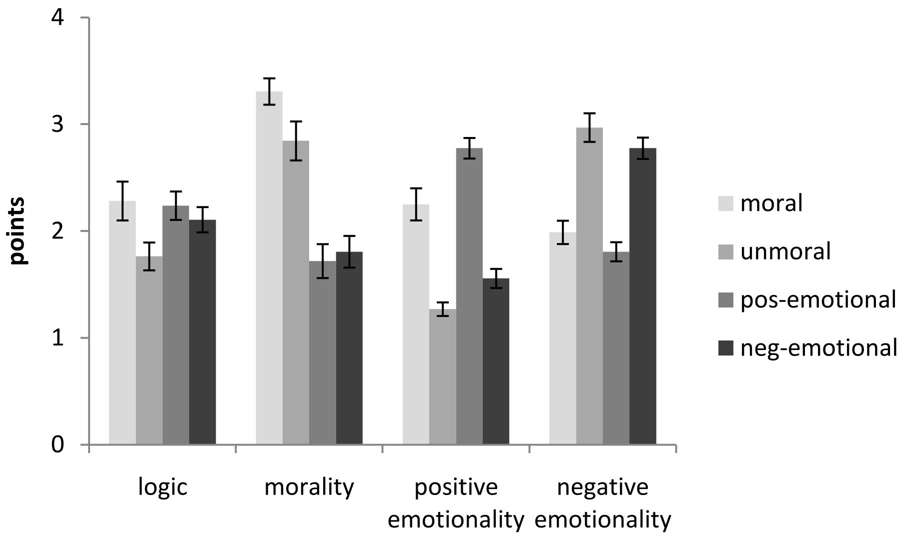


Figure 9: Mean ratings and standard errors for moral, unmoral, positive and negative emotional problems in the dimensions logic, morality, positive and negative emotionality.

4.3 Discussion

The pre-study implied that the problems comprise the intended contents. The logic dimension did not differ between moral-related or emotional reasoning problems, which may seem reasonable since all contents were embedded in deductive inference problems. The moral-related problems were perceived as carrying the most moral content, and positive and negative emotional dimensions were assigned to the corresponding control problems. Therefore, negative emotional problems enabled controlling for negative emotional aspects of moral-related problems, as did the positive emotional ones for positive emotions.

Additionally, participants experienced unmoral problems and negative emotional problems almost equally negative, but interestingly, the moral inference problems were perceived as positively emotional laden. This supports the above made assumptions about morality as a whole. It seems as if moral content refers to the useful functions of morals, and unmoral content refers to the negative effects and potential negative consequences associated with unmoral behavior. Independent of these speculative suggestions, the reasoning problems seemed to contain the anticipated contents and could therefore be used for the following experiments.

5. Experiment II

Since the second pre-study yielded the applicability of the newly developed reasoning problems, the second experiment was carried out to investigate the effects of the emotional controls particularly in contrast to the moral and unmoral problems. Additional hypotheses for this experiment are:

- If the emotional load of morals is much stronger than the knowledge about morals, performance on moral-related and emotional problems should be similar.
- Performance on positive emotional problems is more comparable to that on the moral problems, and performance on the negative emotional problems is more comparable to that on the unmoral problems.

5.1 Method

5.1.1 Participants

In Experiment II 33 newly recruited participants were investigated. The 16 males had a mean age of 23.13 years (S.D. ±3.74) and the 17 females had a mean age of 21.47 years (S.D. ±3.36). All participants were again native German speakers, naïve to deductive reasoning tasks, right-handed according to Salmaso and Longoni’s EHI (1985, modified version of Oldfield, 1971), and showed no previous mental illness or psychotropic medication. Participants came from local universities, and gave informed written consent according to the Declaration of Helsinki (1964/2008). Participants were financially compensated or received course credits for the approximately 30 minutes of testing.

5.1.2 Material

All of the evaluated reasoning problems with moral, unmoral, neutral, abstract, positive and negative emotional content (4 per category) were included. Each problem was shown twice, once with a valid conclusion and in the other instance with an invalid conclusion. This time, a total of 48 reasoning problems was presented, representing a 6 (content) by 2 (validity) factorial within-subject design. As in Experiment I, participants had to deal with the “Freiburger personality inventory” (FPI-R.; Fahrenberg, Hampel, & Selg, 2001) to control for influences or dysfunctions of personality traits.

5.1.3 Procedure

The procedure was identical to that of Experiment I.

5.2 Results

Analyses were equivalent to Experiment I. Since the personality inventory showed no anomalies, no further analyses were conducted.

5.2.1 Error rates

Descriptively, Experiment II yielded increased error rates for moral and unmoral problems as compared to all other conditions. Participants produced 17.80% errors in the moral condition and 13.26% in the unmoral condition, whereas they made 3.03% in the neutral, 5.30% in the abstract, 3.79% in the positive emotional, and 4.92% in the negative emotional condition.

These descriptive values are reflected in significant main effects of content (ME content: $F(2.593, 82.973) = 10.256$, $p < .001$) and validity ($F(1, 32) = 5.408$, $p = .027$) as well as an interaction effect (IE: $F(1.974, 63.161) = 9.578$, $p < .001$). The corresponding post hoc t-tests showed that invalid problems were more prone to errors than valid ones ($t(32) = -2.326$, $p = .027$). Analyzing the content main effect revealed that moral problems produced the highest error rates (moral vs. neutral: $t(32) = 4.695$, $p < .001$; moral vs. abstract: $t(32) = 4.195$, $p < .001$; moral vs. positive emotional: $t(32) = 4.645$, $p < .001$; moral vs. negative-emotional: $t(32) = 4.690$, $p < .001$), whereby the comparison to the unmoral problems did not reach significance ($t(32) = 1.712$, $p = .097$). The unmoral condition also revealed higher error rates than all other conditions (unmoral vs. neutral: $t(32) = 3.032$, $p = .005$; unmoral vs. positive emotional: $t(32) = 2.861$, $p = .007$; unmoral vs. negative emotional: $t(32) = 2.432$, $p = .021$; unmoral vs. abstract: $t(32) = 2.235$, $p = .033$). The abstract problems did not differ from the positive emotional ($t(32) = .780$, $p = .441$), the negative emotional ($t(32) = .215$, $p = .831$), and the neutral control conditions ($t(32) = -1.139$, $p = .263$). Neutral problems showed no differences from both of the emotional inference tasks (neutral vs. positive emotional: $t(32) = -.702$, $p = .488$; neutral vs. negative emotional: $t(32) = -1.153$, $p = .258$), which did not differ from each other too ($t(32) = -.649$, $p = .521$). The post hoc t-tests regarding the interaction effect of content and validity are reported in Appendix G (pp. 177-178). See Figure 10 for illustration of the content main effect.

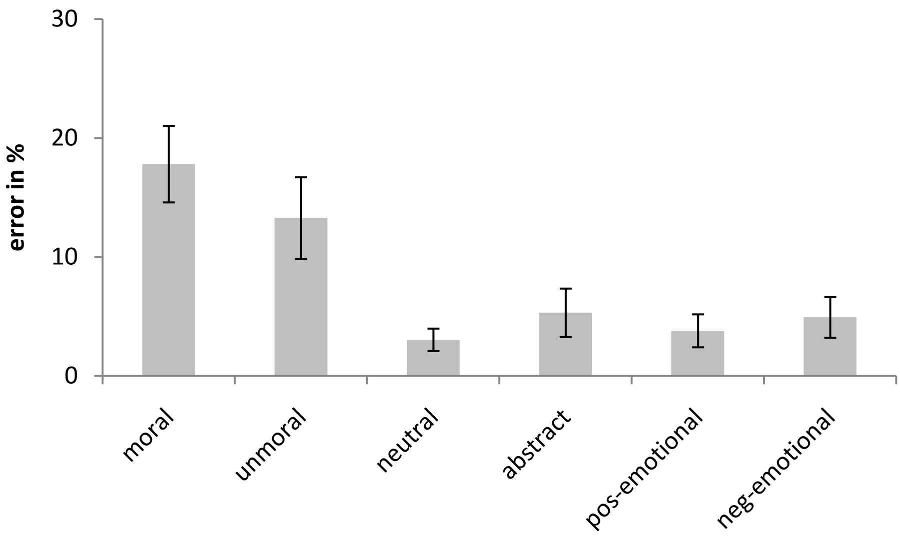


Figure 10: Mean error rates in percent (%) with standard errors for moral, unmoral, neutral, abstract, positive and negative emotional problems.

Finally, only the moral-related inference problems were analyzed again concerning the potential incongruence effect. Again, as in Experiment I, the incongruent problems produced more errors than the congruent ones ($F(1, 32) = 13.361$, $p = .001$; $t(32) = -3.655$, $p = .001$).

5.2.2 Decision times

Participants answered slowest in the moral (3189.29 ms) and unmoral (2831.54 ms) conditions in contrast to the neutral (2058.23 ms) or abstract (1603.11 ms) ones, and also as compared to the newly developed positive (2147.12 ms) and negative (2359.42 ms) emotional controls.

Decision times revealed similar results leading to main effects of content ($F(3.195, 102.225) = 23.974$, $p < .001$) and validity ($F(1, 32) = 34.958$, $p < .001$) as well as an interaction effect ($F(3.079, 98.522) = 7.267$, $p < .001$). Participants needed more time to judge invalid problems compared to valid ones ($t(32) = -5.913$, $p < .001$). The analyses of the content main effect (Figure 11) yielded higher decision times for moral problems compared to abstract ($t(32) = 7.785$, $p < .001$), neutral ($t(32) = 5.389$, $p < .001$), positive ($t(32) = 4.581$, $p < .001$), and negative emotional ones ($t(32) = 3.932$, $p < .001$), however did not differ significantly from the unmoral inferences ($t(32) = 1.789$, $p = .083$). The unmoral problems led to higher decision times than the abstract ($t(32) = 8.513$, $p < .001$), the neutral ($t(32) = 5.659$, $p < .001$), the positive ($t(32) = 4.566$, $p < .001$), and the negative emotional ones ($t(32) = 3.171$, $p = .003$). The abstract condition in turn showed the lowest decision times compared to the neutral ($t(32) = 5.366$, $p < .001$), the positive ($t(32) = -3.335$, $p = .002$) as well as the negative emotional one ($t(32) = -6.410$, $p < .001$). Neutral inferences differed from negative emotional ones (t(32) = -3.593, p = .001), but not from positive emotional problems (t(32) = -.617, p = .541). Finally, the emotional controls revealed no differences (t(32) = -1.378, p = .178). Post hoc t-tests of the interaction effect are shown in Appendix G (pp. 177-178).

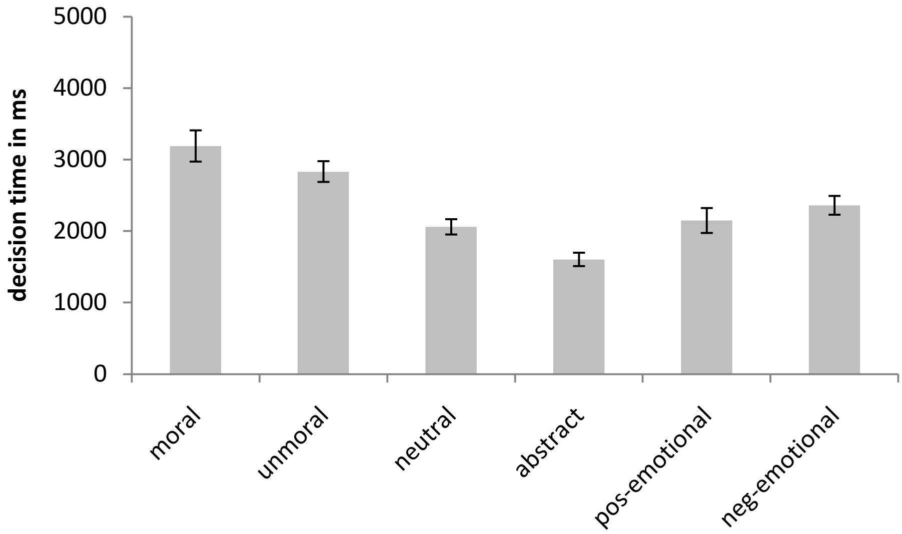


Figure 11: Mean decision times in milliseconds (ms) with standard errors for moral, unmoral, neutral, abstract, positive and negative emotional problems.

Finally, decision times exhibited an incongruence effect (F(1,32) = 9.762, p = .004) with incongruent problems resulting in significantly higher decision times than congruent ones (t(32) = -3.124, p = .004).

5.3 Discussion

Experiment II again showed (see Experiment I) that participants had more difficulties handling the moral-related problems than all of the others. This was reflected in higher error rates and longer decision times for the moral-related problems. Again, the incongruent moral-related inferences played a crucial part in these results, whereby the invalid problems had even higher error rates and longer decision times than the valid ones. The effect that valid deductive problems are easier to solve than invalid ones is well known from former studies (e.g., Evans et al., 1983), and has also been shown in Experiment I. The same accounts for the intended incongruence effect. In contrast, the abstract reasoning problems (i.e., 'pure' deductive) revealed that participants can perform 'pure' deductive reasoning (with the restriction that only the 'simplest' reasoning problems were used, see also discussion on Experiment I above), but the neutral content even facilitates reasoning, however the comparisons were this time insignificant.

Contrasting the moral-related problems with the newly created emotional problems yielded further insights in the potential causes of the content effects of morals. Participants produced about 15 percent errors in the moral-related conditions, whereas they had only 4.5 percent errors in the emotional control conditions. This leads to the assumption that moral-related content is apparently strongly emotionally laden than emotional material itself. In turn, this rather might reflect the influence of prior knowledge and thus support proposals that morality is indeed learned (Bore et al., 2005; Bucciarelli et al., 2008; Churchland, 1998; Goodenough & Prehn, 2004; Hare, 1981; Luo et al., 2006; Moll et al., 2005; Pasupathi & Staudinger, 2001). Strong prior knowledge therefore, that of beliefs or attitudes, might have influenced the reasoning processes of the participants who were apparently unable to ignore the moral-related content and focus only on task demands. This seems to contradict the first hypotheses above. However, taking only the congruent moral-related cases into account reveals a different picture. These problems were solved almost comparably to the emotional ones. Moral-valid problems had comparable error rates to the positive emotional valid ones, and unmoral-invalid problems were similar to the negative emotional invalid ones. Unfortunately, no clear suggestion can be made for the congruent moral-related cases whether participants showed practical reasoning, or whether their theoretical reasoning processes were unaffected. This then limits possible conclusions. Apart from that, the major performance decrements of the moral-related problems can be ascribed to the incongruity effect caused by the incongruent problems. Therefore, the mixture of theoretical and practical reasoning, particularly in conflicting cases, seemed to cause the participants' severities.

Aside from this unresolved issue, moral-valid and unmoral-invalid problems were solved almost comparable to the positive and negative emotional ones. This seems to contradict the conclusion made earlier that morals are more emotionally laden than 'pure' emotional contents. It appears rather plausible that emotional and moral-related contents do not differ much, concerning their impact on deductive reasoning problems. The effects of prior knowledge on reasoning support proposals of intuitive proportions in moral reasoning, but does not explain whether there are strong or weak emotional contributions as well, especially since also the emotional controls were, at least partly, related to prior knowledge. It would have been better therefore to create incongruent emotional reasoning problems and compare them with the incongruent moral-related ones.

Apart of these methodological weaknesses, the results paved way to a few preliminary conclusions. Participants were able to reason rationally, but content facilitated the reasoning process indicating the involvement of intuitive processes in theoretical reasoning. Moral-valid and emotional-valid problems showed performance improvements as compared to the abstract problems, perhaps related to facilitative influences of positive emotional load besides prior knowledge. This shows clear indications for the involvement of emotional and intuitive processes in theoretical and practical reasoning (if participants reasoned according to morals in the moral-valid condition). The same might explain the similar impairing influences on the unmoral-invalid and negative emotional-invalid problems as compared to the abstract-invalid ones, revealing disturbing influences of negative emotional load and prior knowledge. These assumptions however, remain speculative since there is no clear evidence as to whether participants' reasoned morally in the congruent cases or not, particularly since some of these differences did not reach significance.

Another problematic aspect which needs to be addressed might have been the special structure of some moral-related reasoning problems, inherent to the derivation from the human rights articles. This refers to the fact that the last sentence denoted the moral status of a problem and some problems changed their moral status between the first sentence (e.g., formulated as moral) and the last sentence (e.g., formulated as unmoral). This abrupt change in the moral status of an inference problem could have caused some confusion in the reasoning processes of some of the participants. Furthermore, some moral-related and emotional problems might have already produced a content effect in the first sentence. The effect of the first premise refers to the fact that some of the problems had to be constructed to allow their applicability in all possible conditions, therefore avoiding content differences between conditions. The first premise could already contain a conflict between the two partial statements of the 'if' and 'then' part related to prior knowledge/beliefs/attitudes (see Appendix A for examples on the variation of the first sentence of the problem "If a person is a prisoner of war, then he must not be enslaved").

Nevertheless, the results give a first hint that theoretical and practical reasoning could not only be investigated together, but are also based on similar processes, namely emotional/intuitive and rational ones. The evidence for similar processes however, remains vague concerning the experiments conducted up to now.

It would be of interest therefore, to further explore the different processes involved. Another way to test the potential proportions of these processes would be to manipulate one and examine the effects of this manipulation in contrast to a not manipulated one. A well known procedure of manipulation is to train participants in deductive reasoning (for an overview see Klauer & Meiser, 2007). The following experiment implemented a logic training for one group as well as a pseudo training and no training for two other groups, which served as controls.

Since the ‘training’ experiment as well as the following ‘intelligence’ experiment dealt with newly developed neutral problems, the third pre-study will be reported first.

6. Pre-Study III

Experiments I and II used ‘pure’ deductive reasoning problems (i.e., abstract condition) to explore the theoretical reasoning abilities of individuals in comparison to practical reasoning. As could be seen in the results, participants made almost no errors and were very fast in their decisions concerning these theoretical problems. However, a meaningful comparison between decision times of the abstract problems together with the moral-related ones appears to be critical, since sentence length between both conditions extremely differs. In turn, the neutral reasoning problems seemed closely related to the abstract ones, although they also represented some facilitative content effects. This led to the assumption that theoretical reasoning might also be investigated using only these neutral reasoning problems. Hence, for a better comparison on decision times, in the two following experiments, new neutral problems were developed and exactly matched for sentence length between conditions. Furthermore, the emotional controls were excluded since they did not provide so much additional explanatory value.

This pre-study also contained newly developed recognition items as an additional task. The idea to implement a recognition task refers to the fact that also recognition performance depends on two processes such as prior knowledge or emotional load. It only seems rational to obtain further evidences for the involvement of intuitive/emotional processes in practical reasoning due to the performance with these items.

6.1 Method

6.1.1 Participants

In this pre-study, 30 naïve participants took part. The 11 male participants had a mean age of 24.91 years (S.D. ±1.81) and the 19 female participants had a mean age of 24.74 years (S.D. ±4.16). Participants came from local universities and gave informed written consent.

6.1.2 Material

This pre-study was conducted to evaluate the newly developed neutral reasoning problems in comparison to the already evaluated moral-related ones. These neutral problems were now exactly matched in word number/sentence length controlling reading and decision time differences. Since the moral-related problems were further divided into moral and unmoral ones, the neutral problems were adjusted by splitting them into two categories: ‘everyday life’ (neutral1) and ‘leisure time’ (neutral2). This assignment in neutral 1 and 2 was just for analytical reasons and not content related. Thus, the reasoning problems (within-subject factors) for the following two experiments represented a 4 (content) by 2 (validity) factorial design (see Table 10).

Table 10: Design of deductive reasoning problems of Pre-Study III.

		content factor			
		moral	unmoral	neutral1	neutral2
validity factor	valid	4 problems	4 problems	4 problems	4 problems
	invalid	4 problems	4 problems	4 problems	4 problems

The newly developed recognition items should also be evaluated. The additional recognition items were similar to the deductive reasoning problems, except for their second sentence. Their second sentences were rearranged (word replacements) in such a manner where drawing of a logical inference was made impossible. The third sentence of these recognition items literally matched a part of the first sentence ('literal match' condition or short 'match') or not ('no literal match' condition or short 'no match'). These recognition items also contained moral, unmoral, neutral1, and neutral2 contents comparable to the deductive reasoning problems (see all items in German version in Appendix B, pp. 171-172). Thus, these recognition items (within-subject factors) for the following two experiments represented a 4 (content) by 2 (matching) factorial design (see Table 11).

Table 11: Design of recognition items of Pre-Study III.

		content factor			
		moral	unmoral	neutral1	neutral2
matching factor	match	4 recognition items	4 recognition items	4 recognition items	4 recognition items
	no match	4 recognition items	4 recognition items	4 recognition items	4 recognition items

The structure and rating-scale of the questionnaires used was the same as for the second pre-study, but this time containing additional recognition items.

6.1.3 Procedure

The procedure was equivalent to the two former pre-studies.

6.2 Results

6.2.1 Deductive reasoning problems

Analysis of the pre-study corresponded to the ones reported above.

Friedman tests revealed significant main effects for the logic dimension ($\chi^2(7)$ = 118.574, p < .001), the morality dimension ($\chi^2(7)$ = 173.145, p < .001) as well as the positive ($\chi^2(7)$ = 69.102, p < .001) and negative ($\chi^2(7)$ = 146.918, p < .001) emotionality dimensions. Post hoc Wilcoxon signed rank tests for all categories differentiated according to content and validity are reported in Appendix H (p. 179).

Analyzing the ratings according to the content yielded significant effects for the logic dimension ($\chi^2(3)$ = 9.057, p = .029), the morality dimension ($\chi^2(3)$ = 75.111, p < .001), the positive emotionality dimension ($\chi^2(3)$ = 27.938, p < .001), and the negative emotionality dimension ($\chi^2(3)$ = 66.583, p < .001).

Comparing the different content categories in the logic dimension exhibited higher ratings for the moral problems compared to the unmoral (z = -3.217, p = .001) and the neutral2 ones (z = -2.240, p = .025), but not compared to the neutral1 problems (z = -1.277, p = .201). Unmoral problems in turn received lower values compared to the neutral1 (z = -2.464, p = .014) and the neutral2 ones (z = -2.498, p = .012). The neutral problems themselves did not differ in the logic dimension (z = -.863, p = .388).

In the morality dimension, the moral-related problems yielded almost the same ratings (moral vs. unmoral: z = -1.260, p = .208), and so did the neutral ones (neutral1 vs. neutral2: z = -1.312, p = .190). In contrast, the moral problems were rated higher than the neutral1 (z = -4.715, p < .001) and the neutral2 ones (z = -4.712, p < .001), and so were the unmoral problems compared to both neutral conditions (vs. neutral1: z = -4.712, p < .001; vs. neutral2: z = -4.711, p < .001).

In the positive emotionality dimension the moral problems received higher values than the unmoral (z = -3.982, p < .001), the neutral1 (z = -3.675, p < .001), and the neutral2 ones (z = -2.926, p = .003). The unmoral problems were also rated lower than the neutral1 (z = -2.145, p = .032) and neutral2 problems (z = -2.511, p = .012) which did also slightly differ from each other (neutral1 vs. neutral2: z = -2.379, p = .017).

Lastly, in the negative emotionality dimension, the unmoral problems were evaluated more negatively than the moral (z = -4.401, p < .001), the neutral1 (z = -4.545, p < .001), and the neutral2 ones (z = -4.545, p < .001). Also, the moral problems received higher values than the neutral1 (z = -4.310, p < .001) and the neutral2 conditions (z = -4.264, p < .001). Finally, the neutral1 problems differed slightly from the neutral2 ones (z = -2.236, p < .025).

A graphical illustration of the results of Pre-Study III according to the content categories of the deductive reasoning problems is shown in Figure 12 below.

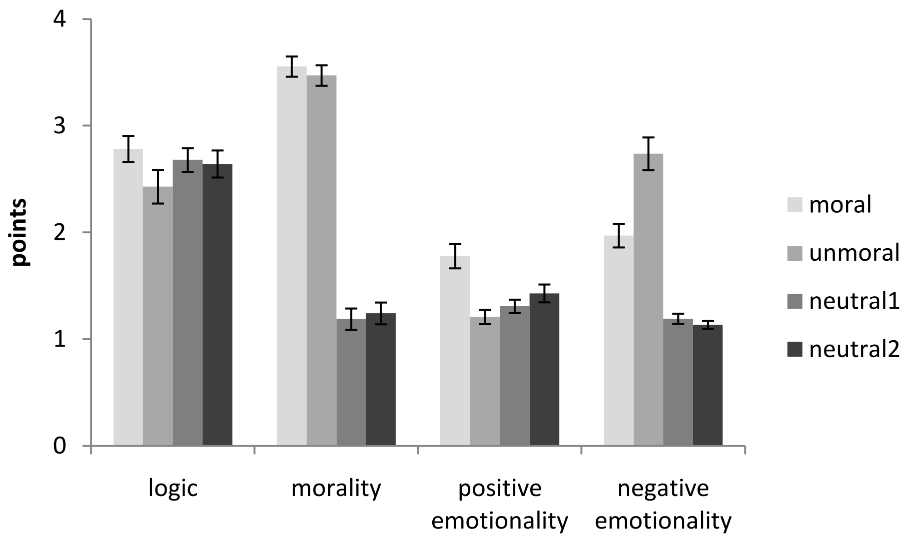


Figure 12: Mean ratings (points 1 to 4) and standard errors for moral, unmoral, neutral1, and neutral2 problems in the dimensions logic, morality, positive and negative emotionality.

6.2.2 Recognition items

The Friedman tests for the ratings of the recognition items yielded significant main effects in the logic dimension ($\chi^2(7) = 29.756$, $p < .001$), the morality dimension ($\chi^2(7) = 162.369$, $p < .001$) as well as the positive ($\chi^2(7) = 34.683$, $p < .001$) and negative ($\chi^2(7) = 147.549$, $p < .001$) emotionality dimensions. Post hoc Wilcoxon signed rank tests for all categories differentiated according to content and literal sentence matching are reported in Appendix H (p. 180).

Analyzing the ratings according to the content yielded significant main effects of the morality dimension ($\chi^2(3) = 72.918$, $p < .001$), the positive emotionality dimension ($\chi^2(3) = 15.652$, $p = .001$), and the negative emotionality dimension ($\chi^2(3) = 69.339$, $p < .001$), but not for the logic dimension ($\chi^2(3) = 7.688$, $p = .053$).

The morality dimension showed that the moral-related recognition items (moral vs. neutral1: $z = -4.684$, $p < .001$; moral vs. neutral2: $z = -4.707$, $p < .001$; unmoral vs. neutral1: $z = -4.707$, $p < .001$; unmoral vs. neutral2: $z = -4.709$, $p < .001$) were rated higher than both of the neutral ones. The moral-related items did not differ from each other ($z = -.367$, $p = .714$), and so did the neutral ones ($z = -.747$, $p = .455$).

Concerning the positive emotionality dimension, the unmoral items were rated lower than the moral ones ($z = -3.085$, $p = .002$) and the neutral2 items ($z = -2.159$, $p = .031$), but did not differ from the neutral1 items ($z = -1.348$, $p = .178$). The moral items showed no significant differences to both of the neutral recognition items (moral vs. neutral1: $z = -1.635$, p = .102; moral vs. neutral2: z = -1.199, p = .230), and the controls were also rated similar (neutral1 vs. neutral2: z = -1.040, p = .298).

The negative emotionality dimension revealed higher ratings for both moral-related recognition items as compared to the neutral ones (moral vs. neutral1: z = -4.291, p < .001; moral vs. neutral2: z = -4.410, p < .001; unmoral vs. neutral1: z = -4.546, p < .001; unmoral vs. neutral2: z = -4.548, p < .001), whereby the unmoral items received even higher ratings than the moral ones (z = -3.825, p < .001). The neutral recognition items yielded no differences (z = -1.405, p = .160). A graphical illustration of the results of the ratings according to the content categories for the recognition items is shown in Figure 13 below.

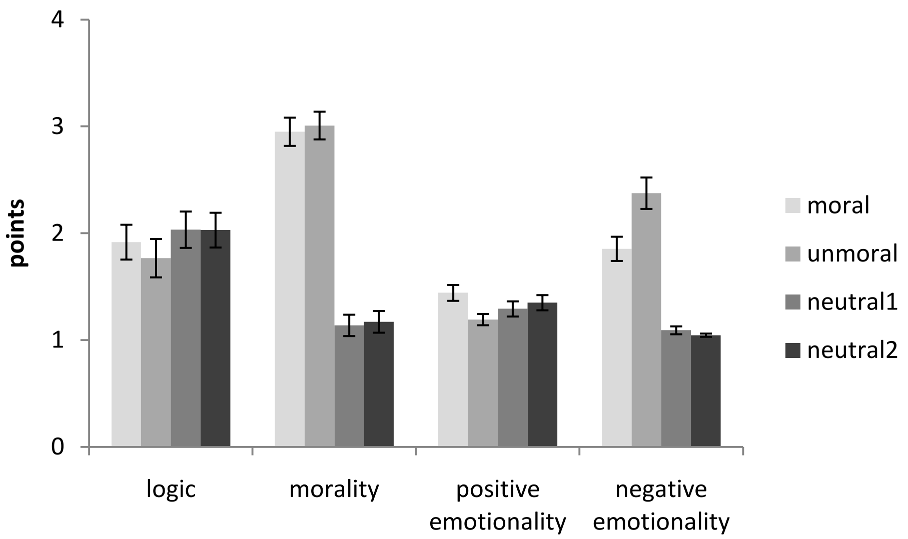


Figure 13: Mean ratings (points 1 to 4) and standard errors for moral, unmoral, neutral1, and neutral2 recognition items in the dimensions logic, morality, positive and negative emotionality.

6.3 Discussion

The pre-study yields evidences that all deductive reasoning problems were rated as intended. They were rated as almost equal in the logic dimension, and moral-related deductive inference problems contained more moral content than the neutral problems. Furthermore, as already found in the second pre-study, the moral-related problems received higher emotional ratings than the neutral controls. Moral problems were rated higher in the positive emotional dimension and unmoral ones in the negative emotional dimension. Therefore, there was a clear distinction between moral-related and neutral content.

The recognition items did not differ in their ratings on the logic dimension, but the ratings were explicitly lower than those for the deductive reasoning problems. The moral-related recognition items received higher ratings in the morality dimension as well as in the emotionality dimension as compared to the neutral items. This time, the emotional ratings of the moral-related items were particularly pronounced in the negative emotionality dimension. Since the moral-related problems were rated as more affectively laden, this seems to once again underpin assumptions that morals are emotionally laden. In totality, the reasoning problems and the recognition items seemed to be useful for the next experiments, which will now be reported.

7. Experiment III – Training[14]

The third experiment implemented a training in logic prior to the computer experiment, manipulating the deductive reasoning abilities of one group of participants. Since this training was based on findings of former training studies, it seems necessary to give a short overview on training studies first, and then report the experiment.

7.1 Excursus - Training in theoretical reasoning

As reported above, misunderstanding the premises of a deductive reasoning problem is the first error a participant can make while solving deductive inference problems. Thus, different studies (e.g., O'Brien & Overton, 1980, 1982; Ziegler, 1990) focused on a better comprehension of the premises and trained participants to avoid such comprehension errors. O'Brien and Overton (1980, 1982) and Overton, Byrnes, and O'Brien (1985) used a contradiction training according to the mental model theory and improved the reasoning abilities of young adults in conditional and syllogism reasoning problems. Ziegler (1990) trained his participants in the Wason Selection Task and found that the group receiving all components of the different training conditions showed the best results as well as transfer effects. Similarly, Klaczynski and Laipple (1993) varied the contents of the selection problems in the trainings and in the later test phase, and found transfer effects as well as facilitation effects of domain independent rules, whereby their training was according to the theory of pragmatic reasoning schemas (Cheng & Holyoak, 1985). Another study investigating the pragmatic reasoning schemas was done by Cheng, Holyoak, Nisbett, and Oliver (1986), who could show that only training with learning and exercising rules via examples produces performance improvements.

Klauer and colleagues (Klauer, Meiser, & Naumer, 2000) compared different training conditions according to different theories of deductive reasoning. They used a syntactic training based on the mental logic theory (Rips, 1994), an abstract semantic training with truth-tables according to the theory of mental models (Johnson-Laird, 1983, 2006), a domain-specific semantic training, and a control training practicing an inductive reasoning task. Both semantic training conditions improved performance, but the syntactic training and the control training did not differ and showed no improvements in performance. Klauer et al. (2000) conclude that the effects are related to a better understanding of the propositional premises and therefore caused facilitation of the construction of appropriate mental models. This assumption is further supported by Klauer et al. (1997), where a truth-table training also

[14] This experiment was conducted within the work for the diploma thesis of Michaela Kandl (Kandl, 2009).

produced substantial training effects in propositional syllogisms. Thus, both of these studies obtained performance improvements in deductive reasoning with trainings aiming at a better understanding of the propositional premises, indicating facilitation of the construction of appropriate mental models. In turn, this supports trainings according to the mental model theory (Johnson-Laird, 1983).

Another approach including training investigates the involvement of working memory in deductive reasoning. In particular, the working memory model of Baddeley (1986, 2003) with its phonological loop, visuo-spatial sketchpad, and central executive has been the focus of diverse experiments. Generally, dual task designs are used presenting a concurrent task according to one specific working memory component testing the disturbing influences of the secondary task and its involvement. The findings concerning the phonological loop are inconsistent (Knauff, 2009a), but for the involvement of the central executive some evidences of studies with conditional and syllogistic reasoning problems exist (e.g., Gilhooly et al., 1993; Klauer et al., 1997).

Using this concurrent task design, Meiser, Klauer, and Naumer (2001) could show that training reduces the impact of heuristics and enhances analytic inference processes which are normally and easily disturbed by the concurrent task. This further supports the mental models theory since mental models also depend on working memory constraints.

To integrate and summarize the different training studies, the main finding is that training can indeed improve deductive reasoning. Hence, the most promising training conditions are those that aid in improving the understanding of the premises by using abstract semantic material like truth tables and/or providing concrete semantic contexts. In addition, training with explanations and examples as well as practical exercises seems to be most effective. Finally, there are hints that trainings according to the mental model theory do not only result in performance increments, but also support this theory due to these improvements.

The training studies described previously yielded evidences that not only rule learning and the like is sufficient in the improvement of reasoning abilities, but also practical exercises and feedback are necessary, thus a very intensive training version was developed for the following experiment. Klauer, and Meiser (2007) stated in their review on training studies in deductive reasoning that it seems helpful to use a training according to the mental model theory, which particularly focuses on improving the ‘pure’ deductive (i.e., rational) reasoning abilities of the participants.

The preceding pre-studies and experiments have revealed some preliminary evidences that theoretical and practical reasoning are based on 'pure' deductive processes (i.e., rational) as well as emotional/intuitive processes. Therefore, a logic training improving the rational proportions of reasoning processes should have no effect in the experiment if reasoning is solely based on emotional/intuitive processes. In contrast, it is not as easy to provide an explanation about the "positive" effects of the training. If the training provides the intended outcomes, participants should not show content effects or incongruity effects. This would endorse assumptions of rational processes involved in theoretical reasoning which could be improved. Since content effects would have occurred without training and underpinned influences of emotional/intuitive processes, such "positive" training effects would strengthen 'dual-process' accounts of theoretical reasoning. With regards to practical reasoning, however, to draw conclusions may seem much more difficult and possible attempts will be reserved for the discussion section below. The newly developed recognition items served as additional task to investigate on the emotional/intuitive processes involved in practical reasoning. Therefore, the concrete hypotheses for the following experiment are:

- Participants receiving logic training should show reasoning improvements due to this training measured by the comparison of the pre- versus the post-test.
- Participants receiving logic training are less prone to errors and show shorter decision times for all deductive reasoning problems versus persons receiving no training or pseudo training.
- Training should result in no performance differences between the three reasoning problem conditions.
- Trained participants should not show an incongruence effect.
- Participants with pseudo training or no training should show an incongruence effect.
- Logic training should not affect the recognition task performance since the training encompasses no components improving recognition memory, thus the groups should not differ in their recognition task performance.
- Moral-related content leads to faster decisions and fewer errors in the recognition task for all participants, especially if the recognition processes involved are solely based on automatic, intuitive, and emotional processes.
- The three groups should not differ in their moral judgment competence (i.e., MJT).

7.2 Method

7.2.1 Participants

Participants, 12 per condition, were randomly assigned to the training, pseudo training, or no training group. In total 36 participants, not formerly tested, participated in the experiment. The group without training consisted of 7 females with a mean age of 23 years (S.D. ±1.73) and 5 males with a mean age of 26.2 years (S.D. ±1.48). The pseudo training group included 7 females with a mean age of 25.43 years (S.D. ±3.51) and 5 males with a mean age of 26.4 years (S.D. ±3.91). The training group encompassed 7 females with a mean age of 23.43 years (S.D. ±3.55) and 5 males with a mean age of 22.40 years (S.D. ±1.14). Again, all participants were native German speakers, naïve to deductive reasoning tasks, right-handed according to Salmaso and Longoni’s Edinburgh Handedness Questionnaire (1985, modified version of Oldfield, 1971) and showed no previous mental illness or psychotropic medication. Participants came from local universities and gave informed written consent according to the Declaration of Helsinki (1964/2008). Participants were financially compensated or received course credits for approximately 80 minutes of testing.

7.2.2 Material

7.2.2.1 Logic Training

The training group received logic training composed of four steps similar to the abstract semantic training used by Klauer and colleagues (2000). First, participants passed a pretest before they received an introductory text about propositional logic, followed by practicing deductive inference problems, and finally, a posttest to measure performance improvements (Figure 14). To avoid additional training effects of working on a computer, especially with deductive inferences, the exercises were done with paper and pencil.

The pretest contained an instruction and an example, followed by 20 deductive inference problems, half of them with a valid or invalid conclusion. These problems had abstract, neutral, absurd, or emotional content, and participants had to judge their logical validity by checking “Yes” or “No” boxes below each problem (problems for the pre-test and post-test as well as the problems and example material of the training can be found in Appendix I, pp. 181-187).

The text about propositional logic, which the participants received after the pretest, started with a brief introduction on the history of logic, followed by an example with an explanatory text about the structure and the elements of a deductive inference problem. After clarifying the terminology and illustrating how deductive inferences are made, emphasizing on the notion that the content is irrelevant for the validity of the inference, simple inference rules were further explained in detail. All contents of the text were supplemented with examples to further support understanding. The text itself was based on an introductory textbook about philosophical logic by Hoyningen-Huene (2006).

After studying the text and clarifying comprehension problems, participants received a worksheet with deductive inference problems and the training phase started. A total of 7 worksheets with 48 problems were used. Half of the deductive inference problems were valid, the other half invalid. Half of the cases contained no conclusion, so participants had to produce it by themselves, whereas the other half included a given conclusion and the validity had to be judged by checking “Yes” and “No” boxes below the problems. The problems used had abstract (4 problems), neutral (4), absurd (8), and emotional (8) content. Abstract problems were presented to familiarize participants with content-free inferences and pointing on the ‘pure’ deductive form of inferences. Neutral content was according to world-/general-knowledge and served as a first step in disrupting the pure analytic reasoning process. Negative emotional as well as absurd content, contradicting our world-knowledge, forced the participants to answer by only using the learned rules while ignoring the content. After completing a worksheet, the answers of the participants were corrected and participants received feedback. If necessary, further explanations were given and difficulties were discussed.

Finally, after all worksheets were processed, a posttest, with the same structure as the pretest, had to be filled out. Comparing the number of errors between pre- and posttest served as an indicator for training success.

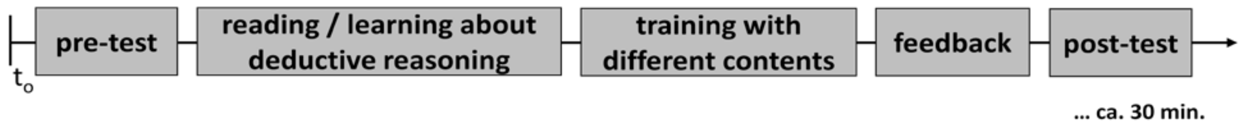


Figure 14: Procedure of the logic training.

7.2.2.2 Pseudo Training

The pseudo training, like the logic training, was similar to that used by Klauer et al. (2000) and was based on the “Bochumer Matrizen Test” (Hossiep, Turck, & Hasella, 1999). The problems of this training were matrices, namely visual analogy problems, and consisted of rectangles with 15 fields (three rows and five columns) containing symbols arranged by a certain principle. One of the 15 fields is empty and the task requires from the participants to select one symbol from six given symbols which logically complements the pattern of the other 14 icons. This pseudo training was chosen since it resembles an inductive reasoning task without any semantic aspects and should therefore not affect the deductive reasoning problems in the experiment. On the other hand, the pseudo training allows to control for general effects of training in problem solving strategies, e.g., systematically following a procedure, as well as for motivational aspects of training per se.

The pseudo training had no pre- or post-test, but was, apart from that, matched to the logic training (see Figure 15). First, participants received an introductory text with examples and a description of the 13 problem solving strategies, as being relevant in solving inductive problems. Then, a practice phase followed in which 10 problems had to be completed by using the learned strategies. Participants had to complete 5 pages with two tasks per page and again, after each page, their tests were corrected and participants received feedback.The duration of both types of training was equivalent. Problems and training material for the pseudo-training are reported in Appendix J (pp. 188-193).

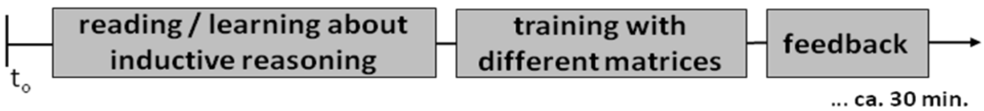


Figure 15: Procedure of the pseudo training.

7.2.2.3 Moral Judgment Test

To obtain an external criterion for the participants' moral judgment level, the 'Moral Judgment Test' (MJT) by Lind (1978, 2008) was used to measure moral judgment competence. The MJT is an advancement of Kohlberg's dilemmas (Kohlberg, 1969) or Rest's 'Defining Issue Test' (DIT, Rest, 1974), as these tests assess individual moral attitudes (Prehn, Wartenburger, Mériau, Scheibe, Goodenough, Villringer, van der Meer, & Heekeren, 2008) rather than moral decision making competences. The MJT is based on a kind of 'dual-process' model with rational and affective components explaining the moral behavior and moral ideals of a person. Both portions of this model influence a so-called competence-score (c-score). This c-score could range from zero to 100 (Lind, 2008), whereby higher values indicate higher moral judgment competence (Table 12). The MJT consists of two dilemmas, each including two competing moral principles. The protagonists involved violate one of these two moral principles. Participants read these stories and are asked to judge the actions of the protagonists in general. Then, six pro and contra arguments concerning the protagonists behavior have to be rated on a scale ranging from "I strongly reject (-4)" to "I strongly accept (+4)". These ratings and the answers to the general questions lead to the c-score.

Table 12: Categorization of c-scores and related moral judgment competence (adapted from Lind, 2008).

Moral Judgment Competence (c-score)			
1 – 9	very low	**30 – 39**	high
10 – 19	low	**40 – 49**	very high
20 – 29	medium	**50 – 100**	extraordinary high

7.2.2.4 Computer Experiment

The stimulus material for the experiment consisted of the deductive inference problems with moral, unmoral, neutral1, and neutral2 content as well as the recognition items with equivalent contents as was evaluated in Pre-Study III reported earlier. All problems appeared twice, once with a valid conclusion and in the other instance with an invalid conclusion, totaling 32 inference problems. The recognition items on the other hand, appeared twice, once with a third sentence literally matching a part of the first sentence presented previously, and in the other instance with no literal match. This third experiment therefore, contained the between-subject factor group (training, pseudo training, and no training) and the within-subject factor task (deductive inference problems and recognition items) representing a 3 (group) by 2 (task) mixed factorial design (Table 13).

Table 13: Design of Experiment III

		group		
		logic training	pseudo training	no training
task	deductive reasoning	32 reasoning problems	32 reasoning problems	32 reasoning problems
	recognition	32 reasoning problems	32 reasoning problems	32 reasoning problems

Separated into categories according to the task, the part with the reasoning problems represented a 4 (content) by 2 (validity) by 3 (group) mixed factorial design containing the within-subject factors content and validity and the between-subject factor group. The part with recognition items also represented a 4 (content) by 2 (matching) by 2 (group) mixed factorial design containing the within-subject factors content and matching and the between-subject factor group (see also Pre-Study III above for graphical illustration of the within-subject factorial design).

7.2.3 Procedure

Participants without training were directly tested on the computer experiment and subsequently completed the MJT. In the training and the pseudo training condition, a maximum of four participants at a time participated. They received their training, followed by a 20 minute break for refreshment. The break was strategically given to avoid a lack of concentration or attention, since it is known that the attention span of students is only about 15 – 20 minutes (e.g., Middendorf & Kalish, 1996). Then, two of the participants were immediately subjected to the computer experiment and the MJT, followed by the two others shortly thereafter.

A fundamental change compared to Experiments I and II pertained to reading and response time limits for this experiment. A preliminary training study, not reported here, had shown that training effects could be obtained easier if time limits are introduced. Thus, the first premise/sentence was presented for 4 seconds, the second one lasted 2 seconds, and finally the conclusion/third sentence offered a time window for making the validity/matching judgment of 3.5 seconds (answers given after 3.5 seconds were labeled as errors). Answers were given on a response-box ("Yes" for valid/match, "No" for invalid/no match), whereby response buttons were counterbalanced for hand preference between participants. After each reasoning problem/recognition item, a fixation cross was shown for two seconds. Reasoning problems and recognition items were presented intermittently and in random. The conclusion of a reasoning problem was presented in red color, the third sentence of a recognition item in green color, to indicate which task had to be executed by the participants. Participants absolved four practice trials after a short introduction on the computer, familiarizing themselves with the procedure. See Figure 16 for an illustration of the procedure.

The remainder of the procedure was identical to Experiment I.

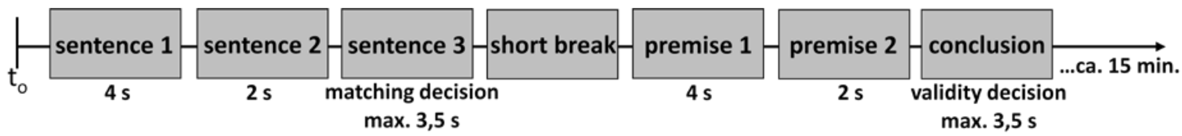


Figure 16: Procedure of the computer experiment showing an example of a recognition item followed by a deductive reasoning problem.

7.3 Results

The analysis included a general linear model to control that the two tasks did not influence the performance of the three groups differently. Then, a pre-post-test comparison of the group receiving logic training was calculated with a paired-sample t-test to explore the intended performance enhancement in deductive reasoning. It was made sure that no moral judgment competence differences between the groups exist. The error rates and decision times of the deductive reasoning problems and the recognition items were analyzed separately, between and within the groups, applying GLMs and post hoc t-tests. Due to the fact that the reasoning problems and recognition items of the two different tasks were presented intermittently and randomly, and that each stimulus sub-category only contained four reasoning problems/recognition items, it was impossible to control influences of one stimulus class on the other concerning participants' performance. A potential analysis failed since no values were available or the number of values varied extremely between stimulus sub-categories and participants that no serious statistical approach could be chosen. Thus, whether reasoning problems preceding recognition items affected participants' performance, or vice versa, could not be controlled and therefore possible influences on performance could not be excluded.

7.3.1 Task by group analysis

On a descriptive level, error rates yielded higher values for the recognition items (32.30%) than for the deductive reasoning problems (11.37%). Trained participants (15.63%) were less prone to errors in both tasks as compared to those with pseudo training (23.83%) and participants without training (25.91%). Figures 17 and 18 illustrate these descriptive results. The GLM for repeated measures of the error rates showed a task main effect ($F(1,33) = 31.119$, $p < .001$), but no group main effect ($F(2, 33) = 1.876$, $p = .169$) and no interaction effect for group by task ($F(2, 33) = .045$, $p = .956$).

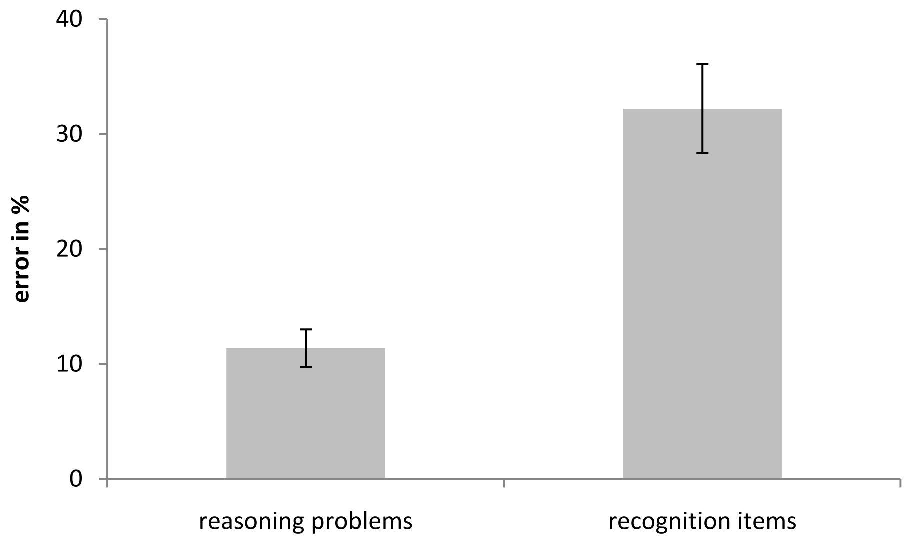


Figure 17: Mean error rates in percent (%) with standard errors for deductive reasoning problems and recognition items.

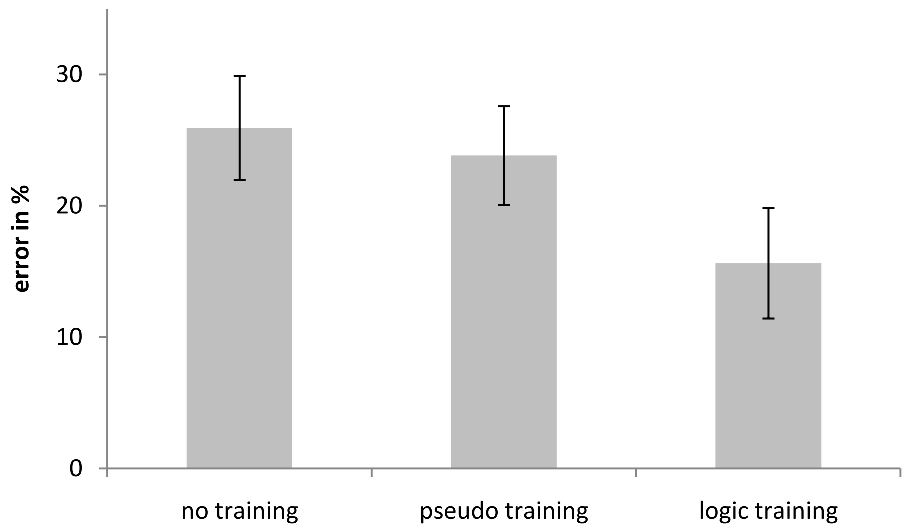


Figure 18: Mean error rates in percent (%) with standard errors for the group without training, the group with pseudo training, and the group with logic training.

Decision times yielded higher values for the recognition items (1735.45 ms) as compared to the deductive reasoning problems (1623.62 ms), and the group receiving logic training (1438.03 ms) was faster than both of the control groups (pseudo training: 1790.90 ms; no training: 1809.69 ms). See Figures 19 and 20 for an illustration. Statistically, a task main effect also occurred ($F(1, 33) = 9.487$, $p = .004$), and a group main effect ($F(2, 33) = 4.319$, $p = .022$), but again no interaction effect could be found ($F(2, 33) = 1.972$, $p = .155$). Since these analyses revealed that the groups were not affected differently by the two tasks, further analyses were conducted separately for the two tasks.

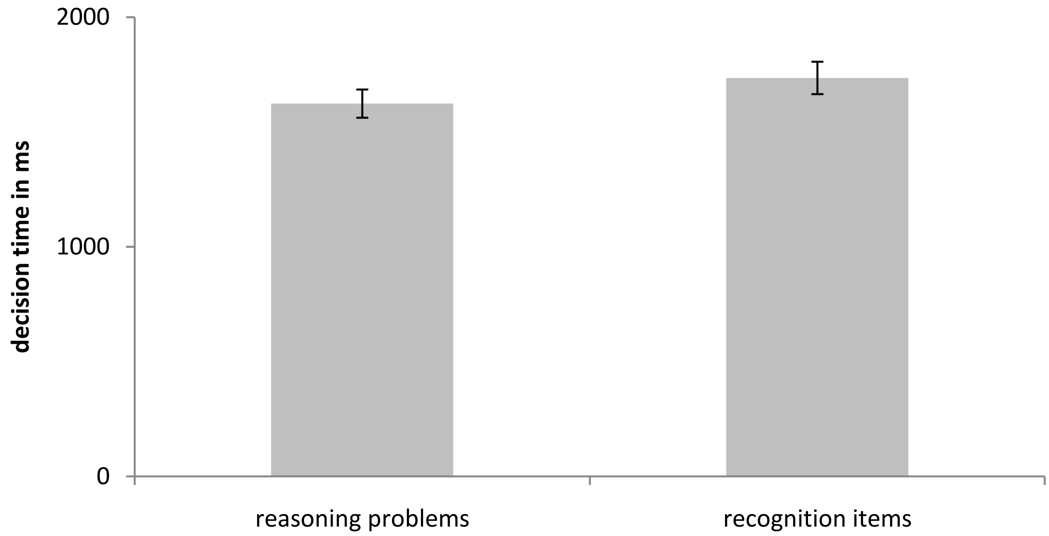


Figure 19: Mean decision times in milliseconds (ms) with standard errors for deductive reasoning problems and recognition items.

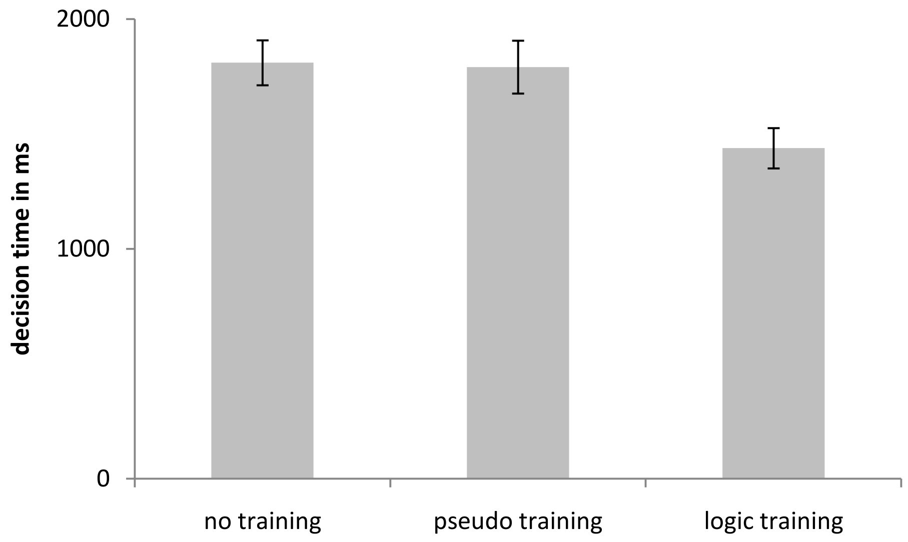


Figure 20: Mean decision times in milliseconds (ms) with standard errors for the group without training, the group with pseudo training, and the group with logic training.

7.3.2 Pre-post-test comparison and MJT

The paired-sample t-test for the comparison of the errors in the pre- versus the post-test revealed a significant difference ($t(11) = 2.327$, $p = .040$) indicating that participants produced less errors in the post test and therefore benefited from the training.

For the Moral Judgment test, the trained group reached a c-score of 29.37 points, the group with pseudo training 23.73 points, and the untrained participants 30.34 points. Although the c-score values were different, these differences did not reach significance. The trained group did not differ from the untrained one ($t(22) = .134$, $p = .895$) and neither from the group with pseudo training ($t(22) = .848$, $p = .406$) as well as the untrained group did not discern from the other control group with the pseudo training ($t(22) = .986$, $p = .335$).

7.3.3 Deductive reasoning problems

7.3.3.1 Error rates

Descriptively, participants with pseudo training (13.80%) or without training (15.89%) were more prone to errors than the trained participants (4.43%). Thereby, the moral (18.06%) and unmoral (16.67%) inference problems had higher error rates than the neutral1 (7.29%) or neutral2 (3.47%) ones.

The general linear model revealed main effects for content ($F(2.568, 84.744) = 18.394$, $p < .001$) and group ($F(2, 33) = 5.930$, $p = .006$), but not for validity ($F(1, 33) = 2.026$, $p = .164$). Content and validity ($F(1.824, 60.207) = 11.706$, $p < .001$) showed an interaction effect, whereas content and group ($F(5.136, 84.744) = 1.461$, $p = .210$) as well as validity and group (F(2, 33) = .533, p = .592) were insignificant. The three-way interaction of content, validity and group (F(3.649, 60.207) = 2.527, p = .055) showed a trend and with the Huynh-Feldt correction (F(4.085, 67.407) = 2.527, p = .047) it reached significance.

Post hoc t-tests for the content effect (Figure 21) revealed that moral (vs. neutral1: t(35) = 3.872, p < .001; vs. neutral2: t(35) = 6.468, p < .001) and unmoral problems (vs. neutral1: t(35) = 3.666, p = .001; vs. neutral2: t(35) = 5.414, p < .001) were more prone to errors than the neutral problems. Additionally, neutral1 problems also resulted in higher error rates than the neutral2 ones (t(35) = 2.142, p = .039). Moral-related problems did not differ (t(35) = .598, p = .554).

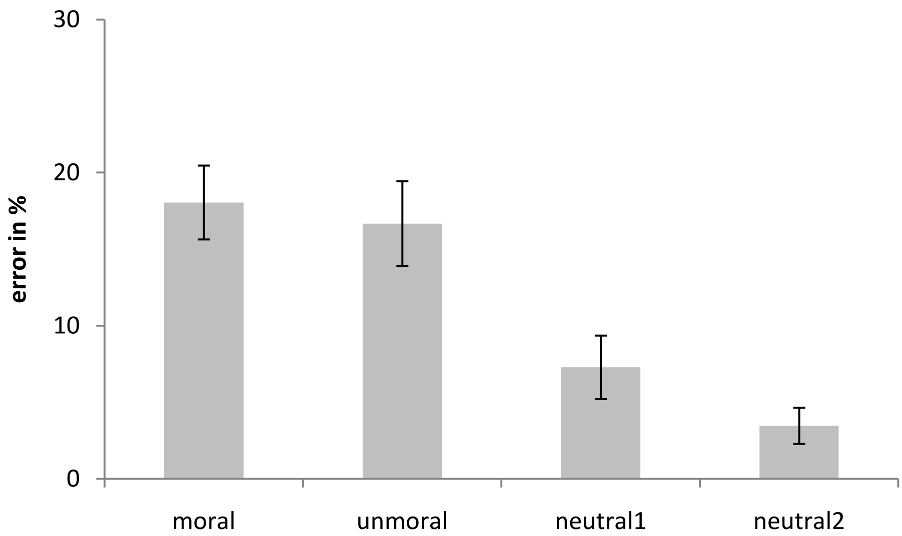


Figure 21: Mean error rates in percent (%) with standard errors for moral, unmoral, neutral1, and neutral2 problems.

Group comparisons (Figure 22) demonstrated that the trained participants made less errors than the group with the pseudo training (t(22) = 2.585, p = .017) and the group without training (t(22) = 4.057, p = .001), whereas the pseudo training and the no training group did not differ (t(22) = .512, p = .614, p = .614).

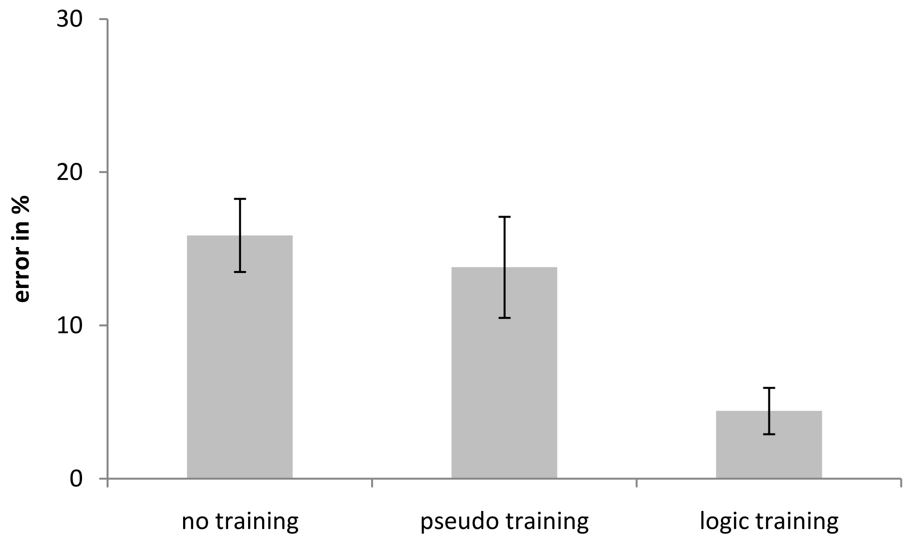


Figure 22: Mean error rates in percent (%) with standard errors for the group without training, the group with pseudo training, and the group with logic training.

The single contrasts concerning the significant interaction of content and validity are shown in Appendix K (p. 194, table also includes decision times).

Following the significant three-way interaction of content, validity and group, the second General Linear Model showed significant interaction effects only for the group without training ($F(1.508, 16.590) = 10.437$, $p = .002$), whereas for the group with the pseudo training it barely failed significance, but revealed a trend ($F(2.159, 23.746) = 2.548$, $p = .096$). For the trained group no significant difference could be found ($F(1.522, 16.739) = .641$, $p = .498$). Paired-sample t-tests for the untrained group are shown in Appendix K (p. 194). It can be seen that mainly the moral-invalid and the unmoral-valid problems (incongruent problems) differed from all others in this group. Descriptively, the same is true for the group with the pseudo training (Figure 23).

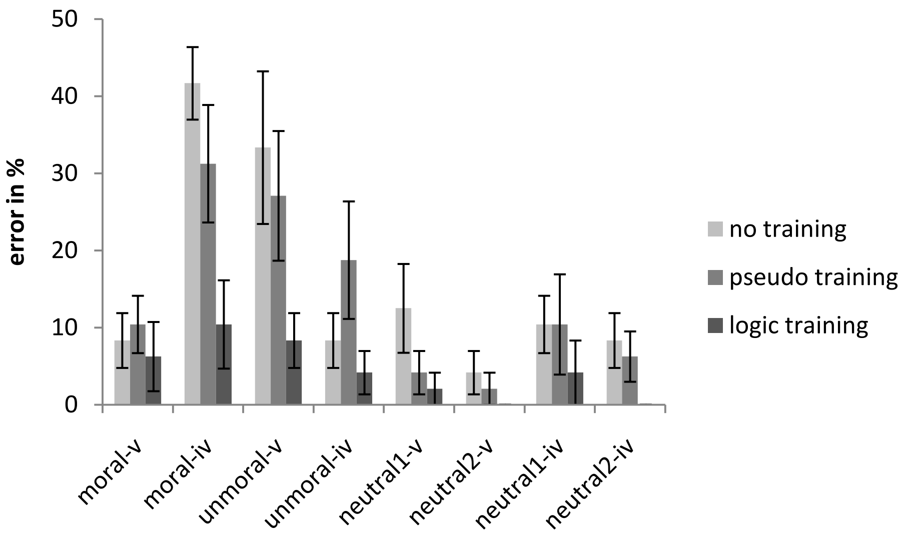


Figure 23: Mean error rates in percent (%) with standard errors for moral, unmoral, neutral1, and neutral2 problems, separated for validity (v = valid, iv = invalid) with the group without training, the group with pseudo training, and the group with logic training.

The independent sample t-tests for the content and validity interaction pair wise comparing the different groups for the single conditions are also reported in Appendix K (p 195).

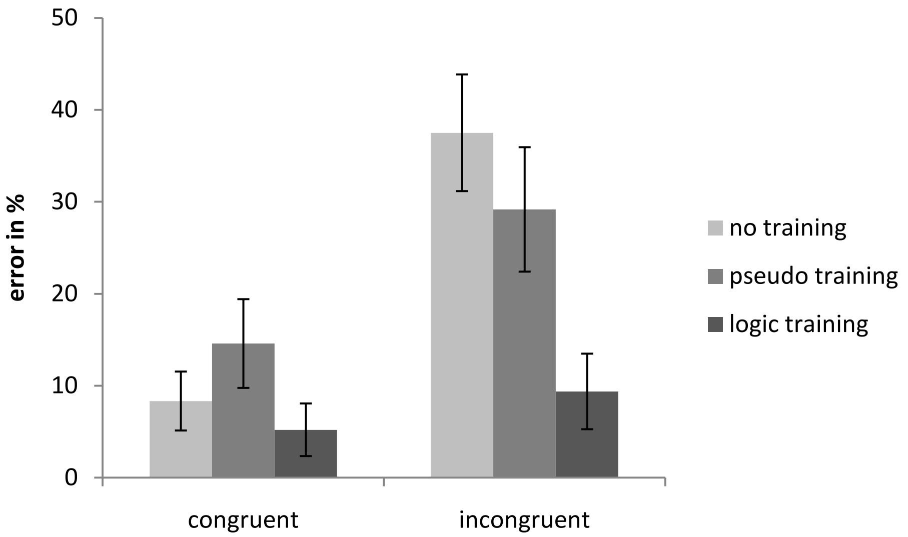


Figure 24: Mean error rates in percent (%) with standard errors for congruent (moral-valid + unmoral-invalid) and incongruent (moral-invalid + unmoral-valid) moral-related problems for the group without training, the group with pseudo training, and the group with logic training.

Analyzing only the moral-related problems concerning the incongruence effect revealed main effects for content ($F(1, 33) = 16.578$, $p < .001$) and group ($F(2, 33) = 6.076$, $p = .006$) as well as an interaction effect ($F(2, 33) = 3.416$, $p = .045$), basically indicating that congruent problems (moral-valid and unmoral-invalid) were less prone to errors than incongruent ones ($t(35) = -3.817$, $p = .001$). For graphical illustration see Figure 24 above.

Independent-sample t-tests showed that training reduced error rates (training vs. pseudo-training: t(22) = 2.727, p = .014; training vs. no training: t(22) = 3.897, p = .001; pseudo-training vs. no training: t(22) = .188, p = .853). However, this effect was especially related to the incongruent problems (training vs. pseudo training: t(22) = 2.501, p = .020; training vs. no training: t(22) = 3.722, p = .001; pseudo-training vs. no training: t(22) = .899, p = .379). This simply means that only the untrained participants (t(11) = -3.626, p = .004) and the participants with the pseudo training (t(11) = -2.028, p = .067) showed the incongruence effect, at least as a trend, whereas the trained participants were not affected by the conflict between validity status and morality status (t(11) = -.886, p = .394). The congruent problems revealed no significant group differences.

7.3.3.2 Decision times

Descriptively, trained participants (1428.66 ms) answered faster than those with pseudo training (1693.69 ms) or without training (1748.51 ms). Mainly the moral (1740.19 ms) and unmoral (1673.65 ms) inference problems were affected compared to the neutral1 (1509.69 ms) or neutral2 (1570.97 ms) problems.

Decision times yielded significant main effects for validity (F(1, 33) = 20.050, p < .001) and content (F(2.285, 75.391) = 12.011, p < .001), whereas the group factor missed significance (F(2, 33) = 2.861, p = .071). In addition, content and group showed a trend for an interaction effect (F(4.569, 75.391) = 2.389, p = .051) significantly in the Huynh-Feldt corrected analysis (F(5.225, 86.213) = 2.389, p = .042). The interaction of content and validity reached significance (F(2.929, 96.664) = 3.582, p = .017) with details of the single comparisons shown in Appendix K (p. 194, together with error rate comparisons). There was no significant interaction of group by validity (F(2,33) = .397, p = .676) and no three-way interaction (F(5.858, 96.664) = .971, p = .448).

Post hoc t-tests revealed that valid deductive inferences were answered faster than invalid ones (t(35) = -4.557, p < .001), whereas moral (vs. neutral1: t(35) = 4.588, p < .001; vs. neutral2: t(35) = 3.176, p = .003) and unmoral problems (vs. neutral1: t(35) = 3.983, p < .001; vs. neutral2: t(35) = 2.241, p = .031) required significantly more time than the neutral ones, which also differed from each other (neutral1 faster than neutral2: t(35) = -2.317, p = .026). See Figure 25 below.

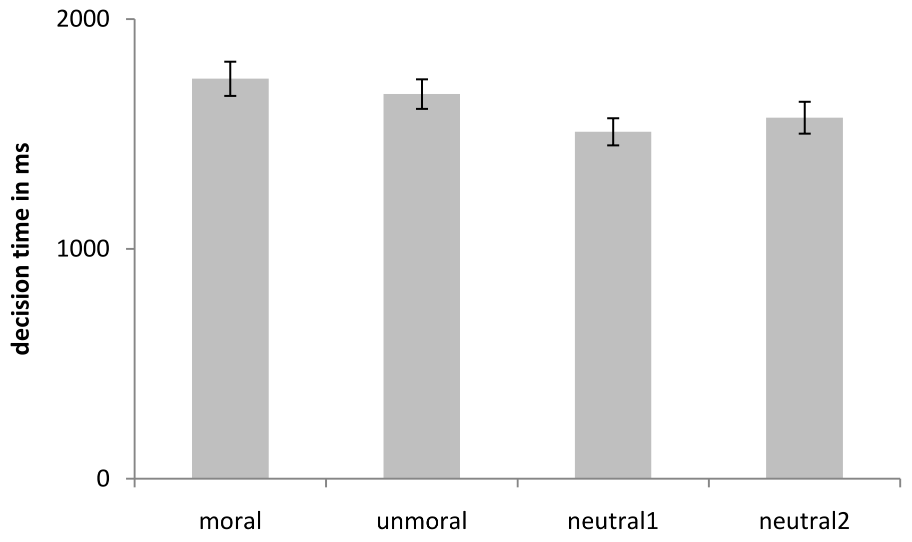


Figure 25: Mean decision times in milliseconds (ms) with standard errors for moral, unmoral, neutral1, and neutral2 problems.

Following the content and group interaction, it could be shown that the group without training and the trained participants differed (unmoral: t(22) = 2.083, p = .049; neutral1: t(22) = 3.672, p = .001; neutral2: t(22) = 3.089, p = .005), indicating that the untrained participants needed more time for decision making than the trained group. The group with logic training was always faster than the groups without training or with pseudo training (Figure 26), even though these differences barely missed significance (see Appendix K, p. 196 for all independent-sample t-test comparisons).

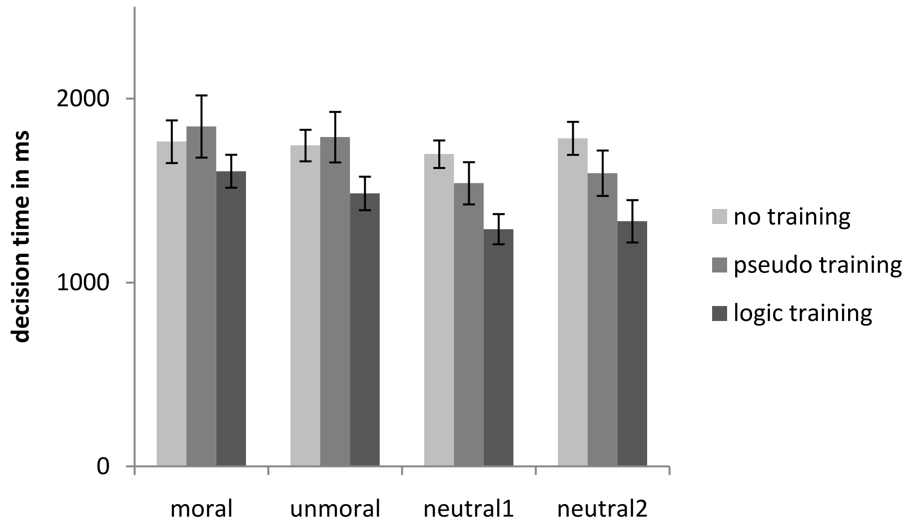


Figure 26: Mean decision times in milliseconds (ms) with standard errors for moral, unmoral, neutral1, and neutral2 problems for the group without training, the group with pseudo training, and the group with logic training.

Separated for the groups, the group with pseudo training had higher decision times for the moral (vs. neutral1: t(11) = 2.983, p = .012; vs. neutral2: t(11) = 2.477, p = .031) and unmoral problems (vs. neutral1: t(11) = 3.304, p = .007; vs. neutral2: t(11) = 2.555, p = .027) compared to the neutral problems. The moral-related problems (t(11) = .606, p = .557) and the neutral ones (t(11) = -1.182, p = .262) did not differ in the pseudo training group. This is similar to what could be found for the trained group (moral vs. unmoral: t(11) = 2.304, p = .042; moral vs. neutral1: t(11) = 6.107, p < .001; moral vs. neutral2: t(11) = 3.411, p = .006; unmoral vs. neutral1: t(11) = 3.034, p = .011, unmoral vs. neutral2: t(11) = 1.790, p = .101; neutral1 vs. neutral2: t(11) = -.943, p = .366). In contrast, the group without training showed almost the same decision times in all categories (moral vs. unmoral: t(11) = .394, p = .701; moral vs. neutral1: t(11) = .805, p = .438; moral vs. neutral2: t(11) = -.248, p = 808; unmoral vs. neutral1: t(11) = .715, p = 489; unmoral vs. neutral2: t(11) = -.697, p = .556; neutral1 vs. neutral2: t(11) = -1.771, p = .104).

Since the three-way interaction and main effect of group failed to reach significance, no incongruence effect with respect to group differences could be reported for decision times. It could only be mentioned that all groups were slower in answering the congruent problems as compared to the incongruent ones (ME content: F(1, 33) = 6.050, p = .019; t(35) = 2.529, p = .016).

7.3.4 Recognition items

7.3.4.1 Error rates

On a descriptive level, trained participants (26.82 %) were less prone to errors than the group with the pseudo training (33.85 %) or the untrained participants (35.94 %). With respect to the different contents, both of the neutral item conditions (35.76 % each) showed higher error rates than the moral (27.78 %) and unmoral ones (29.51 %).

The general linear model for repeated measures analyzing the recognition items revealed a main effect of content (F(2.479, 81.805) = 3.838, p = .018), but no main effects of ‘matching’ (F(1, 33) = 2.375, p = .133) or group (F(2, 33) = .493, p = .615). Furthermore, no two-way interaction effects (content by group: F(4.958, 81.805) = 1.098, p = .368; content by matching: F(2.748, 90.691) = 2.452, p = .074; group by matching: F(2, 33) = .278, p = .759) or three-way interaction occurred (content by group by matching: F(5.496, 90.691) = 1.114, p = .360).

Comparing the different content conditions yielded significant differences of the moral items as compared to the neutral1 (t(35) = -2.242, p = .031) and neutral2 ones (t(35) = -2.360, p = .024) as well as of the unmoral items as compared to the neutral1 (t(35) = -2.168, p = .037) and neutral2 ones (t(35) = -2.311, p = .027). The moral-related (t(35) = -.564, p = .576) and the neutral recognition items (t(35) = .000, p = 1.000) did not differ from each other. A graphical illustration of these results is shown in Figure 27.

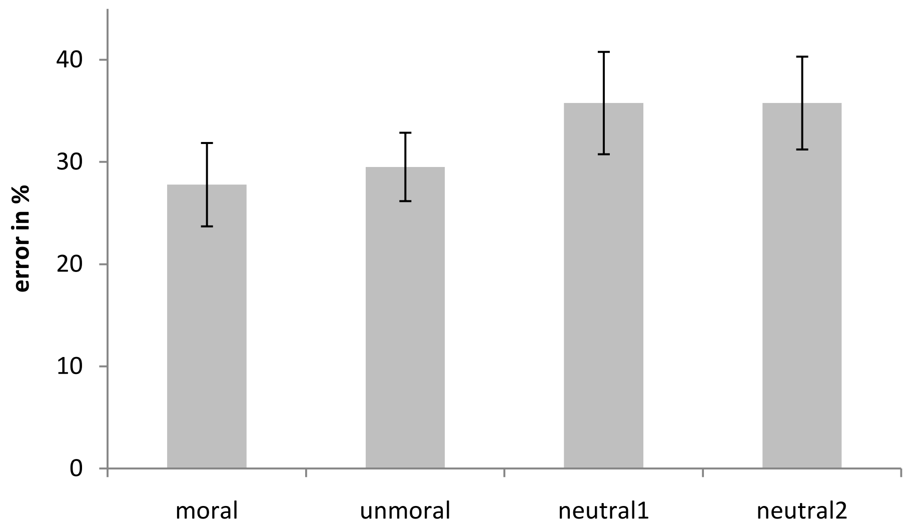


Figure 27: Mean error rates in percent (%) with standard errors for moral, unmoral, neutral1, and neutral2 recognition items.

7.3.4.2 Decision times

Descriptively, the group with logic training (1447.39 ms) had shorter decision times than the groups with pseudo training (1888.10 ms) or without training (1870.86 ms). The content categories did not really differ, but the literal match items (1609.32 ms) were answered faster than the no literal match ones (1861.58 ms).

The general linear model revealed main effects of matching ($F(1, 33) = 20.524$, $p < .001$) and group ($F(2, 33) = 5.172$, $p = .011$), but no effect of content ($F(2.746, 90.606) = 1.026$, $p = .380$). Again, as for the error rates, no two-way interactions (content by group: $F(5.491, 90.606) = 2.025$, $p = .076$; matching by group: $F(2, 33) = .189$, $p = .829$; content by matching: $F(2.502, 82.580) = 1.796$, $p = .163$) or three-way interaction occurred (content by group by matching: $F(5.005, 82.580) = 2.066$, $p = .078$).

The difference of the decisions for the literal match and no literal match items is shown in Figure 28.

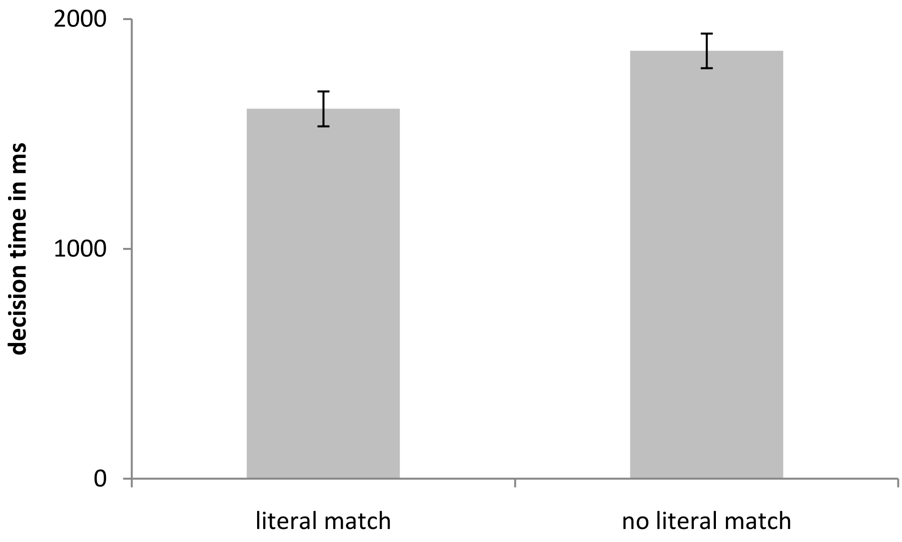


Figure 28: Mean decision times in milliseconds (ms) with standard errors for the literal match and the no literal match recognition items.

The post hoc t-tests according to the different groups showed that trained subjects differed from the group without training (t(22) = 2.724, p = .012) and the group with the pseudo training (t(22) = 2.910, p = .008), whereas the control groups did not differ (t(22) = -.109, p = .914). See Figure 29.

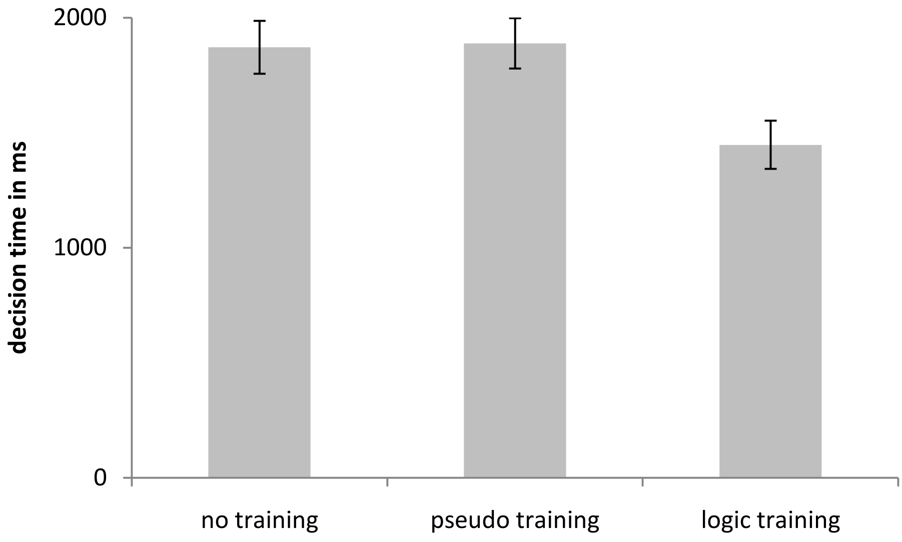


Figure 29: Mean decision times in milliseconds (ms) with standard errors for the group without training, the group with pseudo training, and the group with logic training.

7.4 Discussion

The trained participants made less errors comparing pre- and post-test, and also showed transfer effects and performance increments in the deductive inferences in the computer experiment. They made fewer errors than both control groups, were not subjected to incongruity effects, and were also faster in solving the reasoning problems, but one group comparison (trainings versus pseudo training) barely failed significance. It could be concluded therefore, that the logic training revealed the intended positive outcomes and replicated the effects of the training study by Klauer et al. (2000). This further indicates that an abstract semantic training with practical examples and exercises improves reasoning performance.

Furthermore, the results seem to support the assumption of rational and emotional/intuitive processes involved in theoretical reasoning, thus supporting ‘dual-process’ accounts of theoretical reasoning (e.g., Evans, 2003, 2008). Since participants without or with pseudo training showed content effects, as has been shown also in the two preceding experiments, there is evidence of influences of (perhaps emotional laden) prior knowledge on theoretical reasoning. Moreover, the training of rational reasoning proportions which enabled trained participants to suppress these content influences shows the impact of these rational process proportions on theoretical reasoning. If rational processes were not involved in theoretical reasoning, training these proportions of reasoning processes would have revealed no effects. With regards to practical reasoning however, it seems a little bit more complicated to draw serious conclusions. One might still suggest that practical reasoning is based on emotional/intuitive processes, and especially prior knowledge. This accounts for both of the control groups and the content effects of the moral-related problems, which again indicated that participants did some kind of moral reasoning and moral decision making. Descriptively, the same applies to the trained participants where moral-related problems still differed slightly from the neutral ones. Regarding potential rational proportions of the reasoning processes involved in practical reasoning however, reveals no clear and serious conclusion, since training of rational processes almost eliminated practical reasoning. This practical reasoning has only occurred before due to the influence of prior knowledge on deductive reasoning problems, therefore validating the involvement of emotional/intuitive processes in practical reasoning. Thus, additional experiments seemed necessary, brought upon by the current experimental design, to further explore the potential rational processes contributing to practical reasoning.

The recognition items yielded evidences that prior knowledge, when highly emotional or salient information is taken into account, facilitates recognition of familiar content in contrast to unrelated, less familiar and less salient neutral material (Ferré, 2003; Inaba et al., 2005; Kensinger & Corkin, 2003; Phelps et al., 1997). In particular, participants produced fewer errors with the moral-related recognition items as compared to the neutral ones, although decision times were almost equal between these conditions. These diminished error rates for the moral-related recognition items, probably caused by prior knowledge, provide further evidence for learned morality or strong moral attitudes. However, there is again no possibility of gaining insight via the recognition task performance on the potential involvement of rational processes in practical reasoning.

Interestingly, the training also improved the performance of the trained participants concerning the decision times for the recognition items. Although training should not influence recognition performance, trained participants had shorter decision times (and lower error rates) than both of the control groups. This improvement in performance might be due to the task structure and due to the fact that trained participants have already been familiarized with it especially since differentiation between the two task requirements could already be done after the second sentence appeared which did not match the form of a deductive inference problem in the recognition task. Another interesting point to take note of is that all groups showed worse performance in the recognition task as compared to the deductive reasoning problems, even though recognition is believed to be one of the easiest (memory) tasks among others (Anderson, 1999). This might be due to the fact that all participants were recruited to take part in an experiment on logical thinking and were therefore primed. Furthermore, there is evidence that participants reasoned utilizing emotional/intuitive and rational processes in solving deductive reasoning problems. If the rational proportions of these reasoning processes involved are based on mental models (e.g., Johnson-Laird, 2006), participants might have built mental models after presentation of the first premise. Constructing a mental model means to build a semantic representation of the first premise/sentence, but that the literal wording gets lost. If then the second sentence appears indicating that no deductive inference has to be drawn but recognition is required, the necessary information is no longer available.

So far, there is evidence that theoretical and practical reasoning are based on emotional/intuitive processes, and that theoretical reasoning alone is also based on rational processes. For Practical reasoning, empirical evidence for the involvement of rational processes is still missing. The next experiment tried to investigate this assumption. Concerns related to mediating influences of working memory on deductive reasoning and recognition will be discussed after the next experiment and in the general discussion, respectively.

8. Experiment IV – Intelligence[15]

The fourth experiment was conducted to further explore the up to date yet still unclear involvement of rational processes in practical reasoning as well as the influences of intelligence and working memory on practical and theoretical reasoning. A relation between intelligence level and moral reasoning/judgment competence has already been proposed by Kohlberg (1969) who denoted "higher"/mature moral judgment levels with preceding cognitive development (see also introduction above). Also, connections between deductive reasoning and intelligence have been proposed (e.g., Evans, Handley, Neilens, & Over, 2007). The main focus of this experiment however, was on differences in adult moral judgment and intelligence level, and whether these differences are based on *crystallized* or *fluid* intelligence (Pasupathi & Staudinger, 2001). The concept of fluid and crystallized intelligence proportions was of special interest due to its affinities to working memory capacity (fluid IQ) and accumulated knowledge (crystallized IQ). Thus, not only 'simple' group differences were calculated for the following experiment, but also a correlative analysis to obtain evidence for the particular influences of these two different intelligence components. As external criterion of participants' moral judgment competence, the MJT was applied (Lind, 2008). As was in the preceding experiments, the recognition task was also included to obtain further evidences for the involvement of emotional/intuitive processes in practical reasoning. To obtain clear performance differences, extreme groups were chosen; one with average IQ and the other with superior IQ (see methods section below). The concrete hypotheses for the following experiment are:

- Participants with high intelligence quotient should produce lower error rates and shorter decision times than participants with average intelligence quotient in both tasks.
- Participants with high intelligence should not be subject to an incongruence effect in the moral-related reasoning problems as compared to the group with average intelligence.
- Participants of superior intelligence should have a higher moral judgment competence in the MJT than participants of average intelligence.
- Both groups should show shorter decision times and lower error rates for the moral-related recognition items compared to the neutral ones.

[15]This experiment was conducted within the work for the diploma thesis of Alexandra Schmoranzer (Schmoranzer, 2009).

- The different intelligence components (fluid and crystallized) should correlate specifically with experimental performance on deductive reasoning problems: fluid intelligence with overall performance and crystallized intelligence with knowledge, thus performance on moral-related problems.

8.1 Method

8.1.1 Participants

Participants came from local universities, internet, and the 'Mensa club' (a society for high-intelligence people) as well as the German 'Schülerkademie'. At a first appointment all participants had to fill out two intelligence tests. Fluid intelligence level was obtained by implementing the first part of the 'culture-fair-test' (CFT-20-R; Weiß, 2006) and for crystallized intelligence the 'knowledge test' of the 'intelligence-structure-test' (IST-2000-R, Liepmann, Beauducel, Brocke, & Amthauer, 2001) was applied. After screening 47 participants, 14 participants met the intelligence score criteria. Just a fair match for age or gender could be provided here, due to the difficulty of finding such extreme IQ values. The seven high intelligence participants (2 females) had an IQ-score of 136.57 (S.D. ±5.13) in the CFT-20-R, and 131.92 (S.D. ±9.98) in the IST-2000-R, aged 22 to 37 years, while the average intelligence group (6 females) reached an IQ-score of 106.86 (S.D. ±5.87) in the CFT-20-R, and 100.86 (S. D. ±9.4) in the IST-2000-R, aged 19 to 25 years. Groups differed significantly in CFT ($z = -3.155$, $p = .001$) and IST ($z = -3.148$, $p = .001$). Again, all participants were native German speakers, naïve with respect to logical reasoning tasks, right-handed, and did not participate in one of the previous experiments. In turn, participants gave informed written consent according to the Declaration of Helsinki (1964/2008), and were financially compensated.

8.1.2 Material

Two intelligence tests (see also above) were used to select participants with high and average intelligence. The computer experiment applied the reasoning problems and the recognition items described in Pre-Study III with moral, unmoral, and neutral contents, thus representing a 2 (task) by 2 (group) mixed factorial design with the between-subject factor group and the within-subject factor task. Separated according to the task category, the part with the reasoning problems represented a 4 (content) by 2 (validity) by 2 (group) mixed factorial design, whereby the first two factors represent within-subject factors, and the third factor the between-subject factor. The recognition items in turn represented also a 4 (content) by 2 (matching) by 2 (group) mixed factorial design, with the first two factors as within-subject factors, and the third one as the between-subject factor. Finally, the MJT (Lind, 1978, 2008) was implemented to obtain an external criterion for potential group differences concerning moral judgment competence due to intelligence differences.

The CFT-20-R (Weiß, 2006) is an advanced paper and pencil test based on the concept of fluid intelligence by Cattell (1943, 1963). It is independent of language, but graphically dependent and consists of two parts with four subtests. The tasks involved demand series continuation, classifications, matrices, and topologies, where the correct answers have to be chosen out of five (Weiß, 2006). Since both main parts contain the same thematic subtests, a short form could be applied resulting in a validated IQ-Score for fluid intelligence. This version was chosen for the current experiment.

The IST-2000-R (Liepmann et al., 2001) represents an advancement of the IST (Amthauer, 1953) and IST-70 (Amthauer, 1973) intelligence tests. The basic part measures several intelligence factors according to the intelligence concept by Thurstone, whereas the additional part allows for obtaining crystallized IQ values (Liepmann et al., 2001). The current experiment exclusively dealt with the knowledge test, which measures verbal, numeric, and figural knowledge.

8.1.3 Procedure

At a first appointment, participants were invited for a group testing session to obtain their intelligence scores. Three to five persons were investigated in one session, tested in the same room and at the same time, to avoid influences of forms on the day or setting differences. First, participants had to handle the CFT-20-R which lasted approximately 30 minutes, and then dealt with the IST, lasting for approximately 40 minutes. Afterwards, participants matching the IQ-score criteria were tested on a second appointment with the computer experiment and finally the MJT.

The remainder of the procedure was identical to Experiment III, except of the pre-test selecting the participants.

8.2 Results

Data analysis was done with SPSS© 17.0 again (SPSS Inc., Chicago, Illinois, USA 1989 - 2009). However, since two extreme groups were tested and only 14 participants remained in the final analysis, non-parametric analyses were conducted, which do not presuppose homogeny dispersion and the like (Janssen & Laatz, 2005). As for the training experiment above, first, the possible differential influences of the two tasks on the two groups were controlled, with Friedman tests for a potential main effect of task and for a potential interaction effect of group by task. A potential main effect of group was calculated with the Mann-Whitney-U-test for independent samples. Further analyses were then conducted for the two tasks separately. Potential group differences (ME group) were also calculated with Mann-Whitney-U-tests for independent samples and Friedman tests for dependent samples were implemented for a potential content main effect. A potential main effect of validity or matching was computed with the Wilcoxon signed rank test. Wilcoxon signed rank tests were also executed for detailed post hoc analyses of the significant main effects, whereby post hoc calculations of detailed group differences still required Mann-Whitney-U-tests. Since non-parametric statistics does not allow a direct calculation of interaction effects, difference values between groups and/or problems/sentences according to their dimensions of 'content' and 'validity/matching' were built and then inserted in further Friedman analyses. Results were again calculated with Wilcoxon signed rank tests. The significance level for comparisons with Friedman tests was set to $p = .05$ (asymptotic significance) with three degrees of freedom (df) and $\chi^2 = 7.815$ as critical value. For Mann-Whitney-U-tests (exact significance) and Wilcoxon signed rank tests (asymptotic significance) critical z-values with 1.96, and accordingly also $p = .05$ were used. The final correlation analysis was based on non-parametric Spearman's rho with a significance level of $p = .05$.

8.2.1 Task by group analysis

Evidently, the participants with superior intelligence made fewer errors (2.46%) than those with average intelligence (18.75%), whereby the deductive reasoning problems were less prone to errors (9.38%) than the recognition items (11.83%). See graphical illustrations in Figures 30 and 31 below. The subsequent Friedman test for a potential task main effect was not significant ($\chi^2 = .333$, $p = .564$), whereas the Mann-Whitney-U-test yielded a main effect of group ($z = -3.148$, $p = .001$). No interaction effect of task by group occurred ($\chi^2 = .667$, $p = .414$).

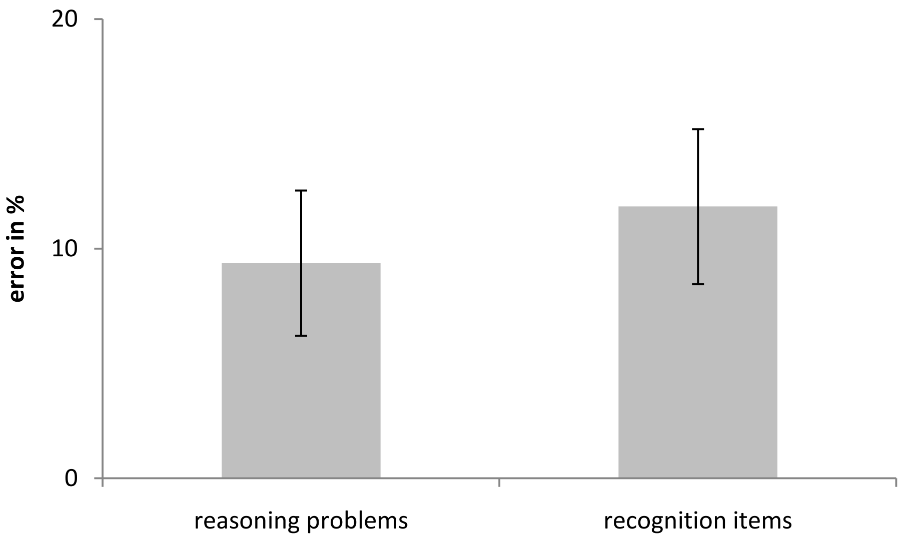


Figure 30: Mean error rates in percent (%) with standard errors for deductive reasoning problems and recognition items.

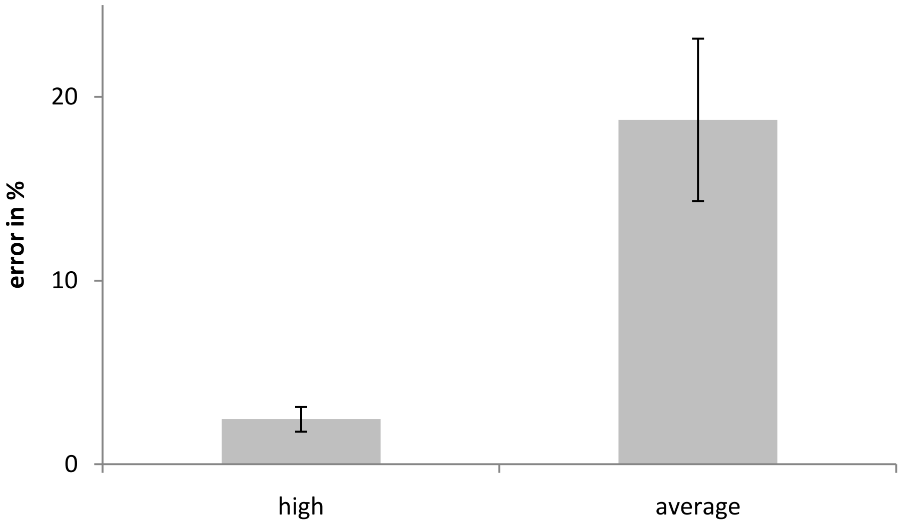


Figure 31: Mean error rates in percent (%) with standard errors for the group with high intelligence and the group with average intelligence.

Concerning decision times, the deductive reasoning problems required more time (1802.52 ms) than the recognition items (1772.45 ms), whereby the group with high intelligence (1644.85 ms) was faster than the group with average intelligence (1930.12 ms). See Figures 32 and 33 below. The calculated Friedman test revealed no main effect of task ($\chi^2 = .000$, $p = 1.000$), and the Mann-Whitney-U-test yielded no main effect of group ($z = -1.597$, $p = .128$). In addition, no interaction effect occurred ($\chi^2 = 3.571$, $p = .059$). Therefore, as in the training experiment, further analyses were conducted for the two tasks in separation.

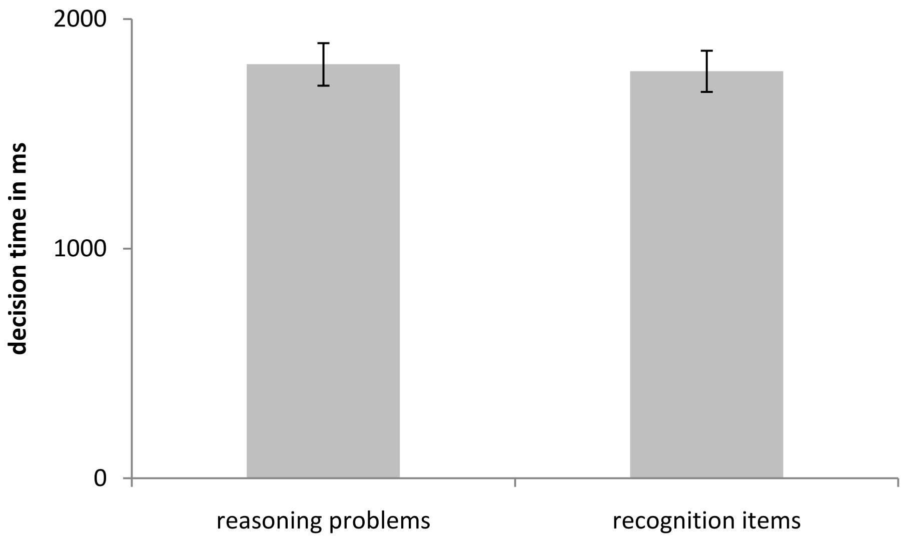


Figure 32: Mean decision times in milliseconds (ms) with standard errors for deductive reasoning problems and recognition items.

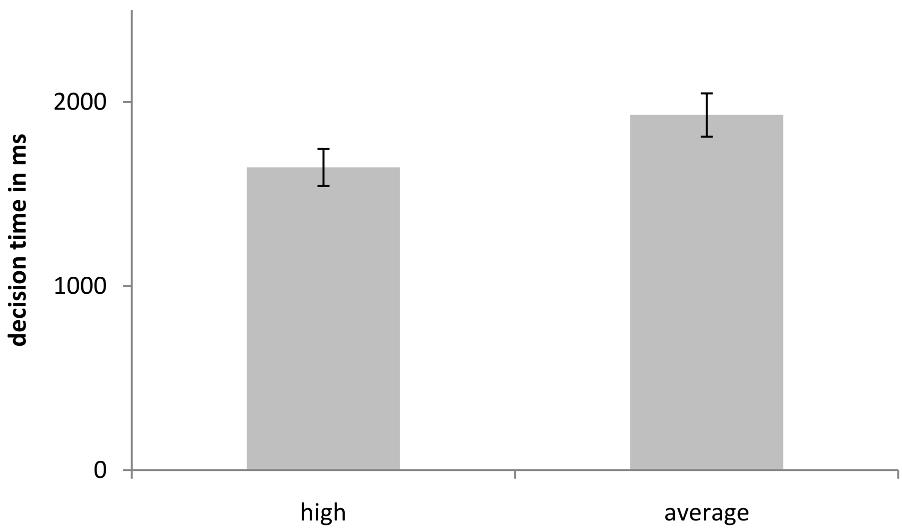


Figure 33: Mean decision times in milliseconds (ms) with standard errors for the group with high intelligence and the group with average intelligence.

8.2.2 MJT

The high intelligence group reached an average c-score in the MJT of 35.91 points (S.D. ±12.76; high, according to standardization; Lind, 2008) whereas the average intelligence group achieved a value of 22.21 points (S.D. ±17.63; average). However, since there was an outlier in the average intelligence group (58 points), the groups did not differ significantly ($z = -1.725$, $p = .097$). Since the outlier did not show any anomalies in the error rates or decision times, it remained in the analyses of the computer experiment, but was excluded for the correlation analysis, because the MJT was also taken into account in this analysis. After exclusion of the outlier in the average intelligence group and the person with the lowest c-score of the high intelligence group (to retain equal group sizes), the groups differed significantly in the MJT (z = -2.562, p = .009). Now, the high intelligent participants had an average c-score of 38.77 points (S.D. ±11.26), and the average intelligent participants reached 16.31 points (S.D. ±9.01).

8.2.3 Deductive reasoning problems

8.2.3.1 Error rates

Participants with high intelligence made almost no errors (1.34%) compared to participants with average intelligence level (17.41%). See Figure 34. Thereby, moral (10.71%) and unmoral (13.39%) inference problems led to higher error rates compared to both of the neutral conditions (neutral1: 8.04%; neutral2: 5.36%).

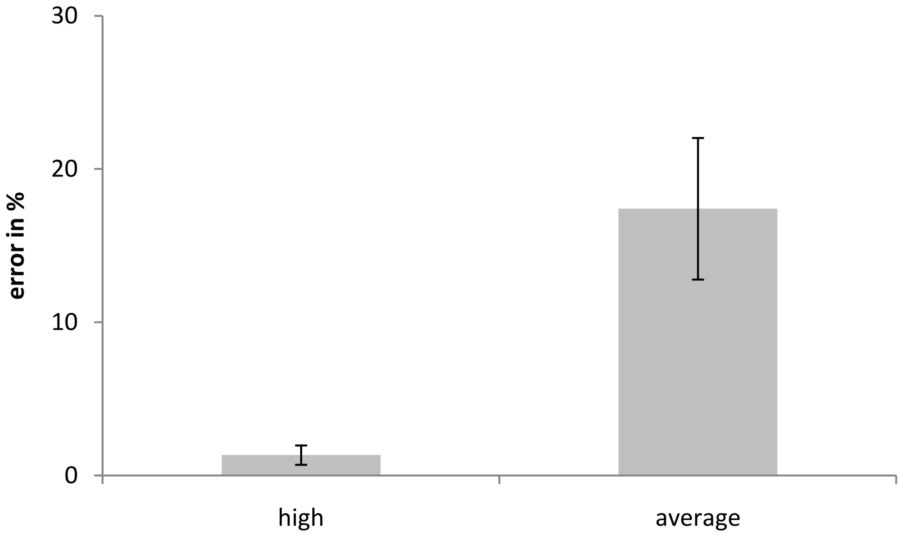


Figure 34: Mean error rates in percent (%) with standard errors for the group with high intelligence and the group with average intelligence.

The statistical analysis revealed a significant main effect of group (z = -2.842, p = .004; Figure 34), whereas no main effect of validity (z = -.741, p = .458) or content (χ^2 = 3.446, p = .328) occurred. The interaction of content and validity yielded significant differences (χ^2 = 8.520, p = .036, Figure 35), while there was no interaction of content and group (χ^2 = 2.143, p = .543) or validity and group (z = -1.029, p = .303).

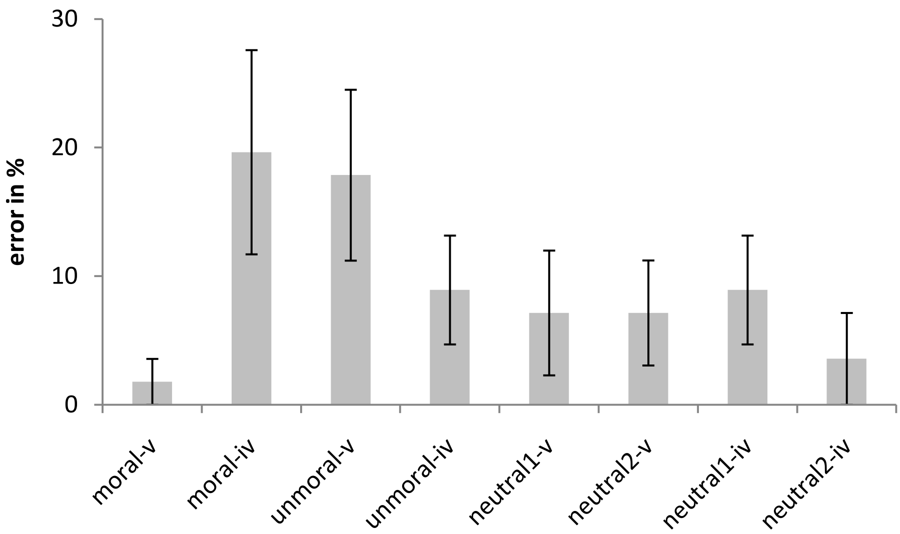


Figure 35: Mean error rates in percent (%) with standard errors for moral, unmoral, neutral1, and neutral2 problems, separated for validity (v = valid, iv = invalid).

Post hoc Wilcoxon signed rank tests for the interaction effect of content and validity are reported in Appendix L (p. 197). The three-way interaction of content, validity, and group also reached significance (χ^2 = 7.981, p = .046). The resulting tests concerning the interaction of content and validity separated for the groups showed no significant effect for the high intelligence participants (χ^2 = 1.000, p = .801), but for the average intelligence participants (χ^2 = 9.000, p = .029). The single contrasts for this group with the effects of content and validity are reported in Appendix L (p. 197). The significant three-way interaction also allowed calculating differences between groups for each single problem category. Thereby, only moral-invalid problems revealed a significant effect (z = -2.992, p = .004) indicating that both groups made more errors in this condition compared to all others.

Regarding only moral-related problems concerning a potential incongruence effect yielded trends for a content main effect (z = -1.774, p = -.076) and an interaction of content and group (z = -1.802, p = .072) as well as a significant group difference (z = -2.695, p = .007). This demonstrates, at least descriptively, that all participants were affected by the incongruent problems compared to the congruent ones, but participants with average intelligence were even more prone to errors than the high intelligent participants (Figure 36).

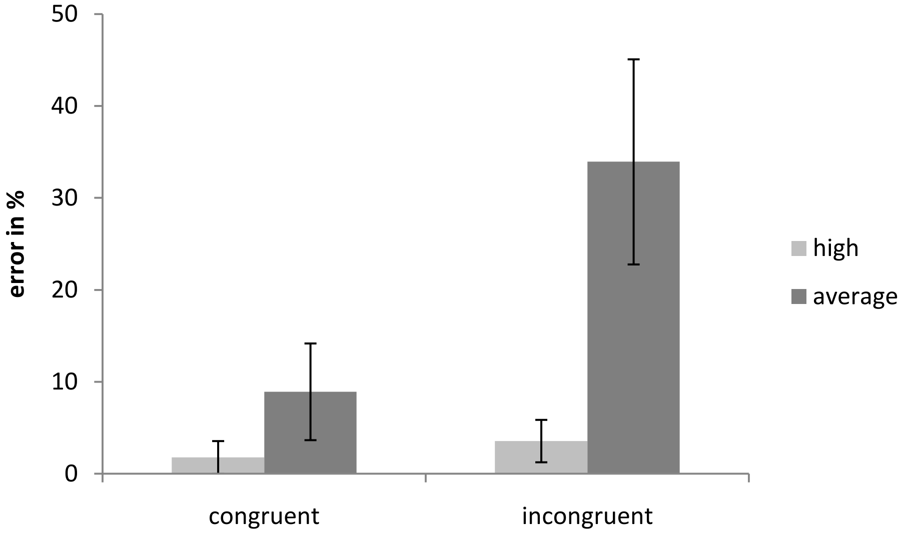


Figure 36: Mean error rates in percent (%) with standard errors for congruent (moral-valid + unmoral-invalid) and incongruent (moral-invalid + unmoral-valid) moral-related problems for the group with high intelligence and the group with average intelligence.

8.2.3.2 Decision times

High intelligent participants answered faster (1614.71 ms) than participants with average intelligence (1990.33 ms). Moral problems had the highest decision times (1912.22 ms), followed by neutral2 (17.87.86 ms) and unmoral ones (1770.57 ms), and finally the neutral1 inference condition (1739.43).

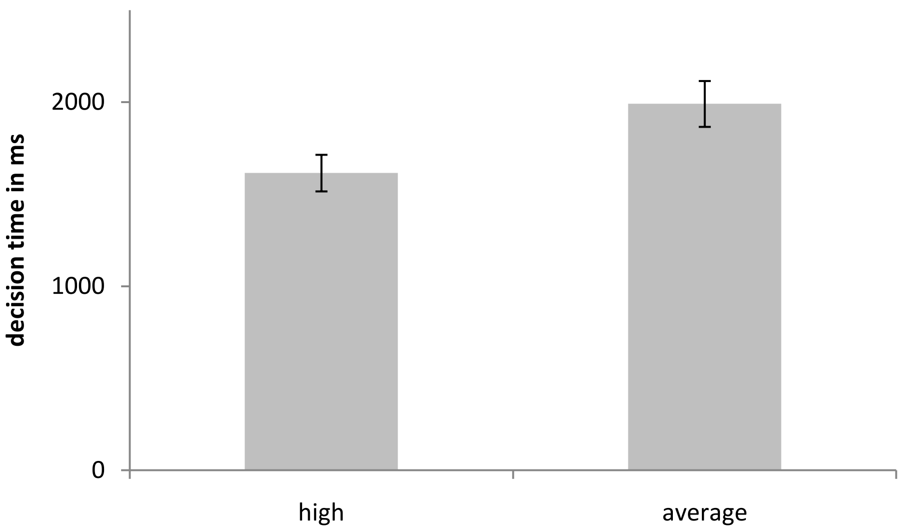


Figure 37: Mean decision times in milliseconds (ms) with standard errors for high and average intelligence group.

The main effect of group showed only a trend (z = -1.853, p = .073) indicating that high intelligent participants were faster than the average intelligence group (Figure 37). The validity main effect did not reach significance (z = -.031, p = .975), whereas content revealed a significant main effect (χ^2 = 10.371, p = .016). Post hoc Wilcoxon signed rank tests then yielded a difference between moral and neutral1 problems (z = -2.229, p = .026) and a trend for the comparison of moral and neutral2 ones (z = -1.726, p = .084) which demonstrates that moral problems had the highest decision times (see Figure 38). The other contrasts were insignificant (moral vs. unmoral: z = -1.601, p = .109; unmoral vs. neutral1: z = -.973, p = .331; unmoral vs. neutral2: z = .408, p = .683; neutral1 vs. neutral2: z = -.659, p = .510). Neither the two-way interactions of content and validity (χ^2 = .943, p = .815), content and group (χ^2 = 2.143, p = .543), and validity and group (z = -1.859, p = .856), nor the three-way interaction of content, validity and group (χ^2 = .771, p = .856) yielded significant effects.

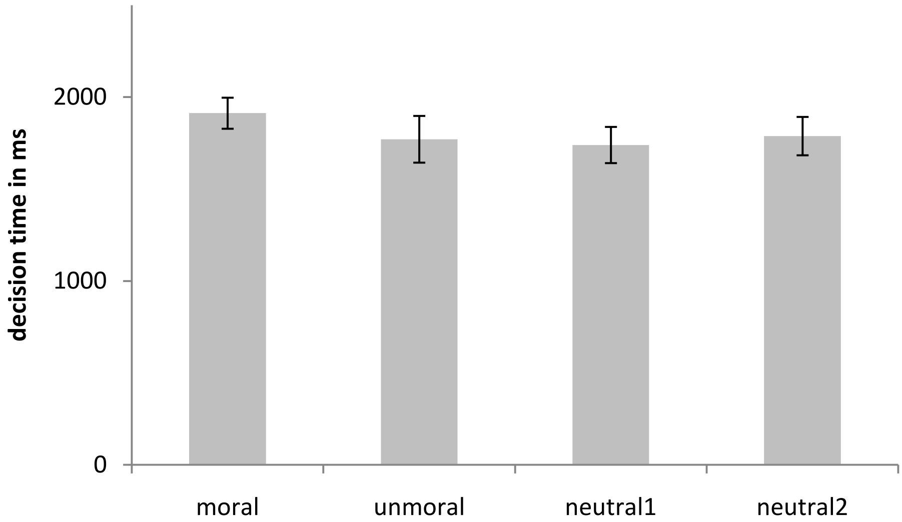


Figure 38: Mean decision times in milliseconds (ms) with standard errors for moral, unmoral, neutral1, and neutral2 problems.

Concerning the moral-related problems, there was a trend for a group difference (z = -1.725, p = .097) illustrating that participants with average intelligence showed the incongruence effect, whereas participants with high intelligence did not (Figure 39).

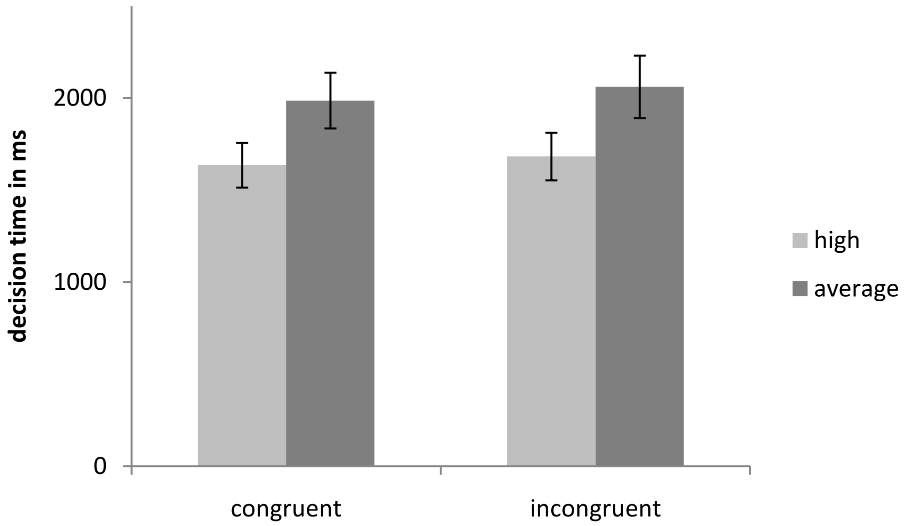


Figure 39: Mean decision times in milliseconds (ms) with standard errors for congruent (moral-valid + unmoral-invalid) and incongruent (moral-invalid + unmoral-valid) moral-related problems for the high and the average intelligence group.

8.2.4 Recognition items

8.2.4.1 Error rates

On the descriptive level participants with high intelligence (3.57 %) were less prone to errors than those with average intelligence (20.09 %). Moral recognition items (10.71 %) resulted in fewer errors than the unmoral ones (14.29 %) and the neutral1 items (14.29 %). The lowest error rate had the neutral2 items (8.04 %).

The Mann-Whitney test revealed a group main effect (z = -3.033, p = .001), but no main effects of matching (Wilcoxon signed rank test: z = -.522, p = .602) or content (Friedman test: χ^2 = -719, p = .869) occurred. Furthermore, there were no significant interaction effects (content by matching: χ^2 =5.300, p = .151; content by group: χ^2 = 3.141, p = .370; group by matching: z = -1.706, p = .088; group by content by matching: χ^2 = 4.918, p = .178). The group difference is shown in Figure 40 below.

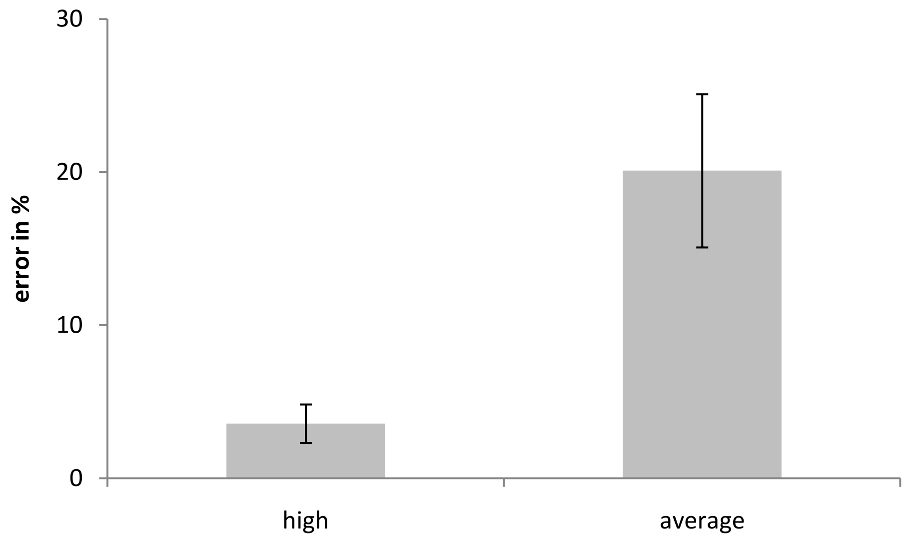


Figure 40: Mean error rates in percent (%) with standard errors for the group with high intelligence and the group with average intelligence.

8.2.4.2 Decision times

Descriptively, the participants with high intelligence (1674.99 ms) were faster than those with average intelligence (1869.91 ms). Thereby, moral (1893.90 ms) and unmoral items (1838.43 ms) required more decision time than the neutral1 (1620.64 ms) and neutral2 ones (1736.84 ms). In particular, the literal match recognition items were answered faster (1623.74 ms) than the ones with no literal match (1921.16 ms).

Statistical analysis yielded a main effect of content (χ^2 = 13.457, p = .004) and matching (z = -3.170, p = .002, see Figure 41), but no group effect (z = -.575, p = .620). No interaction effects occurred (content by group: χ^2 = 5.914, p = .116; content by matching: χ^2 = .943, p = .815; group by matching: z = -.676, p = .499; content by group by matching: χ^2 = 4.543, p = .208).

Post hoc analysis of the content main effect showed that moral recognition items differed from the neutral1 (z = -2.982, p = .003) and neutral2 ones (z = -2.166, p = .030), as did the unmoral items from the neutral1 (z = -2.542, p = .011) and the neutral2 ones (z = -2.229, p = .026). Neither the moral-related recognition items (z = -1.161, p = .245), nor the neutral ones (z = -1.789, p = .074) differed from each other (see Figure 42).

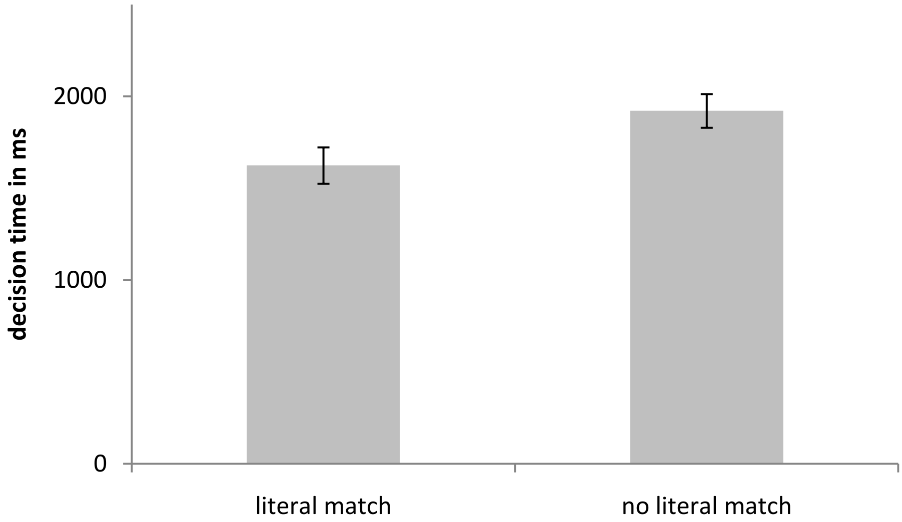


Figure 41: Mean decision times in milliseconds (ms) with standard errors for the literal match and the no literal match recognition items.

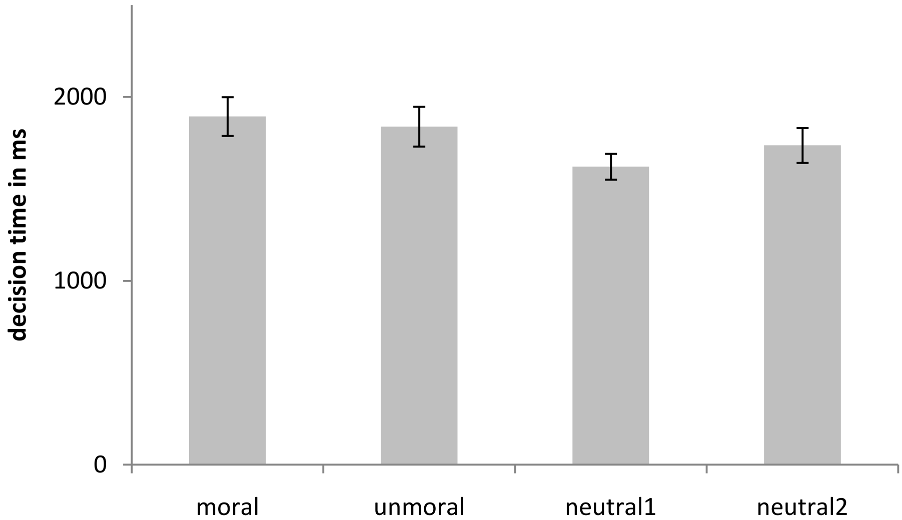


Figure 42: Mean decision times in milliseconds (ms) with standard errors for moral, unmoral, neutral1, and neutral2 recognition items.

8.2.5 Correlation

Finally, a correlation analysis was carried out including decision times and error rates for all reasoning problems as well as for the moral-related and neutral ones, together with both intelligence scores (CFT/IST), and the c-score of the MJT, looking for differential aspects of fluid versus crystallized intelligence components. Significant correlations were obtained from IST and CFT (r_s = .863), IST and c-Score (r_s = .621), CFT and c-score (r_s = .827). Error rates of moral-related problems correlated with the c-score (r_s = -.689), with the CFT (r_s = -.794), and with the IST (r_s = -.653). See Appendix L (pp. 198-199) for statistical details and Figures 43 and 44 for graphical illustration.

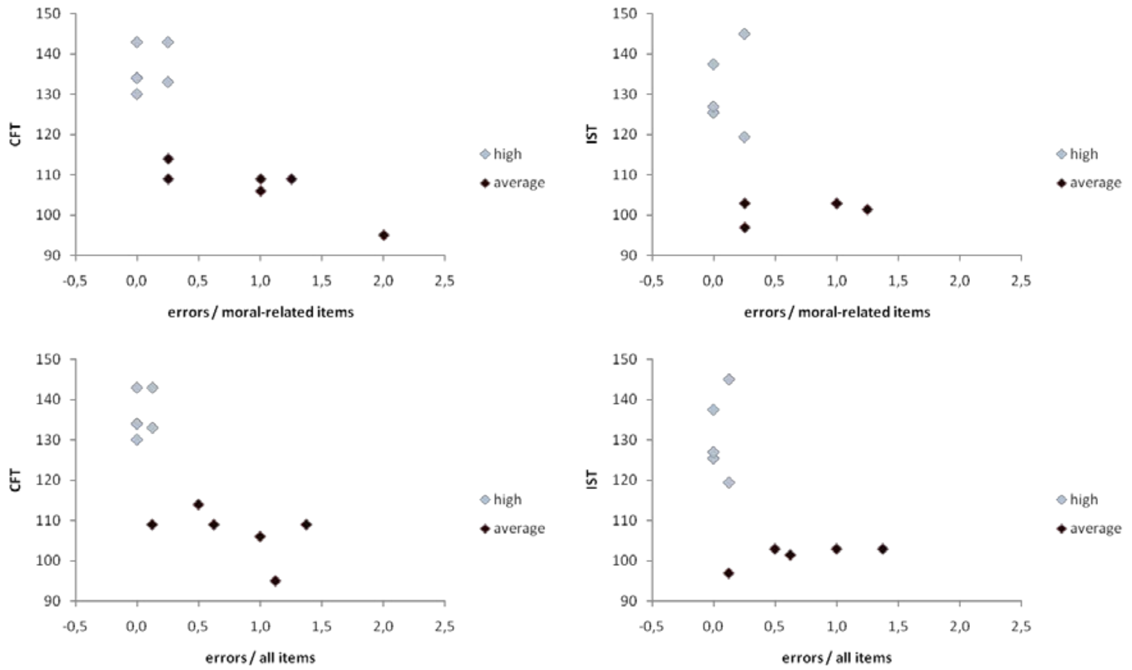


Figure 43: Scatter plots of mean CFT-20-R and IST-2000-R IQ-values of 6 participants with mean error rates of all problems or moral-related problems.

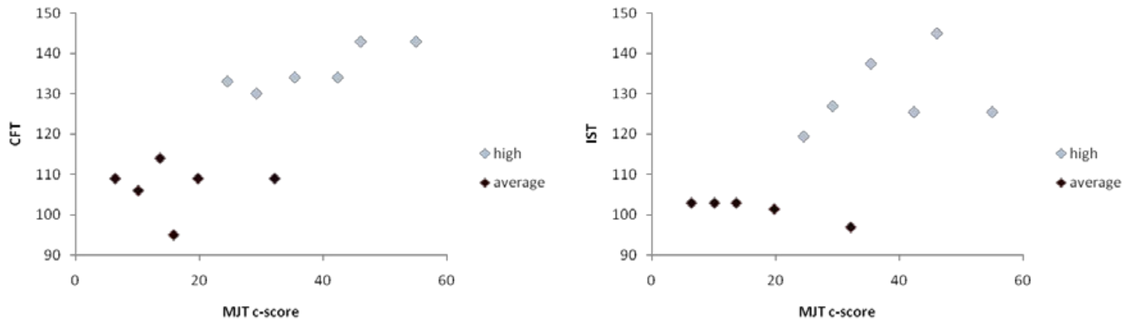


Figure 44: Scatter plots of mean CFT-20-R and IST-2000-R IQ-values of 6 participants with mean c-scores of MJT.

Scatter plots with regression lines comparing IST, CFT, and c-score values of 6 participants with mean error rates of moral-related as well as error rates of all problems are shown in Appendix L (pp. 200-201). Furthermore, scatter plots with regression lines for IST and CFT, c-score and IST, c-score and CFT, and c-score with moral-related errors are reported in Appendix L (pp. 200-201).

8.3 Discussion

The fourth experiment revealed group differences between participants with high and average IQ values. Participants with superior intelligence had higher c-scores, namely higher moral judgment competence, shorter decision times and made fewer errors (although decision time differences did not reach significance). They showed no incongruity effect in the error rates of the moral-related problems and interestingly, there was also a general effect of intelligence as shown in the correlations of the CFT with all error rates and decision times of all deductive reasoning problems (moral-related and neutral problems). The IST in contrast was almost exclusively correlated with the error rates of the moral-related problems and MJT c-score.

The results indicate that intelligence has indeed influences on performance in deductive reasoning problems (Johnson-Laird, 2006). Fluid intelligence in particular, seems to be responsible for these overall better reasoning abilities. Fluid intelligence in turn is associated with working memory capacity (Engle, 2002). Both reflect "[…] the ability to keep a representation active, particularly in the face of interference and distraction." (Engle, Tuholski, Laughlin, & Conway, 1999, p. 309). Thereby, "WM capacity, or executive attention, is most important under conditions in which interference leads to retrieval of response tendencies that conflict with the current task." (Engle, 2002, p. 19). Moreover, there is also a connection between speed of task processing and intelligence (Fry & Hale, 2000), although there was only a trend in decision times that high intelligent participants answered faster than participants with average intelligence. The reason why this difference did not reach significance might be due to the fact that the high IQ group was significantly older than the group with average intelligence ($z = -2.567$, $p = .007$), since it is known that decision times slow down with age (Der & Deary, 2006; Tun & Lachman, 2008). Another explanation might be that high intelligent participants took more time to perform better as was reflected in the lower error rates. In relation to these group differences, one might wonder that there were no content effects in the error rates but only in the decision times. A potential content effect might have vanished since the high intelligent participants made almost no errors at all. This appears to have reduced the overall error rates for both groups, although the group with average intelligence made almost twice as much errors in the moral-related condition as compared to the neutral one. The decision times in turn revealed a significant main effect that could be related to the additional time the high intelligent participants needed for the moral-related problems to avoid errors. Thus, both groups needed more time to process the moral-related problems, but that resulted only in the group of superior intelligence in lower error rates. This additional time the later group applied to the moral-related problems might be due to the emotional load of these problems (see also evaluation of Pre-Study III). These emotional laden contents might have activated personal attitudes (Lefford, 1946), which had to be processed additionally. It can be suggested that these attitudes are related to prior knowledge about morals, thus reflect moral attitudes (cp., Luo et al., 2006). The correlations of the crystallized intelligence components particularly to the error rates of the reasoning problems and the c-score of the MJT (Lind, 2008) seem to further support these assumptions. Thus, all participants seemed to reason practically in the moral-related conditions.

It might be concluded that intelligence influences both theoretical and practical reasoning. This influence seems to be mediated by fluid and crystallized intelligence components. Fluid intelligence is more related to working memory capacities (i.e., rational processes), whereas crystallized intelligence represents emotionally laden prior knowledge (i.e., emotional/intuitive processes) or is related to moral attitudes based on this emotionally laden knowledge. Perhaps, the WM capacities helped suppressing prior knowledge / moral attitudes to focus on task demands more easily, or 'just' enabled appropriate task processing in general.

With regards to the recognition items, there was only a main effect of group for the error rates reflecting again the overall better performance of the high intelligent participants. With respect to the content, both groups were slower in the moral-related condition and no differences concerning error rates could be found. In contrast to the postulated effects and the evidences seen in the training experiment reported earlier, the moral-related content did not improve recognition performance. However, taking the given explanation of moral attitudes and upcoming prior knowledge into account as well as working memory capacities might give a hint to understand these results. The overall performance of the average intelligence group was worse compared to that of the superior intelligence group, independent from moral-related or neutral contents, and also independent from the specific task at hand. In contrast, the performance of the high intelligent participants was almost error free independent from contents and tasks, but they needed in both tasks more time to handle the moral-related problems/sentences in order to keep error rates low.

Besides these interesting results and evidences, a few limiting factors of this experiment have to be mentioned. The first factor is the small number of participants. This reduced statistical power extremely, thus rendering potential effects not significant. Furthermore, the correlation analysis reveals insights in potential relations between intelligence and reasoning performance, but does not allow assumptions about causal relationships. In addition, the groups differed with regards to age and gender as well as concerning potential prior knowledge of intelligence tests and presumably logical reasoning in general (although explicit knowledge of deductive reasoning tasks could be excluded). They were also intentionally chosen to differ significantly in intelligence level. A further experiment and correlation seems necessary however, in order to compare many more persons with varying intelligence and performance levels to obtain evidence whether performance in theoretical and practical reasoning as well as moral judgment competence rises with increasing IQ or not. These critical points constrain potential generalizations of the current findings. Therefore, the results are impressive, but have to be interpreted with caution.

An even better indicator of the reciprocal involvement of rational and emotional/intuitive processes on theoretical and practical reasoning can be obtained by measuring participants' brain activations. Thus, the last experiment dealt with the brain processes and associated cognitive processes which participants show while they handle theoretical and practical reasoning problems.

9. Experiment V – fMRI

This last experiment investigated 'normal' participants dealing with the deductive inference problems used in Experiment II while measuring their brain activity with event-related functional magnetic resonance imaging (fMRI). This experiment should yield further evidence of potential similarities or even dissimilarities in cognitive processes and associated brain areas for theoretical and practical reasoning.

In order to determine similarities and dissimilarities in the functional neuroanatomy of reasoning and its associated processes with moral content (i.e., 'practical reasoning') and with abstract content (i.e., 'theoretical reasoning') Regions of Interest (ROIs) were defined for the brain imaging data analysis according to the main findings of the above mentioned studies. It was only focused on the main structures which revealed to be involved in theoretical and practical reasoning. However, regarding the differences between studies in labeling the frontal brain areas, only the DLPFC as region rather associated with rational processes (e.g., Greene et al., 2004; Schaich Borg et al., 2006), the OFC as region rather associated with emotional/intuitive processes (e.g., Dolan, 2002; Prehn et al., 2008), and the mPFC as region rather associated with integration of both of these processes (especially the ventral parts of the mPFC; e.g., Adolphs, 1999, 2006; Damasio, 2006) were taken into account. Apart from the frontal regions, further brain structures and areas were proposed and investigated: the anterior cingulate cortex (ACC) as region rather associated with working memory processes like conflict detection and monitoring (e.g., Moll et al., 2003; Young & Koenigs, 2007), the superior temporal sulcus (STS) as region associated with Theory of Mind (ToM) processes and sensory integration (e.g., Frith & Frith, 1999; Heekeren et al., 2003; Luo et al., 2006), temporal areas associated with emotional, intuitive, and knowledge-based processes (e.g., Goel & Dolan, 2003b; Moll et al., 2005), and the parietal cortex associated with rather rational and working memory processes or reasoning with and without content (e.g., Canessa et al., 2005; Goel & Dolan, 2003b; Greene et al., 2001).

The assumption was to obtain increased activations in the following brain regions for reasoning with (moral) and without (abstract) content: the anterior cingulate cortex (ACC), the dorsolateral prefrontal cortex (DLPFC), the medial prefrontal cortex (mPFC), the orbitofrontal cortex (OFC), the parietal lobe, the precuneus, the STS-region, and the temporal pole, since these structures have consistently been identified across various studies. The concrete hypotheses related to this assumption are:

- Theoretical and practical reasoning are at least partly based on activations of similar brain areas involving frontal, temporal, and parietal regions (see concrete ROIs above).
- These regions in turn are associated with rational and emotional/intuitive processes as well as with working memory~, knowledge~, and ToM related processes.

The STS-region ROI (in the following, short: STS) included the superior temporal gyrus, the angular gyrus, and the supramarginal gyrus.[16] The parietal lobe ROI (in the following, short: parietal) included the superior and inferior parietal lobe as well as the angular and supramarginal gyrus. The occipital brain regions were neglected, since increased activations due to the visual presentation of the reasoning problems were expected there anyway. The positive and negative emotional content problems as well as the neutral content ones served as controls identifying regions only dealing with emotional or 'simple' content effects.

Note that the main focus of this experiment was on overlapping brain regions, not single x,y,z-coordinates, and thus on the main involved cognitive processes (rational and emotional/intuitive). Further sub-divisions of brain structures/regions and further associated sub-processes are necessary, but beyond the scope of the current experiment. Details concerning methods are reported in the following section.

9.1 Method

9.1.1 Participants

In this last experiment, 31 healthy participants meeting the inclusion criteria (no previous mental illness, no psychotropic medication, and medical suitability for fMRI) volunteered in the study and gave informed written consent according to the Declaration of Helsinki (1964/2008). The 17 female participants had a mean age of 21.76 years (S.D. ±3.13), and the 14 male participants had a mean age of 25.36 years (S.D. ±6.52). The experiment was approved by the ethics committee of the German Psychological Society (DGPs). Participants came from local universities. For the approximately 45 minute's experimental session, the participants were financially compensated or received course credits. Again, all participants were native German speakers, naïve to logical reasoning tasks, right-handed according to Salmaso and Longoni's (1985) modified version of the Edinburgh Handedness Questionnaire (Oldfield, 1971), and had normal or corrected-to-normal visual acuity.

---

[16] Since the STS is sometimes also labeled as temporal parietal junction (TPJ, e.g., Young, Cushman, Hauser, & Saxe, 2007) and its real dimension is unspecific, the inferior parietal lobe was also included in the ROI for the STS region to cover all possible activations associated with this area.

9.1.2 Material

The deductive inference problems applied here were the same as in Experiment II. Thus, a total of 48 problems were presented in the scanner, representing a 6 (content) × 2 (validity) within-subject factorial design.

9.1.3 Procedure

Prior to the brain imaging experiment, participants read the instruction and practiced with three examples not used in the experiment outside the scanner on a standard personal computer. Subsequently, the participants were put into the scanner, and ran through the technical measurements (field mappings, anatomical scan) before engaging in the main experiment. Problems were randomly presented via an LCD projector onto a screen that could be viewed through a mirror mounted on the head coil (visual field = 18°). The sentences were presented successively, whereby the first premise lasted 5 s, the second 4 s, followed by a 5 s pause with a ±1.125 s jitter, and finally, the conclusion lasted for 14 s. The third sentence (presented in red) had to be judged as valid or invalid using a response box which was positioned on the right side of the participants'. The participants used their index and middle fingers to answer via button press on the response box within this 14 s presentation of the conclusion. This time window for the conclusion was also jittered with ±1.125 s before the next problem appeared (Figure 45). The jitters were implemented to avoid anticipation effects of the participants in order to control possible sustained steady states of the BOLD signal as well as to avoid measuring always the same slices at the same time. Each trial lasted approximately 33 s. The software Presentation 9.0 (©Neurobehavioral Systems Inc., Albany, USA) was used for stimulus presentation as well as for the recording of participants' responses and decision times.

Figure 45: Procedure of one trial of the imaging experiment.

9.1.4 Analysis of behavioral data

Behavioral data concerning error rates (almost no missing values occurred, thus correct answers correspond to the difference of 100% minus error rates) and decision times were analyzed by setting up a general linear model for repeated measures and post hoc paired-sample t-tests for the 6 × 2 factors (content × validity). Analysis were done with SPSS© 17.0

(SPSS Inc., Chicago, Illinois, USA 1989 - 2009) and significance level was set to $p = .05$ (F-values of the GLM were Greenhouse-Geisser corrected).

9.1.5 Image acquisition and analysis

Functional and anatomical brain images were obtained using a 1.5 T whole-body tomograph (Siemens Magnetom Symphony, Erlangen, Germany) with a standard head coil. Structural image acquisition consisted of 160 T1-weighted sagittal images (MPRage, 1 mm slice thickness). Field mapping images were measured in the same space and resolution as the functional images. For the functional imaging a total of 684 volumes were registered using a standard T2*-weighted gradient echo-planar imaging sequence (EPI) with 25 slices covering the whole brain (slice thickness = 5 mm, 1 mm gap, descending acquisition, TE = 55 ms, TR = 2.5 sec., flip angle = 90°, field of view = 192 mm × 192 mm, matrix size = 64 × 64). The orientation of the axial slices was parallel to the AC-PC line.

The first six volumes were discarded to control for saturation effects. Preprocessing and data analyses were done with the statistical parametric mapping software package (©SPM5, Wellcome Department of Cognitive Neurology, London, UK) using Matlab 7.6 (©Mathworks Inc., Natick, Massachusetts, USA). After converting scanner DICOM images with MRIConvert 2.0 (©Lewis Center for Neuroimaging, University of Oregon, USA), preprocessing was carried out. The measured fieldmaps were used in the realignment and unwarping steps, followed by a slice time correction and the normalization steps, where images were matched to the standard ICBM/MNI (Montreal Neurological Institute) template. Finally, the images were smoothed using an isotropic three-dimensional Gaussian filter with a full width at half maximum of 9 mm.

For each participant, the experimental conditions were modeled by a stick function convolved with a canonical hemodynamic response function (hrf) using a general linear model. This function was applied to the 6 (content) by 2 (validity) factors building the factorial design at the single subject level. The two types of premises and the six movement parameters of the rigid body transformation, applied by the realignment procedure, were introduced as covariates of no interest in the model. The time series were filtered with a high pass filter of 128 s and serial correlations were controlled by an AR(1) process. T-contrasts were set for all 12 conditions (content by validity), which were then analyzed at the group level.

At the group level, a factorial within-subject ANOVA design was built to model the 12 conditions. Comparisons of conditions were done by calculating appropriate contrasts.

Conjunction analyses were conducted according to the 'minimum statistics approach' (Nichols, Brett, Andersson, Wager, & Poline, 2005). The resulting voxel-wise tests were corrected for multiple comparisons for the predefined ROIs as well as for the explorative whole brain analysis. Results were considered significant at $p < .05$ ($p < .025$ when the ROIs were split in left and right hemisphere). ROI creation was done with MARINA (Walter, Blecker, Kirsch, Sammer, Schienle, Stark, & Vaitl, 2003).

Labeling for the brain regions activated in the F-test were done with MARINA (Walter et al., 2003) and its included anatomical tool (Tzourio-Mazoyer, Landeau, Papathanassiou, Crivello, Etard, Delcroix, Mazoyer, & Joliot, 2002). To obtain the Brodmann labels, SPM5 MNI/ICBM-based x, y, z data were transformed into Talairach-space with GingerALE 2.0 implemented in BRAINMAP (Laird, Fox, Price, Glahn, Uecker, Lancaster, Turkeltaub, Kochunov, & Fox, 2005), whereby the newest ALE-algorithm was used (Eickhoff, Laird, Grefkes, Wang, Zilles, & Fox, 2009). Coordinates were then defined with the TALAIRACH DAEMON (Lancaster, Woldorff, Parsons, Liotti, Freitas, Rainey, Kochunov, Nickerson, Mikiten, & Fox, 2000). Reported Brodmann Areas should be handled carefully (Talairach & Tournoux, 1988) and are only mentioned as supplemental information. Finally, MRIcro (Rorden & Brett, 2000) was used to create figures of activated ROI contrasts for better illustration.

9.2 Results

9.2.1 Behavioral data

9.2.1.1 Error rates

On the descriptive level participants made 18.15% errors with the moral problems, 16.53% with the unmoral ones, 7.26% with the abstract problems, 5.24% with the neutral ones as well as 8.06%, and 13.71% with positive and negative emotional problems, respectively (Figure 46, left). The general linear model revealed a significant main effect of content (ME content: $F(2.556, 76.695) = 7.612$, $p < .001$) and an interaction effect of content and validity (IE: $F(1.926, 57.772) = 7.032$, p .002) for the error rates. No main effect of validity occurred (ME validity: $F(1, 30) = .005$, $p = .946$).

The corresponding post hoc t-tests of the content effect (Appendix M, p. 202) showed that moral problems (vs. neutral: $t(30) = 3.906$, $p < .001$; vs. abstract: $t(30) = 2.979$, $p = .006$; vs. positive emotional: $t(30) = 2.843$, $p = .008$) and unmoral problems (vs. neutral: $t(30) = 4.119$, $p < .001$; vs. abstract: $t(30) = 2.923$, $p = .007$; vs. positive emotional: $t(30) = 2.695$, $p = .011$) were more prone to errors than all other conditions (except the negative emotional control problems, where analysis did not reach significance; moral vs. negative emotional: t(30) = 1.283, p = 209; unmoral vs. negative emotional: t(30) = 1.022, p = .315). In contrast, the abstract problems revealed the lowest error rates as compared to all others (except for the neutral problems, t(30) = -1.541, p = .134), whereby the comparison with the positive emotional problems was insignificant this time (vs. positive emotional: t(30) = -.441, p = 662; vs. negative emotional: t(30) = -2.794, p = .009). Post hoc t-tests for the significant interaction are only shown in Appendix M (p. 203), since the analysis of the brain imaging data was restricted to the content categories.

Analyzing only the moral-related problems concerning their moral status (moral, unmoral) and their validity (valid, invalid) yielded an incongruence effect. The effect showed that participants made more errors with the incongruent (moral-invalid/unmoral-valid) problems as compared to the congruent ones (moral-valid/unmoral-invalid) (F(1, 30) = 9.990, p = .004; t(30) = -3.161, p = .004).

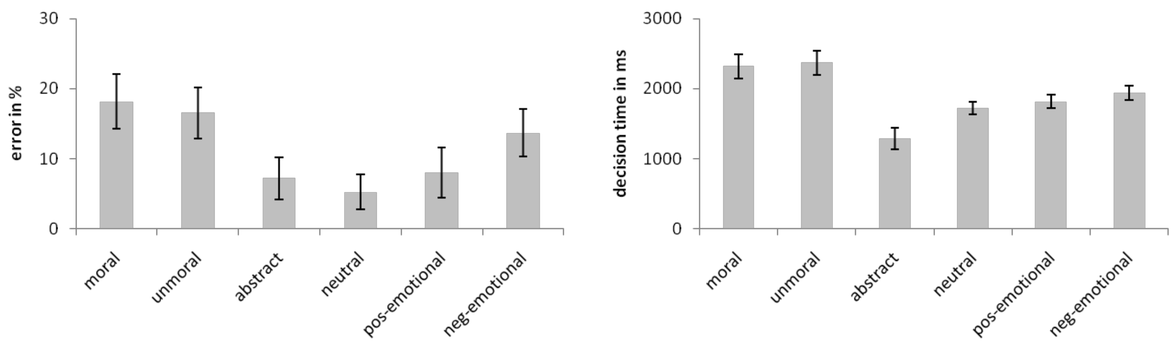


Figure 46: (left) mean error rates in percent (%) with standard error for moral, unmoral, neutral, abstract, positive emotional, and negative emotional problems; (right) mean decision times in milliseconds (ms) with standard error for the same conditions.

9.2.1.2 Decision times

Descriptively, participants showed increased decision times for moral (2321 ms) and unmoral problems (2372 ms) as compared to abstract (1290 ms), neutral (1723 ms), positive emotional (1816 ms), and negative emotional (1939 ms) problems (Figure 46, right). Decision times yielded a significant main effect for validity (F(1, 30) = 29.808, p < .001) and content (F(3.141, 94.226) = 22.496, p < .001), but no interaction (F(2.169, 65.066) = .662, p = .531) in the general linear model. Post hoc t-tests confirmed that invalid problems showed higher decision times than valid ones (t(30) = -5.460, p < .001). The same accounts for moral problems (vs. neutral: t(30) = 4.445, p < .001; vs. abstract: t(30) = 6.900, p < .001, vs. positive-emotional: t(30) = 3.765, p = .001; vs. negative-emotional: t(30) = 3.072, p = .004) as well as unmoral ones (vs. neutral: t(30) = 4.882, p < .001; vs. abstract: t(30) = 8.058, p <

.001; vs. positive-emotional: t(30) = 4.262, p < .011; vs. negative-emotional: t(30) = 3.103, p = .004) in comparison to all the others. The abstract problems in turn differed from all others resulting in the shortest decision times (vs. neutral: t(30) = 3.612, p = .001; vs. positive emotional: t(30) = -5.417, p < .001; vs. negative emotional: t(30) = -4.780, p < .001). All results are reported in Appendix M (p. 202). Decision times showed no incongruence effect of the moral-related problems (F(1, 30) = .287, p = .596).

9.2.2 Imaging data

Since the neural similarities and dissimilarities of theoretical (i.e., deductive reasoning) and practical reasoning (i.e., moral reasoning), as well as in the associated processes (emotional/intuitive and/or rational) were of special interest, the analysis included four steps. First, the effects of single conditions (moral, unmoral, neutral, abstract, positive emotional, and negative emotional) were analyzed, regardless of validity. Subsequently, the focus was put only on conditions showing any corresponding neural activation. Second, a conjunction analysis for these conditions was done. Third, differences between these conditions were calculated. Finally, a whole-brain explorative F-test was conducted in order to test for further activations in structures outside the pre-defined ROIs.

9.2.2.1 Results of the Region of Interest (ROI) analyses

Analyzing the activation during the six content conditions revealed significant effects for moral, abstract, and neutral problems, but not for unmoral and both of the emotional ones in all predefined ROIs (see table in Appendix N, p. 204, and Figure 47 below for details).

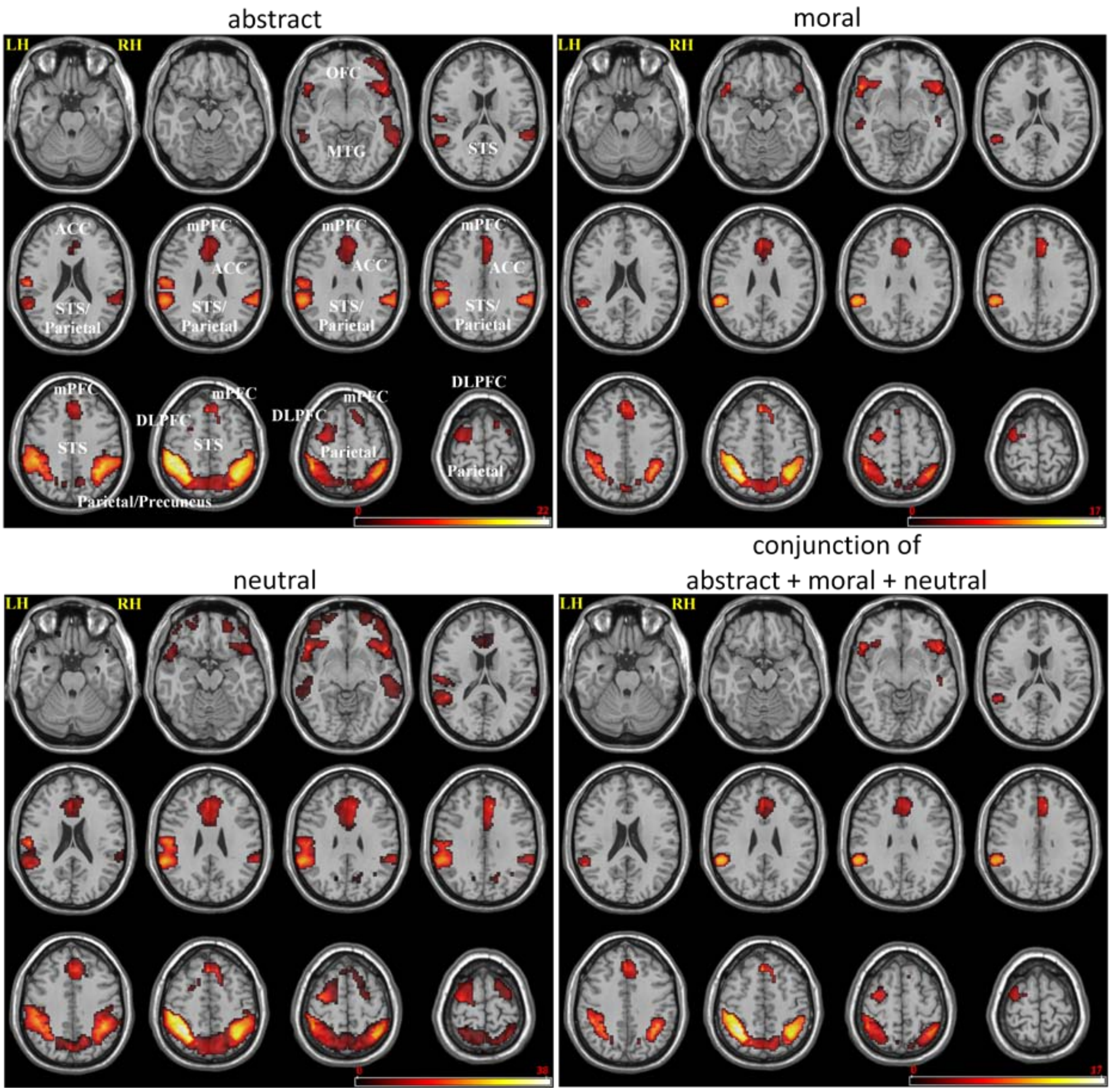


Figure 47: Significant activations of calculated ROIs (voxel-based, FEW-corrected with p = .025) for single conditions (abstract, moral, neutral); LH = left hemisphere, RH = right hemisphere, each figure contains 12 slices with activation density color bar; slices in each figure correspond to the slice lines shown in the sagittal image on the right bottom image shown in Figure 48 below. Activated brain regions are: OFC = orbitofrontal cortex, MTG = middle temporal gyrus, STS = superior temporal sulcus region, ACC = anterior cingulate cortex, mPFC = medial prefrontal cortex, parietal cortex, precuneus, and DLPFC = dorsolateral prefrontal cortex, (for better identification of activations, only the first image includes labeling of activated brain areas).

Next, the common activations in abstract, moral, and neutral problems were tested. Again, all predefined ROIs revealed significant bilateral activations (see table in Appendix N, p. 205, and Figure 47 above for details). Note that pair-wise conjunctions are not necessary since the conjunction of the three conditions also encompasses the results of possible pair-wise conjunctions.

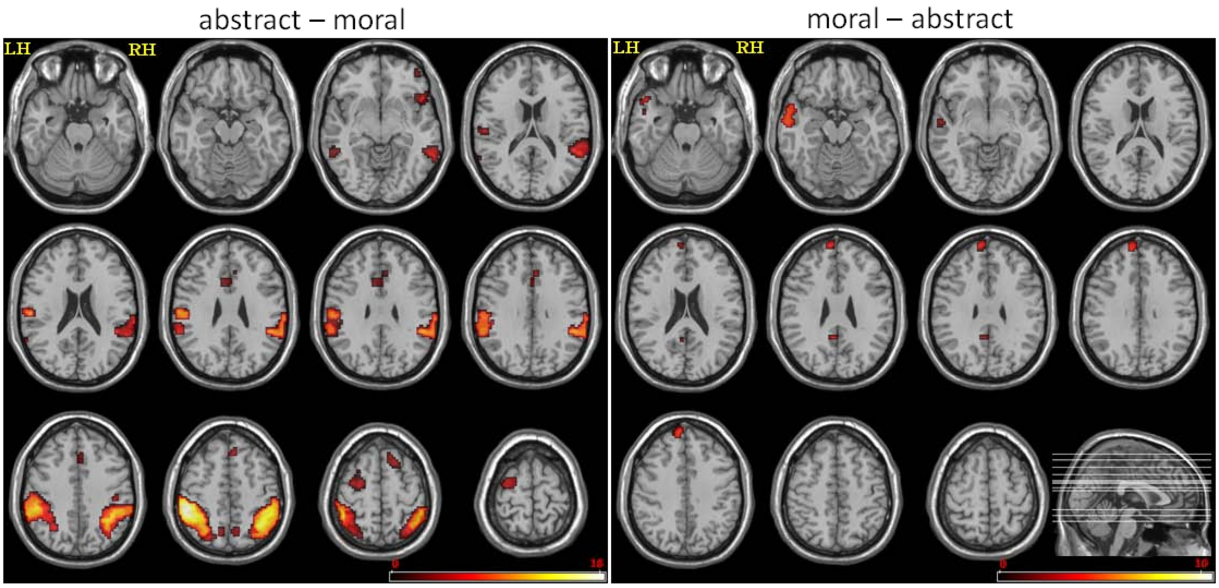


Figure 48: Significant activations of calculated ROIs (voxel-based, FEW-corrected with p = .025) for difference comparisons between abstract and moral problems as well as moral and abstract problems; LH = left hemisphere, RH = right hemisphere, each figure contains 12 slices with activation density color bar; slices in each figure correspond to the slice lines shown in the sagittal image on the right bottom image. Activated brain regions are labeled in Figure 47 above.

In the third analysis, the differences between the conditions were analyzed. Contrasting abstract problems with the moral ones yielded increased bilateral activations in the ACC, the DLPFC, the mPFC, the MTG, the parietal lobe, the Precuneus, the STS, the Temporal Pole, and the right OFC. The inverse contrast (moral – abstract) showed increased left-hemispheric activations in the DLPFC, the mPFC, the MTG, the precuneus, the STS, and the temporal pole (see both comparisons in Figure 48 above). Comparing the abstract with neutral problems again resulted in increased activations in all ROIs bilaterally, except for the OFC and the temporal pole, which only showed right hemispheric activations. Conversely, neutral problems minus abstract ones revealed increased activations in the left temporal pole. Subsequently, the contrast of the moral problems with the neutral ones identified increased bilateral activations in the ACC and the mPFC as well as left hemispheric activations in the DLPFC, the MTG, the OFC, parietal, and the STS. The inverse comparison yielded no activations at all (see all comparisons in Appendix N, pp. 206-207).

9.2.2.2 Results of the explorative whole-brain analysis

Finally, an explorative whole-brain analysis (F-test) was conducted to test, whether further regions are activated during the different reasoning processes. The activated regions matched almost all of the predefined ROIs except for an inferior and medial occipital lobe activation (F-value: 13.21; cluster-size: 5; x y z coordinates: -15 -93 -3; BA 17). Therefore, no further analyses were conducted.

9.3 Discussion

9.3.1 Behavioral Results

The results of the behavioral data showed that participants again produced more errors and needed more time in answering the moral-related problems as compared to all others. In contrast, the abstract problems were almost comparable to the neutral problems and yielded fewer errors and shorter decision times than the emotional controls or the moral-related problems. Thus, as shown in Figure 46, the content per se affected the reasoning performance, but it was especially heightened for the moral-related problems. This content effect normally facilitating deductive reasoning (e.g., Wason & Shapiro, 1971; Johnson-Laird et al., 1972), disturbed the reasoning performance in the current experiment, as it was already shown in Experiment I and II. Thus, the fMRI experiment also yielded the moral-related, problem inherent incongruence effect. In turn, this incongruence effect, which the participants revealed also in this experiment, again supports assumptions of knowledge about morals or personal moral attitudes due to learning involved in practical reasoning (e.g., Luo et al., 2006; Moll et al., 2005).

However, this moral-related content effect could not explain the increased error rates of the negative emotional and, to a limited amount, the positive emotional problems. These increments might rather have resulted from the emotional content per se, since Blanchette and Richards (2004) as well as Blanchette (2006) had shown that emotional content (positive and negative) impeded the reasoning process. Another possible explanation for the incongruity effect might also be related to the emotional load of the moral-related problems. If morals are learned and stored, at least partly, as 'somatic markers' (Damasio, 2006), dealing with morality seems to activate these markers again. Therefore, the incongruence effect could also reflect the emotional load of the moral-related problems, since emotional laden knowledge is remembered better (e.g., Ferré, 2003). Perhaps, the emotional load is responsible for the performance of the participants in the emotional problems, whereas a mixture of emotional and intuitive influences, and thus related processes, caused the performance on the moral-related problems. Since participants were able to deal with the abstract reasoning problems almost error-free, and solved most of the others problems as well (about 80 % correct), there is further evidence for rational processes involved in theoretical and practical reasoning.

Independent of these behavioral results and the drawn assumptions, the main foci of the fifth experiment were the neural correlates of theoretical and practical reasoning, and whether there are similarities between these processes and associated brain areas. This will be discussed in the following section.

9.3.2 Imaging data

9.3.2.1 Abstract problems

The bilateral activations found for the abstract reasoning problems are in line with the idea of a widespread network of cortical structures involved in theoretical reasoning (Goel, 2007, see introduction). This network, encompassing frontal, temporal, and parietal structures, presumably associated with emotional/intuitive and rational processes as well as ToM and working memory, accordingly supports ‘dual-process’ accounts of theoretical reasoning (e.g., Evans, 2008; Goel, 2007, Liebermann, 2003). In turn, ‘dual-process’ models account for an involvement of emotional/intuitive and rational processes in reasoning. However, Goel (2003) could only identify a bilateral fronto-parietal network activated by abstract reasoning, but is lacking temporal structures. Other studies in turn, could show activations for reasoning per se in a bilateral fronto-temporo-parietal network (Fangmeier, Knauff, Ruff, & Sloutsky, 2006), or a fronto-temporo-occipital network also including the basal ganglia and cerebellar structures (Goel & Dolan, 2003a). With regards to the activated clusters in the left DLPFC and the left parietal lobe, the results further support the findings of Knauff et al. (2003) who also discovered activations in a bilateral fronto-temporo-parietal network, where the left hemispheric activations in the frontal and parietal regions were increased as compared to the right hemispheric activations.

9.3.2.2 Moral problems

Problems with moral content also yielded bilateral activations in a widespread fronto-temporo-parietal network. This provides support for ‘dual-process’ accounts of practical reasoning (e.g., Greene et al., 2004) requiring a large cortical network (e.g., Casebeer, 2003a; Greene & Haidt, 2002; Raine & Yang, 2006). The results corroborate the notion of Goodenough and Prehn (2004) who stated that almost all brain imaging studies (e.g., Greene et al., 2001; Heekeren et al., 2003; Moll et al., 2001) –concerned with morals– found similar areas to be activated, although different tasks with different stimuli and task demands were used. Therefore, the results are in line with such findings by also identifying similar brain activations. This might be taken as further evidence for the applicability of the behaviorally validated reasoning problems. In contrast, it would be possible to argue that the actual experiment did not directly investigate moral reasoning since it used deductive inference problems. In fact, deductive inferences with a moral content (not moral reasoning in the literal sense) were explored. As already mentioned above however, the behavioral results and the difference contrasts (discussed below) yield further evidence that participants executed some kind of moral reasoning/judgment making.

Another interesting point is that the clusters of activated voxels for the abstract problems were (descriptively) often almost twice as large as those for the moral ones. If morals are learned, the moral problems might have required less brain activity than the abstract ones, whereas most people actively need to do mental executions while solving a deductive/logical problem (we seldom reason without contents in such an abstract way). This assumption is supported by studies showing that over learned, practiced, or automated tasks could lead to decreased brain activity (e.g., Ramsey, Jansma, Jager, Van Raalten, & Kahn, 2004; Van Raalten, Ramsey, Duyn, & Jansma, 2008). Therefore, also the brain activations found for moral reasoning provide evidence for rational and emotional/intuitive processes as well as ToM and working memory processes involved in practical reasoning. This further supports 'dual-process' accounts of practical reasoning.

9.3.2.3 Neutral problems

The neutral problems led to bilateral fronto-temporo-parietal network activations. This seems to reflect the above mentioned results of the reasoning literature (e.g., Goel, 2007; Knauff, 2007, 2009a). Since the neutral problems served as controls to filter out 'simple' content effects, it appears astonishing that these controls activated almost comparable cluster-sizes as the abstract problems. The neutral material applied here might have been not familiar enough to facilitate reasoning not only behaviorally (reduced error rates), but also with respect to brain activity. Therefore, participants seemed to handle neutral problems in a similar way as the abstract ones.

9.3.2.4 Conjunction analysis

The major aim of the last experiment was to obtain evidence whether theoretical and practical reasoning is based on similar brain structures, and thus, similar processes or not. The results seem to support these assumptions since bilateral activations in all of the predefined ROIs for the conjunction of the abstract, neutral, and moral problems could be found. The main argument against these assumptions may be that the actual task only required deductive reasoning and thus led to common activations of the different stimulus sub-categories. However, as shown in the behavioral data as well as in the differences of the single effects or the difference contrasts discussed in the following, there are similarities and dissimilarities in the reasoning processes and their underlying neural substrates. It seems legitimate therefore, to assume that participants reasoned about (and judged sometimes according to) the content. Therefore, the moral content rather than the validity was decisive for the participants´ responses on the moral problems. Furthermore, according to the above mentioned brain imaging results of deductive and moral reasoning research, the results seem to be in line with the findings of both of these research domains. Moreover, for the first time this shows similarities between theoretical and practical reasoning.

9.2.3.5 Difference analysis

Comparing the abstract problems with the moral ones yielded bilateral activations in all of the predefined ROIs. These activations however, were more accentuated in the left hemisphere (larger voxel clusters activated with higher t-values). This especially applies to the DLPFC, the temporal and the parietal regions, except for the OFC which only showed right hemispheric activations in this comparison. Almost identical results were obtained from the contrast of the abstract and neutral problems. These results seem to support assumptions of Fangmeier et al. (2006), who proposed a bilateral fronto-temporo-parietal network for reasoning. Furthermore, the results partly match the accentuated left hemispheric activations in frontal and parietal regions in the study of Canessa et al. (2005) or Noveck, Goel, and Smith (2004) as well as of Knauff et al. (2003). However, the right-hemispheric and the temporal activations challenge the distinctions made by Goel (2003) who claimed that abstract content requires a more universal system involving a bilateral fronto-parietal network, whereas reasoning with semantic content activates a left hemispheric fronto-temporal network. Such lateralization could not be found in the current experiment. The data are in better agreement with findings of Monti and colleagues (Monti, Parsons, Martinez, & Osherson, 2007; Monti, Parsons, & Osherson, 2009), who also proposed common, but especially left hemispheric accentuated, ‘core and support’ brain areas for deductive inferences.

In contrast, Goel and Dolan (2003b) showed that a left fronto-temporal network is activated with familiar content, and bilateral parietal lobes and dorsal PFC are required for unfamiliar content. Since one can assume that abstract content is different from semantic content (i.e., moral problems), and differs even more from ‘simple’, not so familiar content (i.e., neutral problems), further studies seem necessary to resolve the open question whether the hemispheric specialization really exists or not. If a hemispheric specialization exists, as proposed by Goel and Dolan (2003a, 2003b) for example, studies should focus on the precise origins of this lateralization. Moreover, comparing the neutral problems with the abstract ones indeed yielded a left temporal activation, but the left frontal activations were absent. Providing therefore, only partial support for such lateralized network proposals again.

The comparison of the moral problems with the abstract ones only revealed left hemispheric activations, especially accentuated in temporal regions and the mPFC. These activations in the left mPFC seem to correspond to the assumptions of Goel, Shuren, Sheesley, and Grafman (2004) who proposed an asymmetry for deductive reasoning depending on the content. Thereby, they assumed that the left PFC responds more to social content and the right PFC to arbitrary content, which is comparable to the abstract condition. Although no increased unilateral activation could be found for the abstract problems in the right DLPFC or mPFC, the results are (partially) in accordance with the proposal of Goel and colleagues (Goel et al., 2004) that reasoning with different contents results in a hemispheric asymmetry. This asymmetry however, needs to be further explored since it might not be differentiated only by semantic content versus ‘pure’ abstract reasoning (see also arguments above). This assumption gets further support from the bilateral activations of almost all cortex regions in the single effects analysis for all conditions. Conversely, the left hemispheric activations in the contrast between moral and abstract problems provide thorough evidence for an additional left hemispheric activation for familiar (Goel & Dolan, 2003b) or social content (Goel et al., 2004).

Regarding the research on moral reasoning, the left DLPFC activation of the moral problems compared to the abstract ones seems to match the results of Heekeren et al. (2003), Moll et al. (2001), and Schaich Borg et al. (2006), who also found left DLPFC activation with moral stories and scenarios. The mPFC activation of the same comparison partly matches the VMPFC activations of diverse studies using personal moral dilemmas (Greene et al., 2001; Greene et al., 2004), moral pictures (Harenski & Hamann, 2006), simple ethical sentences (Heekeren et al., 2003), or socio-normative judgments (Prehn et al., 2008). The dorsal fractions of the mPFC ROI and their activations seem to correspond to the findings of Fiddick et al. (2005) concerning their social contract trials. The increased temporal activations for moral problems also replicate, at least partly, several studies (e.g., Greene et al., 2004; Heekeren et al., 2003; Prehn et al., 2008).

Canessa et al. (2005) for example found right hemispheric activations for social contents, while Heekeren et al. (2003) detected also bilateral temporal activations for moral sentences compared to controls. Thus, it seems necessary to investigate potential hemispheric asymmetries associated with varying contents more systematically to gain further insights in these possible hemispheric specializations.

10. General Discussion

The work presented here investigated human reasoning using behavioral and brain imaging experiments. Thereby, the main goal was to find similarities of theoretical (i.e., how humans do or can reason) and practical reasoning (i.e., how humans should use their reasoning abilities to act reaching 'eudaimonia', or to act morally, respectively). These similarities should be based on common cognitive processes (i.e., rational and emotional/intuitive) and underlying brain structures. Further influencing factors mediating these processes were also suggested and discussed (i.e., working memory and intelligence for example).

The reported experiments reveal indeed evidence for similar but also distinct cognitive processes and brain structures involved in theoretical *and* practical reasoning. These findings seem to support assumptions already made by Damasio (2006), LeDoux (2006), Thagard (2006), or Monti et al. (2007, 2009), whereby the later postulated 'core and support' brain areas and associated processes involved in reasoning focusing exclusively on theoretical reasoning. Therefore, the current experiments seem to fulfill a requirement already proposed by Evans (2008). He suggested a reunification of the theoretical and practical reasoning research to explore potential similarities of these two reasoning domains. Evans (2008) derived this idea by referring to the various 'dual-process' accounts that exist in both of these research domains. However, they were developed rather in isolation than in combination. Thus, adapting Evans (2008) proposals, a suggestion for the current findings is to interpret them as providing evidence for 'Type 1' and 'Type 2' processes (derived from the terms 'System 1' and 'System 2') involved in theoretical and practical reasoning. It seems important to further adapt Evans' notion that these two processes/systems/types involved in human reasoning can both be conscious or unconscious, and can work together, or separated, or in opposition to each other. Even though there is some dispute what kind of terms and associated processes could really be ascribed to the two different systems, a distinction can be made along the potential involvement of working memory processes that are only related to 'Type 2' processes (cp., Evans, 2008). Therefore, 'Type 1/System 1' processes seem to be rather fast and automatic, whereas 'Type 2/System 2' processes might be rather slow and effortful (cp., Samuels, 2006). For the current work and in accordance with the literature of both research domains, it could be further assumed that 'System 1' processes are rather associated with emotions and (over) learned prior knowledge (i.e., intuitions), whereas 'System 2' processes' rather refer to rationality and 'pure' reason. The actual experiments provide no evidence however, which of these potential processes involved in human reasoning is conscious or unconscious. Furthermore, whether or not terms such as 'somatic markers' (Damasio, 2006), or ‘informed intuition’ (Thagard, 2006), or ‘high and low road’ as well as ‘mental models’ are better suitable in describing some of these processes involved cannot be judged with the experiments conducted within this thesis. Apart from this, the current results appear to be in line with such a cautious phrased ‘dual-process’ account based for example on the finding of the incongruence effect and its variable occurrence. “Normal” participants showed this effect whereas trained participants or participants with superior intelligence did not. This leads to the assumption of two different processes involved which can conflict, or interact, or work in isolation, respectively. Yet, the concrete mechanisms remain unclear and have to be explored further. The same accounts for other processes and influencing factors, not in the focus of the current work (e.g., situational context, learning history, personality traits). Nevertheless, the conclusion one might draw from the current findings seems rather to extend the suggestion by Monti et al. (2007, 2009) of ‘core and support’ units involved in theoretical reasoning to the domain of practical reasoning, instead of adapting the interpretation of Fiddick et al. (2005) to theoretical reasoning. They assumed various content-dependent processes for practical reasoning and rejected the idea of ‘dual-process’ accounts.

The following sections shall discuss the reported findings and interpretations of the various experiments in detail. It aims to provide a critical integration and starts with the three pre-studies.

10.1 Pre-Studies

The pre-studies indicated the applicability of the deductive inference problems with different contents. All problems seemed to represent their intended contents. Concerning the logic dimension, the abstract problems received the highest rating in Pre-Study I, and the unmoral problems the lowest ratings in all three pre-studies. Moral content problems were almost equivalent to neutral, positive emotional, negative emotional, neutral1, and neutral2 problems across the three pre-studies. The invalid problems received lower ratings than the valid ones, within their corresponding contents. In addition, invalid problems were most often also rated lower than valid ones regardless of the content categories. Even though the participants of the pre-studies were not supposed to draw inferences, the problems were evaluated also according to their inherent logical structure. This yields evidence that the distinction between problems where the conclusion follows from the premises, and those where drawing an inference is impossible, was rated appropriately. Moreover, rating the abstract problems in the logic dimension highest seems to reflect their relation to ‘pure’ deductive reasoning, thus logic. The participants of all pre-studies and experiments were presumed to be no logicians or trained in logic (prior to the current study). To recognize the relation between abstract problems and logic so directly, might therefore reflect some “inborn” or “learned” affinity to logic, hence their rational reasoning abilities/capacities. However, this assumption remains speculative, even though Kant for example (1781/1998) proposed something very similar in his ‘critique of pure reason’, whereby his metaphysics did not refer to inborn but to ‘a priori’. In addition, more empirically based evidence corresponds to the chosen reasoning problems. These were the ‘simplest’ ones available which are almost always solved without errors (cp., Knauff, 2007). Therefore, rating the abstract problems as most ‘logical’ might be also due to their obvious and easy inherent structure. This is particular when all three sentences are presented at the same time (i.e., on one page together).

Another interesting point concerning the logic dimension is that unmoral problems received the lowest ratings. This could be interpreted as first evidence that parts of the knowledge about morals, persons might possess, are according to the content of the human rights articles. Without such ‘moral’ prior knowledge participants should have rated the moral and unmoral problems as equally logical. However, the occurring reasoning process of a participant rating an unmoral problem might have been equivalent to “this is unmoral, thus illogical”. Vice versa, moral problems might have been sensed as serving our human purposes reaching a fruitful and peaceful flourishing life, thus being logical. It could also be even simpler. It is morally “OK” since the participants learned statements like these in institutions (i.e., school, church), thus rating moral problems as logical. These hypotheses are on the one hand very speculative, but receive further support from the ratings of the moral or unmoral problems in the positive and negative emotionality dimensions.

The unmoral problems were rated as more negative than all other ones (except the negative emotional problems), whereas the moral problems were rated as more positive than all other inferences (except the positive emotional ones). Basically, one might assume that moral-related content is highly emotional. This was confirmed by the current findings, especially in Pre-Study I, where only emotionality was interrogated. However, a further speculation is that morals are negatively emotional laden. Since education and therefore learning of morality are often strongly intertwined with duties and obligations, and failings are almost always immediately penalized, children might evaluate and associate morals with negative feelings. However, when children begin to realize why morals (e.g., in form of laws, social rules, etc.) are useful for a peaceful and flourishing coexistence, their negative perception of morals might shift towards a more positive one. Therefore, normative statements and morally proper actions seem to be accompanied by positive feelings, whereas only unmoral actions violating “our moral sense” lead to negative appraisals. Interestingly, these speculations seem to have several implications for a theory of morality. The fact that emotions (positive and negative) play an important role in moral reasoning and decision making challenges proposals of ‘pure’ rational process accounts of morality (e.g., Kant, 1785/2000, 1797/2007; Kohlberg, 1969; Piaget, 1932; Schaich Borg et al., 2006). In contrast, the assumed shift in children’s learning processes –understanding and internalizing morals– seems to require more rational processes for the evaluation and the recognition of real emotions, followed by the insight how useful morality is for “a better life”. This notion clearly contradicts “emotional theories” of morals (e.g., Haidt, 2001; Hume, 1751/2003; Hutcheson, 1755/1968; Shaftesbury, 1711/1963). Another assumption that has to be made refers to these ratings which revealed evidence that participants in the pre-studies were also influenced by the content of the problems, thus reasoned about and perhaps also answered/rated according to this content. Therefore, the pre-studies present first evidences for emotional/intuitive and rational processes involved in theoretical and practical reasoning. However, whether or not morals are learned and represent therefore prior knowledge/intuitions, or reflect some automatic moral attitudes (cp., Luo et al., 2006), it cannot be decided within this current work (see also discussion in the introduction and below).

Nevertheless, the concept of emotions being involved in morals, especially as a kind of “moral emotions” for approval and disapproval of actions, has already been proposed by different researchers (Casebeer, 2003b; Goodenough & Prehn, 2004; Haidt, 2001; Moll, De Oliveira-Souza, Eslinger, Bramati, Mourao-Miranda, Andreiuolo, & Pessoa, 2002; Nichols & Mallon, 2006; Smith, 1759/2004). Unfortunately, this idea has not been linked to emotional evaluations in “normal” reasoning processes which are not concerned with moral contents. This seems astonishing, in particular, since a few proposals state that emotions support learning and serve as an appraisal tool for all reasoning and decision making processes, perhaps represented as ‘somatic markers’ (Damasio, 2006).[17] Emotions seem therefore to be an important part of our cognitions (see also LeDoux, 2006; or Thagard, 2006).

These assumptions made for the pre-studies will be discussed further in the following sections starting with the first two behavioral experiments.

[17] The current pre-studies and experiments provide thorough evidence for the involvement of emotionally laden prior knowledge (i.e., emotional/intuitive processes) on reasoning, but could not account for the whole ‘somatic marker’ hypothesis by Damasio, which requires the entire body to be activated and involved, when storing and retrieving this emotional laden knowledge. To investigate this, was beyond the scope of the current experiments. However, the concept of Damasio seems the most plausible one explaining influences of emotional laden prior knowledge on cognitive processes, thus the involvement of emotional/intuitive processes in reasoning.

10.2 Experiments I + II

The first two experiments show that moral-related reasoning problems, especially in the conflicting incongruent condition, resulted in more errors and higher decision times (cp., Evans, 2008; Greene et al., 2004). Incongruity effects have often been investigated in deductive reasoning with valid conclusions but unbelievable content (and vice versa), hence contradicting prior knowledge (Blanchette et al., 2007; Goel & Dolan, 2003b; Knauff, 2007; Lou et al., 2006; Manktelow & Evans, 1979). As reported in the introduction, these effects were thus called ‘belief bias’ (Evans, 1989, 2008). This ‘classical belief bias’ however, does not match the current findings. As already discussed above, the incongruity effects appearing in the current experiments might therefore, if at all, be labelled ‘new belief bias’. Since this incongruity was embedded in the current experiments to ensure that participants reason and answer according to morals and not to logic; this occurring incongruence effect is a hint that participants indeed reasoned and answered according to the moral-related content. Hence, the reasoning problems contained aspects of practical as well as theoretical reasoning.

A possible different explanation of the appearing incongruity effect has been mentioned already. It refers to automatic moral attitudes (Luo et al., 2006). These attitudes represent activated stimulus-reinforcement associations, thus an automatic evaluation of a moral stimulus as good or bad. In incongruent cases, two opposing valence expectancies are built, leading to a response conflict which has to be solved. However, Luo et al. (2006) refer also to patient studies which revealed evidence that moral attitudes can only be created adequately if persons were able to learn and therefore form the necessary stimulus-reinforcement associations. Therefore, moral attitudes seem also to be based on learned prior knowledge or intuitions associated with emotional processes. This knowledge comes into mind automatically and emotionally laden. For this reason, the current work adapted the ‘somatic marker’ hypothesis by Damasio (2006). Moreover, also Luo and colleagues (Luo et al., 2006) refer to more ‘rational’ processes and associated brain structures which are activated in addition to the areas associated with moral attitudes. These additional brain areas are required to solve a potential incongruity-based conflict, in particular, when explicit reasoning is required. Thus, it remains an open question, whether (over) learned and emotionally laden prior knowledge leads to attitudes or just represents emotionally laden intuitions. It is also an unresolved issue, whether knowledge has to be separated from emotional evaluations/loads or not. Further research seems therefore necessary.

In contrast, the assumption that morality represents (over) learned knowledge is supported more and more (Churchland, 1998; Goodenough & Prehn, 2004; Kohlberg, 1969;

Moll et al., 2005; Pasupathi & Staudinger, 2001). This has also been shown in various patient studies (Koenigs et al., 2007; Mendez, Chen, Shapira, & Miller, 2005; Raine & Yang, 2006). Since prior knowledge means that certain facts have been (over) learned, these (over) learned facts seem to create the necessary basis interfering with task demands. In turn, these intuitive processes might then have been responsible for the incongruity effects found in the current experiments.

Furthermore, the cited patient studies yield evidences that moral knowledge could be preserved after brain injuries, but ‘correct’ moral decisions also require emotions, since morality represents highly emotional material. Koenigs et al. (2007) for example, investigated moral decision making of six patients with ventromedial prefrontal cortex (VMPFC) lesions as compared to brain-damaged comparison participants (no damage in areas associated with emotions) and healthy controls. They used moral-personal, moral-impersonal and non-moral dilemmas. The VMPFC-patients had problems only with emotional responses and emotion regulation, but normal intelligence, normal deductive reasoning abilities, and preserved knowledge of social and moral norms. Koenigs et al. (2007) only found differences between the controls and the VMPFC patients for the moral-personal dilemmas (especially for the highly conflicting ones) indicating that these patients were impaired in moral judgments involving emotionally salient actions. The decisions of the patients represented a kind of utilitarian judgment in these dilemmatic scenarios. This means that they approved killing one person in order to save the lives of five other persons, for example. The authors (2007) suggest that their results show intact knowledge of social and moral norms in VMPFC patients. This was proven due to the missing emotional reaction to harm of others in the moral-personal dilemmas. In these dilemmas the patients had to rely on their explicit knowledge in dealing with the decision and consequently revealed a kind of utilitarian judgment. Therefore, Koenigs et al. (2007) conclude that “[…] the present findings are consistent with a model in which a combination of intuitive/affective and conscious/rational mechanisms operate to produce moral judgments.” (p. 910). As mentioned above, the authors also claim that the VMPFC is rather an integration area of rational and emotional processes in moral reasoning than an exclusive ‘moral decision area’. It seems still unclear however, whether all emotional abilities of these patients were impaired since also the moral-impersonal dilemmas contained emotional load, where patients did not show any performance differences compared to both of the control groups.

Apart from this, these patient studies support the findings of the current experiments that practical reasoning requires emotional/intuitive and rational processes. In turn, learning morality, highly emotional material as well as executing moral tasks after years of learning might create and further activate ‘somatic markers’ (Damasio, 2006). Thus, a possible explanation concerning the emotional/intuitive proportions of moral reasoning processes in these patients might be missing or impaired ‘somatic markers’. These markers contribute to cognitive processes besides the rational proportions.

With regards to the ‘pure’ deductive reasoning problems (i.e., abstract inferences), participants were fastest and made almost no errors at all. This indicates that participants were able to reason theoretically although the applied problems were the easiest available (cp., Knauff, 2007). This missing variation in problem complexity however, restricts these assumptions. Moreover, the abstract problems were sometimes more prone to errors than the neutral ones. Thus, there is evidence that persons normally do not reason according to ‘pure’ logic in such an abstract way. In turn, these facilitative effects of the neutral problems as compared to the abstract ones support proposals of positive content effects in deductive reasoning (e.g., Wason & Shapiro, 1971; Johnson-Laird et al., 1972). Taking the emotional controls into account complicates this picture. Indeed, it shows the reverse effects. In particular, the positive and negative emotional invalid problems showed increased decision times and error rates that contradict facilitative content effects. This rather seems to be in line with the findings of Blanchette and colleagues (Blanchette, 2006; Blanchette & Richards, 2004; Blanchette et al., 2007) that emotionally laden content could lead to positive, but also to negative performance in deductive reasoning. Again in turn, the valid emotional problems showed the potential facilitative content effects. Apart from an abstract version of deductive reasoning problems according to classical logic, this leads to the suggestion that theoretical reasoning which is normally not content-free, is influenced by emotions caused by the emotional load of the reasoning problems. So, theoretical reasoning seems to be based on rational and emotional processes, as has already been shown for practical reasoning as well. One might object that all theoretical reasoning problems produced lower error rates than the practical ones, and therefore, the real influences on these reasoning problems were solely based on prior knowledge. However, as shown in the moral-valid and the unmoral-invalid conditions, where prior knowledge also played a role, the ‘intuitive’ proportions of the involved reasoning processes could not account for all results of the moral-related problems. Indeed, the moral-valid problems showed facilitative content effects, whereas the unmoral-invalid ones showed slightly negative content effects. The latter finding might account for disturbing negative emotions on reasoning. In turn, the fact that the participants could deal with the moral-valid inferences might endorse the human ability to reason deductively and to deregulate emotional influences (cp., Pizarro et al., 2003).

The findings discussed so far yielded first evidences for rational and emotional/intuitive processes involved in theoretical and practical reasoning, but remain unsatisfactory. Therefore, further support seems necessary.

10.3 Experiment III – Training

The third experiment not only showed that training in deductive reasoning improves performance (cp., Cheng et al., 1986; Klaczynski & Laipple, 1993; Klauer et al., 1997; Klauer et al., 2000), but also that transfer effects are possible (cp., Klacynski & Laipple, 1993; Ziegler, 1990). The trained participants improved their reasoning performance in the post test as compared to the pretest as well as their results in the computer experiment, thus transferring their learned knowledge with abstract, absurd, and emotional problems to moral-related and neutral reasoning problems. Also the trained participants however, were more prone to errors and had higher decision times in the moral-related conditions as compared to the neutral ones, even though these effects could only be shown descriptively.

The content effects caused by the practical reasoning problems, in particular by the moral-invalid and the unmoral-valid ones, where content and logical form conflict represented again an incongruence effect in deductive reasoning. Hence, the moral-invalid problems produced most errors while moral-valid inferences were solved almost perfectly. This incongruity effect might have been increased for the groups without training or pseudo training due to the time pressure in Experiment III (cp., Evans & Curtis-Holmes, 2005). The time limit for evaluating the conclusions was set in Experiment III to enhance training differences. This time pressure discerned not only the groups of the third experiment better, but also increased the error rates of both of the controls groups as compared to the error rates of the also untrained participants in the first two experiments. These time pressure effects seemed to have been especially heightened for the group without training. Thus, if one had argued that for example the group with the pseudo training performed worse due to an exhaustive but unhelpful training, the time pressure effects of the group without training could contradict such an assumption.

Interestingly, all participants, sometimes also the trained group, were influenced by the moral-related content, thus showing practical reasoning. Furthermore, these content influences, in particular the incongruence effect, might represent ‘intuitive’ processes involved in practical reasoning. Additionally, the effects of the moral-valid and neutral reasoning problems are a hint for facilitative content influences on theoretical reasoning (cp., Cox & Griggs, 1982; Griggs & Cox, 1982; Johnson-Laird et al., 1972; Wason & Shapiro, 1971). However, the slightly enhanced error rates of moral-valid problems compared to the neutral ones, for example in the trained group, also seem to contradict facilitative content effects. At least, if one assumes that theories could explain reasoning that postulate that social content by itself should have been helpful (Canessa et al., 2005; Cheng & Holyoak, 1985; Cosmides, 1989; Manktelow & Over 1991). In contrast, one might suggest that there was a mixture of practical and theoretical reasoning, owed to the problem inherent structure, and both were not only influenced by prior knowledge but also by rational processes. Such assumption receives further support from the obtained training effects, since training improved only the rational proportions of the involved reasoning processes. An objection might be that the training applied did not only improve rational reasoning processes. However, the training was derived from the one used by Klauer et al. (2000) which explicitly focused with their training on improving deductive reasoning abilities (especially in the version that worked best, and was thus applied here). Furthermore, the participants of the current experiment were not trained in moral reasoning, or taught moral knowledge, or instructed how to control and to guide their emotions. It is clear that this cannot definitely exclude different factors that might have been trained. There is evidence however, that only the rational reasoning abilities of the training group were improved.

The fact that all groups seemed to apply rational and intuitive reasoning processes yields evidence for ‘dual-process’ accounts of theoretical and practical reasoning (e.g., Evans, 2008; Greene, 2007). This appears to contradict mental rule theories (e.g., Braine & O’Brien, 1998; Braine et al., 1995; Rips, 1994). In accordance to the mental rule theories, participants should have ignored the content and should have dealt only with content-free, abstract rules. This was clearly not the case for both of the control groups but also for the trained participants in the moral-valid condition. Moreover, the results also object the domain-specific accounts like the pragmatic reasoning schema theory (e.g., Cheng & Holyoak, 1985) or the logic of social exchange theory (e.g., Cosmides, 1989). These theories contain the assupmtion that reasoning should be built on acquired (via typical life experiences) or innate rule structures. With regards to deontic questions in particular, these rules/schemas should have produced facilitative content effects and even less incongruity effects (see also critics by Manktelow & Over, 1991, 1995). Therefore, according to these theories, participants should have had performance gains in the reasoning problems especially in the moral-related conditions just because of content (cp., Canessa et al., 2005; Cheng & Holyoak, 1985; Cosmides, 1989).

Vice versa, there is evidence that also the mental model theory (e.g., Johnson-Laird, 1983, 2006; Johnson-Laird & Byrne, 1991; Johnson-Laird et al., 1992), besides 'dual-process' accounts (e.g., Evans, 2008), might at least partly be in line with some of the results. The weakest evidence for the mental model theory is based on the fact that the logic training of the current experiment was derived from former trainings showing that training according to the mental model theory works best (for an overview see Klauer & Meiser, 2007). However, the current experiment did not intend to investigate the success of different training methods. Thus, this evidence remains rather weak, since no comparison trainings according to other reasoning theories were applied. Further support for the mental model theory could be derived from the fact that it deals with mental models containing semantic content. This means that the reasoner does not use abstract, content-free formal rules for his reasoning processes, but rather relies on the content of the given premises. The participants of the current experiment therefore, might have dealt with the moral-related content of the premises building mental models for their reasoning processes. Furthermore, the mental model theory also postulates influences of prior knowledge on the reasoning process and provides a shortcut to working memory (e.g., Baddeley, 1986, 2003) involvement in deductive reasoning. So, the participants might have built mental models of the premises and related these to prior knowledge of morals. Now, the mental model theory postulates that "[...] reasoners may fail either because they miscomprehend the premises and construct inappropriate models or because the required manipulations of models overtax their working memory." (Klauer & Meiser, 2007, p. 213). There is evidence that the performance of the participants with pseudo training or no training was disrupted by working memory constraints. Meiser and colleagues (Meiser et al., 2001), for example, showed that working memory capacity plays an important role in deductive reasoning after being trained according to the mental model theory in contrast to being trained with regards to other reasoning theories. Thus, Meiser et al. (2001) denote the influence of working memory capacity on reasoning as inherent in the mental model theory account of deductive reasoning. So, an explanation for the trained participants of the current experiment might be that their working memory capacities were improved due to the prior training. Hence, they could better focus on the logical structure of the deductive reasoning problems and suppress upcoming intuitive/heuristic processes (i.e., moral knowledge) more easily than the untrained participants. An advanced "moral" theory, based on the mental model theory, referring to deontic propositions and using such an explanatory proposal has been presented by Bucciarelli et al. (2008). This theory however, deals with 'intuitive'/heuristic and emotional proportions in the reasoning process, thus it seems connected to a 'dual-process' account rather than to the mental model theory which is almost exclusively based on rational and conscious reasoning processes (Manktelow, 2004). They try to circumvent such difficulties by proposing that moral reasoning is deontic reasoning. Deontic reasoning is on the one hand ‘normal’ reasoning and on the other hand it is specialized for morals. This seems to be quite confusing. Furthermore, Bucciarelli et al. (2008) postulate that their theory is based on mental models, but that decisions are sometimes based on emotional evaluations, and sometimes on deontic evaluations. Finally, they differentiate between intuitions and emotions, which also appear to be misleading. If morals are learned and concerned with highly emotional issues (Bucciarelli et al., 2008 share this view), it does not seem to be useful to discern intuitions (i.e., prior knowledge/learned morality) from emotions so extremely. In turn, this also represents a potential critique of the ‘classic’ mental model theory, which struggles with a possible explanation of the involvement of ‘intuitive’ and emotional processes in theoretical and practical reasoning. As Verschueren and colleagues (Verschueren et al., 2005a) noted, it might be reasonable to integrate the mental model theory in a ‘dual-process’ account representing the involved ‘System 2’ processes (e.g., Evans, 2003) to circumvent such difficulties.

Another point supporting the assumption of a working memory involvement in the current experiment refers to the decision times of the trained participants. The group receiving the logic training was always faster than both of the control groups, but needed more time answering the moral-related problems than the neutral ones. This might indicate that activated emotionally laden prior knowledge (i.e., heuristic, emotional, and intuitive processes) had to be suppressed by analytic processes (i.e., rational processes) first, before the learned rule could be applied. Suppression mechanisms might have caused a higher working memory load, since analytic processes are associated with working memory capacity (e.g., Evans, 2008; Johnson-Laird, 2006). In turn, working memory involvement in theoretical reasoning, mediated via prior knowledge provides evidence again for ‘dual-process’ accounts of theoretical reasoning. The influence of ‘intuitive’, perhaps emotional laden, processes on practical reasoning also seems evident. However, as mentioned above, the rational proportions of the reasoning processes involved in practical reasoning could not be tested with the current experiment. Therefore further experiments were conducted, which will be discussed in the following section.

10.4 Experiment IV - Intelligence

The results of Experiment IV seem to reinforce the previous assumptions. The fact that high intelligent participants did not show an incongruence effect as well as less errors and shorter decision times in all reasoning problems compared to the group with average intelligence indicates a general effect of intelligence on reasoning. This suggestion is supported by the correlation analysis, where especially the fluid IQ values were correlated with the decision times and error rates of all problems. An explanation for this general effect of intelligence is that theoretical reasoning performance is often associated with ‘System 2’ processes, which are in turn related to working memory capacity/performance (cp., Evans, 2008). For working memory, a connection between higher intelligence and higher performance levels was shown earlier (e.g., Colom et al., 2004; Johnson-Laird, 2006). One might conclude therefore, that participants with higher fluid intelligence possess higher working memory capacities, thus better theoretical reasoning abilities.

In contrast, the crystallized intelligence component was only correlated with the c-score and the error rates of the moral-related reasoning problems, whereby crystallized intelligence is assumed to reflect accumulated knowledge, thus ‘System 1’ processes (cp., Evans, 2008). Former studies showed that there are also connections between IQ and moral reasoning/competence levels (e.g., Kohlberg, 1969; Rest, 1979). Furthermore, a higher IQ allows for better learning performance and earlier achievement of higher moral competence levels (Pasupathi & Staudinger, 2001; Stojiljković, 1998). Thus, it can be concluded that participants with higher crystallized IQ values had better learning and memory capacities for morals, perhaps stored as ‘somatic markers’ (Damasio, 2006). This would indicate better practical reasoning abilities.

So far, there is evidence that persons with high fluid and crystallized intelligence levels have higher working memory capacities. This leads to an improvement of rational reasoning processes in general as well as higher knowledge-related performance levels, therefore an improvement of (emotional) intuitive reasoning processes as compared to persons with average intelligence level. For the current experiment however, one might assume that the higher moral competence of the high intelligent participants should have conflicted with the better general reasoning abilities. With respect to the results, this was not the case, even though the incongruity effect seems to be based on activations of ‘intuitive’ processes dealing with (over-) learned knowledge. Thus, to avoid a potential incongruence effect, one has to suppress these upcoming, automated processes by more ‘rational’ and deliberate processes. Regarding the correlations of the fluid intelligence component of the high intelligent participants may provide an explanation. The fluid IQ values correlated with almost all decision times and error rates. Therefore, high intelligent participants possess high 'rational' reasoning abilities. Although they also possess high practical reasoning abilities at the same time (compare c-score values), the rational reasoning abilities might have even trumped the 'intuitive' ones, thus suppressed these automatic knowledge-based processes. Another speculative explanation refers to potential rational influences on practical reasoning which were used to guide moral decisions in the "right" direction (i.e., to reason and subsequently act according to the human rights articles). Evidence for this assumption is available in the few errors made by the high intelligent participants. These exclusively occurred within the unmoral reasoning problems. Nevertheless, these assumptions remain speculative and further evidence is necessary regarding the potential influences of 'rational' processes on 'emotional/intuitive' ones in reasoning.

Potential critics concerning this experiment have been mentioned already, but a few points have to be discussed more extensively. The main point is the small sample and thus the weak power of the statistics applied as well as the limited possibility to generalize the results, which were also sometimes based on a trend rather than clear significant effects. Concerning the sample, there was not only a huge mean age difference between the two groups, but also a gender mismatch. The group with the high intelligent participants consisted almost exclusively of men (5 men, 2 women), whereas the group with the average intelligent participants was composed mainly of women (6 women, 1 man). The former group represented a mean age of 29 years and the later had a mean age of 21 years. Thus, it might be that besides intelligence based differences, confoundations due to age and/or sex were responsible for the results in the experiment and in the MJT. However, there is evidence that the IQ values of men and women are equal on average and that no sex differences exist (e.g., Colom, García, Juan-Espinosa¸ & Abad, 2002; Hyde, Lindberg, Linn, Ellis, & Williams, 2008). In turn, men normally show a greater variance performance in IQ tests than women, representing more extreme values in the tests applied (Amelang & Bartussek, 1990). Therefore, more women can be found with average IQ values, whereas more men are available with extreme IQ values. This might explain why it was easier to acquire men for the superior intelligence group than women and vice versa. Furthermore, it seems possible to exclude potential sex-related differences concerning logical reasoning abilities. A study by Lynn, Backhoff, and Contreras-Niňo (2004) for example showed that there are no sex differences in logical reasoning performance. Berzonsky and Ondrako (1974) also reported no sex differences in deductive reasoning abilities. With regards to the age differences, there is also the assumption that they did not have a confounding influence. In accordance to Piaget's proposals, humans reach the formal operational stage (i.e., abstract reasoning abilities) from age 12 onwards (Miller, 1993), and a possible decline of logical reasoning abilities is assumed not to occur before the age of 50 years or older (Viskontas, Morrison, Holyoak, Hummel, & Knowlton, 2004). Finally, one might suggest that the moral judgment competence levels differed between the groups due to age or sex influences. But again, also for moral judgment competence, Lind and colleagues (Lind, Grocholewska, & Langer, 1987) could not find any sex differences, and also no effects due to age differences are reported. It seems therefore plausible to relate the groups differences found in the actual experiment to intelligence differences. Nevertheless, a few doubts remain due to the small sample size. To gain further evidences for the main assumptions of the current work, a final experiment was conducted, which shall be discussed below.

10.5 Experiment V - fMRI

The behavioral results of this last experiment were similar to those of the former ones, particularly of the second experiment, where the same reasoning problems were applied. Of more interest in this investigation were the neural correlates of theoretical and practical reasoning. Several Regions of Interest (ROIs) were postulated according to the existing findings of theoretical and practical reasoning research: dorsolateral prefrontal cortex (DLPFC), medial prefrontal cortex (mPFC), orbitofrontal cortex (OFC), anterior cingulate cortex (ACC), temporal pole, medial temporal gyrus (MTG), superior temporal sulcus (STS), parietal lobe, and precuneus. The different brain areas assumed to be involved in both reasoning domains were associated with specific cognitive processes according to the existing research results. The DLPFC is associated with rather rational processes (e.g., Greene et al., 2004; Schaich Borg et al., 2006), the OFC is associated with rather emotional processes (e.g., Dolan, 2002; Prehn et al., 2008), and the mPFC (especially its ventral parts) counts as integration area for both of these processes (; e.g., Adolphs, 1999, 2006; Damasio, 2006). Furthermore, the ACC is associated with working memory processes and executive functions (e.g., conflict detection and monitoring; e.g., Moll et al., 2003; Young & Koenigs, 2007), whereas the STS is associated with ToM processes and sensory integration of 'information' (e.g., Frith & Frith, 1999; Heekeren et al., 2003; Luo et al., 2006). The temporal areas are often related to emotional~, intuitive~, and knowledge-based processes (e.g., Goel & Dolan, 2003b; Moll et al., 2005), and the parietal cortex seems to be responsible for rather rational and working memory processes or reasoning with and without content (e.g., Canessa et al., 2005; Goel & Dolan, 2003b; Greene et al., 2001).

The results of this last experiment seem to confirm the assumption of similar brain areas to be involved in theoretical and practical reasoning. Brain activations were found in all of the predefined ROIs for the practical (i.e., moral) and theoretical (i.e., abstract) reasoning problems, both in the single effects analysis and the conjunction analysis. This gives evidence for similar cognitive processes to be involved in theoretical and practical reasoning, namely 'rational' (and working memory related) *and* emotional/intuitive processes.[18] The brain structures found to be activated by theoretical and practical reasoning problems thus encompassed a widespread fronto-temporo-parietal network. Therefore, these brain activations and their associated cognitive processes reveal further evidence for 'dual-process' accounts representing and explaining the reasoning processes involved in theoretical reasoning (e.g., Goel, 2007) and in practical reasoning (e.g., Greene, 2007) best. This also endorses the suggestion already made by Evans (2008), who derived his proposal however only theoretically, and awaited empirical confirmation. The current experiments, especially this last one, might have provided first evidence for his suggestions, although further experiments appear to be necessary (see critics section below).

Apart from these common activations, the last experiment provided also some distinct brain activations for theoretical and practical reasoning, with bilateral activations for the first one, and exclusively left-hemispheric activations for the latter one in almost all of the ROIs. These obtained hemispheric asymmetries could lead to an additional separation, not just between 'emotional/intuitive' and 'rational' reasoning processes, but rather between different reasoning 'sub-processes' (reasoning without content and reasoning with content, i.e. moral, social, emotional, etc.). This seems to be in line with a suggestion originally made only for theoretical reasoning by Monti et al. (2007, 2009) who postulated 'core and support areas' to be involved in reasoning. The investigation of reasoning about specific contents (events, actions, goals) with various reasoning problems (i.e., syllogisms, conditionals, dilemmas) could be means to further explore the cognitive and neural 'sub-categories' of the human reasoning system (theoretical, practical). For example, the frontal cortex seems to include much more specific brain areas than have so far been defined and investigated (cp., Moll, De Oliveira-Souza, Moll, Ignacio, Bramati, Caparelli-Daquer, & Eslinger, 2005). Such segregation of brain structures could be an interesting step for future reasoning research.

[18] Note that particularly the fMRI experiment yielded evidence for emotional laden prior knowledge, thus influences of emotions on theoretical and practical reasoning, since the orbitofrontal and temporal areas found to be activated are particularly associated with emotional and intuitive processes.

10.6 Critics and limitations

The preceding experiments dealt with deductive inference problems containing various contents. The aim of all experiments was to find similarities and maybe a common basis for theoretical and practical reasoning concerning cognitive processes and related brain structures. A first possible critique on all experiments is that they only encompassed theoretical reasoning problems, thus only investigated theoretical reasoning. Although the embedded possibility of an incongruity effect, indeed appearing in all experiments, should ensure that participants partly reason and decide according to practical reasoning, one cannot finally exclude the possibility that only theoretical reasoning occurred. The fact however, that the trained participants produced the most errors with incongruent moral-related problems as compared to all others seems to support the assumption of occurring practical reasoning processes, even though these effects could only be seen descriptively. Apart from this unsolved question, the moral-related problems resulted in huge content effects (except for the participants with superior intelligence). Content effects are known from many studies (cp., Manktelow, 2004). One content effect often found is the so called ‘belief bias’ (e.g., Evans et al., 1983). This effect represents the fact that prior knowledge/beliefs (activated by the content of the problem) interact with rational processes (activated by the logical structure of the problem). This means that persons tend to rather accept a given conclusion, if it is valid and/or believable, and to reject a given conclusion, if it is invalid and/or unbelievable. The effect of believability is especially heightened on invalid problems. In the current work, it was at least partly assumed that the ‘belief bias’ effect corresponds to the results of the moral-related problems. Indeed, the results indicate an interaction effect of intuitive and rational processes involved in solving the reasoning problems. Furthermore, the invalid problems were almost always solved worse than the valid ones. This seemed to be in line with the ‘classical belief bias’ effect. Having a more detailed look on the results however, revealed that they did not match this effect. The results of the moral-valid, the moral-invalid, and the unmoral-valid problems were in accordance with it, but those of the unmoral-invalid problems were not. Although these latter problems contradict common moral beliefs and represent invalid conclusions, they were treated similar as the moral-valid ones. In turn, this does not contradict the assumption of prior knowledge/intuitive processes interacting with rational processes while solving the task, and was thus labeled ‘new belief bias’ effect. This interaction is related to the problem inherent congruence-incongruence dichotomy. The question remains however, as to what caused the results, particularly of the unmoral-invalid problems. This issue could be extended, at least partly, to the results regarding the emotional problems, especially the invalid ones. It appears that several factors might have been responsible for the various results of the different experiments. Since different factors contributed to the effects found, no single explanation related to former findings of other studies could account for the current findings in total.

It seems helpful to have a more detailed look on the problems and their respective structure and contents. As shown in Appendix A, a few of the moral-invalid and unmoral-invalid problems contained a negation (i.e., ‘not’) in the conclusion. From former studies it is known that sentences containing a negation require more time to be processed and result in poorer reasoning performance (e.g., Gough, 1966; Just & Carpenter, 1971; Kaup & Zwaan, 2003; Kaup, Zwaan, & Ludtke, 2007). The reading comprehension in particular seems to be affected by negations inherent in the wording of a sentence (Just & Carpenter, 1992). Just and Carpenter (1992) assumed that sentence comprehension depends on the amount of information that could be processed in working memory at the same time. Thus, working memory capacity constrains comprehension. However, comprehension also varies according to individual WM capacity differences (Just & Carpenter, 1992). Since negation might be particularly taxing working memory, more capacity seems necessary to process negated sentences compared to non-negated ones. Therefore, problems containing a negation in the conclusion appeared to show a worse performance than those without a negation. This might also explain why participants with superior intelligence and assumed higher working memory capacity did not reveal such performance decrements. The structure of all three sentences of the reasoning problems shows that a few of them, e.g., the positive emotional and negative emotional invalid ones, also contained a negation in the “then” part of the first premise. Thus, there might have already been a comprehension problem for the participants when reading the premises. Further explanations however, even complicate this picture. As discussed in the introductory part on logic, reasoning problems could be varied according to their logical form by affirming or denying the antecedent or consequent of the conditional in the minor premise, leading to valid or invalid conclusions. A negation in the conditional sentence, i.e., the major premise, can also affect the logical form of the whole problem, and thus the form of the possible conclusions. Evans (1977) investigated persons presenting reasoning problems with affirmed or negated antecedent or consequent of the conditional sentence. He found that participants approved conclusions with a negation rather than those without. This effect depending on the manipulation of negating the antecedent or the consequent of the major premise was originally called ‘conclusion bias’ and thought to reflect a pure response bias (Evans et al., 1993). It affected AC, DA, and MT inferences but not the MP ones (under implication). Later, Evans, Clibbens, and Rood (1995) could show that this bias occurs mainly with DA and MT inferences, but not AC, and labeled it then 'double negation effect'. The 'double negation effect' refers to the fact that participants are unable to deal with negated propositions that lead to affirmative conclusions. This effect might account for a few of the results of the current experiments besides the already mentioned influences of working memory capacity due to negated sentences per se. These factors referred so far, represent a weakness of the reasoning problems applied in the current experiments, which were not controlled and matched for linguistic complexity. Problem complexity influences performance on deductive reasoning (e.g., Monti et al., 2007, 2009) and thus also participants' performance in the current experiments might have been affected due to the different complex reasoning problems applied. As mentioned above, the problems, especially the invalid moral-related ones as well as the invalid emotional ones contained further weaknesses. To match the content of the human rights articles and to create problems with similar wording and contents in all possible forms (e.g., moral and unmoral, valid and invalid), the problems sometimes changed their affiliation to a specific content-condition from the first to the third sentence. For example, a moral major premise resulted in an unmoral conclusion and a positive emotional major premise ended up in a negative emotional conclusion. This might have caused some irritations for the participants while solving the reasoning problems. Another confounding factor, specific to these problems could have resulted in even more dramatically deteriorating effects on reasoning performance. There appear to be content effects within the major premises which could have conflicted with the content effects of the overall problem. A statement like "If a person is happily in love, then she perceives her life as not reasonable." already contains conflicting contents in the major premise. Those statements could immediately cause complicated and conflicting interactional processes between prior knowledge (i.e., intuitive processes) about such propositions and emotional evaluations of these statements as well as rational processes trying to mediate and monitor conflicting thoughts. In turn, this could have also influenced working memory capacity and performance on the reasoning problems. Further points that have been already discussed above shall only be remembered. Decision times for Experiments I, II, and III have to be interpreted with caution, since sentence length was not matched and therefore could not be controlled adequately due to the abstract problems. Furthermore, no incongruent neutral, abstract or emotional control problems were embedded in the experiment. Although the incongruity for the moral-related ones was only chosen to ensure that participants reason practically, this would have ensured better comparability of the results and reduced potential complexity differences between the distinct reasoning conditions. Moreover, no "real" moral reasoning control condition was applied. It is assumed that at least the wrong answers of the moral-related problems relate to the fact that participants reasoned and answered practically instead of theoretically. However, testing moral reasoning directly would have revealed better evidence of the potential processes involved in practical reasoning. Further studies should match the problems so far, so that participants could and should have to answer according to practical and not according to theoretical reasoning (i.e., questioning whether the last sentence is morally correct or not). All of these critics limit and constrain potential interpretations and generalizations of the results. Moreover, four problems per category seem to be very few, resulting in low statistical power and perhaps restricted explanatory variance in the occurring performances. This refers in particular to the moral-related problems where 26 other articles of the Declaration of Human Rights would have been available to be implemented in the current experiments. It remains unclear as to what kind of effects further and/or other articles would have produced. This also caused constraints on the interpretation and potential generalization of the results. Finally, there is one additional critical factor that might have influenced the performance of the reasoning problems, even though it was not inherently based on them. This factor concerns potential influences of the performance of the recognitions items on the performance of the reasoning problems (and vice versa). Although the recognition items applied in experiments III and IV did not influence the groups differently, contributions and influences on the distinct reasoning performances could not be excluded.

Other critics refer to the small sample size of the intelligence experiment or to the assumed training effects which should have only affected rational reasoning processes, but were already discussed in the preceding sections.

In total, these critics, especially on the reasoning problems applied in the preceding experiments, restrict the interpretations and conclusions that could be drawn from the results. The critics indicate some confoundations and weaknesses of the paradigm used. However, since the current experiments represent a first attempt to combine theoretical and practical reasoning research and to create deductive reasoning problems according to relative strictly formulated statements (i.e., human rights articles), these first experiments and related preliminary results might outweigh the critics. In particular, since new, explorative research projects seem to be accompanied often by starting problems and limited potential generalizations of the findings, thus requiring replication and further evidences. So, a first step is made and awaits further contributions to this interesting and interdisciplinary research field.

11. Conclusions and outlook

Apart from the just mentioned critics, it seems possible to draw a few preliminary conclusions with respect to the current findings. First, the experiments yielded evidences that theoretical and practical reasoning are based on common and distinct brain areas and associated reasoning processes. Both reasoning domains appear to require 'rational' and 'emotional/intuitive' cognitive processes and underlying brain areas, namely a fronto-temporo-parietal network. Therefore, 'dual-process' accounts (e.g., Evans, 2008; Goel, 2007; Greene, 2007) seem to explain and to account for theoretical and practical reasoning processes best. Indeed, if both of these reasoning domains are based on similar cognitive processes and related brain structures, a single 'dual-process' model seems sufficient. Furthermore, theoretical and practical reasoning processes depend on and are influenced by cognitive abilities (e.g., intelligence), ToM related processes, and working memory capacity. In addition, morals seem to be learned, at least partly according to the contents of the articles of the 'Universal Declaration of Human Rights', perhaps stored as 'somatic markers' (Damasio, 2006). Thus, the preceding experiments also reveal evidence for the applicability of these articles to investigate moral reasoning with validated items according to an (almost) worldwide accepted standard.[19] Finally, the current experiments provide first support that theoretical and practical reasoning could and should be investigated together, as has been suggested by Evans (2008) for example.

The experiments reported here revealed first evidence for most of the hypotheses and assumptions derived from the literature on theoretical and practical reasoning research findings. Although these evidences were obtained using different experimental approaches and investigating numerous and various participants, an important refinement has to be mentioned. There was no big variation in the items applied (i.e., only 'simple' deductive reasoning problems filled with the content of four different human rights articles were used). Thus, it seems apparently clear that further experiments should implement different reasoning tasks as well as further human rights articles to obtain additional evidences for the up to now preliminary conclusions drawn above. Furthermore, as noted above, moral control tasks might be used as well as matched incongruent problems for all conditions to exclude potential critics or alternative explanations. Moreover, problem complexity and problem inherent content

[19] The Declaration of Human Rights is not a philosophical proof or logical derivation embedded in a comprehensive theory according to philosophical demands. However, the lack of such a theory forces scientists to use the best material available and a worldwide agreement seems to be more appropriate than any speculative and not evaluated approaches (e.g., dilemmas). In turn, the influences of experimental research findings on morals and ethics has to be discussed, but is far beyond the scope of this thesis (for further discussion of these issues, see for example Greene, 2003).

effects have to be controlled or, if necessary and possible, be eliminated. However, to postulate common reasoning processes and underlying brain areas for theoretical and practical reasoning (besides distinct ones), which vary and are varied by and from different factors (e.g., intelligence, working memory capacity, ToM, task complexity, task contents), instead of proposing different reasoning processes and brain areas for each potential sub-domain of reasoning, seems to match proposals of other researchers in the field (e.g., Damasio, 2006, LeDoux, 2006, Thagard, 2006) as well as new suggestions made for example in the memory research domain. New ideas suggest that material stored in long-term memory (LTM) and activated by a current task requires or even produces no new storage areas/systems between LTM and working memory but just the known associative storage areas/systems interact with the process-guiding, manipulative ones (e.g., Postle, 2006). Therefore, it seems prominent to take the current finding as a primary step for further investigations of the human reasoning system. Not only to explore its distinct, but also its common reasoning processes and underlying brain areas involved (i.e., the ‘core units’ of human reasoning involved in all reasoning domains).

Bibliography

Adolphs, R. (1999). Social cognition and the human brain. *Trends in Cognitive Sciences, 3,* 469-479.

Adolphs, R. (2006). The social brain. *Engineering & Science, 1,* 2-8.

Allison, T. (2001). Neuroscience and morality. *The Neuroscientist*, *7*, 360-364.

Amelang, M., & Bartussek, D. (1990). *Differentielle Psychologie und Persönlichkeitsforschung.* Stuttgart: Kohlhammer.

Amthauer, R. (1953). *I-S-T Intelligenz-Struktur-Test* [I-S-T Intelligence-structure-test]. Göttingen: Hogrefe.

Amthauer, R. (1973). *Intelligenzstrukturtest I-S-T 70* [Intelligencestructuretest I-S-T 70]. Göttingen: Hogrefe.

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology, 6,* 451-474.

Anderson, J. R. (1999). *Learning and memory: An integrated approach.* Hoboken: John Wiley & Sons, Inc.

Anderson, S. W., Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1999). Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nature Neuroscience*, *2,* 1032-1037.

Aristoteles (2002). *Die Nikomachische Ethik* [Nicomachean Ethics] (O. Gigon, Trans.). München: Deutscher Taschenbuch Verlag.

Baddeley, A. D. (1986). *Working memory.* Oxford: Clarendon Press.

Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4,* 829-839.

Baddeley, A. (2007). *Working memory, thought, and action.* New York: Oxford University Press.

Baddeley, A. D., Emslie, H., Kolodny, J., & Duncan, J. (1998). Random generation and the executive control of working memory. *Quaterly Journal of Experimental Psychology, 51,* 818–852.

Baddeley, A. D., Eysenck, M. W., & Anderson, M. C. (2009). *Memory.* New York: Psychology Press.

Bara, B. Bucciarelli, M., & Johnson-Laird, P. (1995). The development of syllogistic reasoning. *American Journal of Psychology, 108,* 157-193.

Barrouillet, P., & Lecas., J.-F. (1999). Mental models in conditional reasoning and working memory. *Thinking and Reasoning, 5,* 289-302.

Bentham, J. (1789/2007). *An introduction to the principles of morals and legislation.* Oxford: Clarendon Press.

Berzonsky, M. D., & Ondrako, M. A. (1974). Cognitive style and logical deductive reasoning. *The Journal of Experimental Education, 43*, 18-24.

Björklund, F. (2003). Differences in the justification of choices in moral dilemmas: Effects of gender, time pressure and dilemma seriousness. *Scandinavian Journal of Psychology, 44,* 459-466.

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

Blanchette, I., & Campbell, M. (2005). The effect of emotion on syllogistic reasoning in a group of war veterans. In *Proceedings of the XXVIIth Annual Conference of the Cognitive Science Society* (p. 1401). Mahwah, NJ: Erlbaum.

Blanchette, I., & Richards, A. (2004). Reasoning about emotional and neutral materials: Is logic affected by emotion? *Psychological Science, 15,* 745-752.

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.

Bore, M., Munro, D., Kerridge, I., & Powis, D. (2005). Selection of medical students according to their moral orientation. *Medical Education*, *39,* 266-275.

Braine, M. D. S., & O'Brien, D. P. (1998). *Mental Logic.* Mahwah, NJ: Erlbaum.

Braine, M. D. S., O'Brien, D. P., Noveck, I. A., Samuels, M. C., Lea, R. B., Fisch, S. M., & Yang, Y. (1995). Predicting intermediate and multiple conclusions in propositional logic inference problems: Further evidence for a mental logic. *Journal of Experimental Psychology: General*, *124,* 263-292.

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

Bucciarelli, M., Khemlani, S., & Johnson-Laird, P. N. (2008). The psychology of moral reasoning. *Judgment and Decision Making*, *3,* 121-139.

Canessa, N., Gorini, A., Cappa, S. F., Piattelli-Palmarini, M., Danna, M., Fazio, F., & Perani, D. (2005). The effect of social content on deductive reasoning: An fMRI study. *Human Brain Mapping, 26,* 30-43.

Capon, A., Handley, S., & Dennis, I. (2003). Working memory and reasoning: An individual differences perspective. *Thinking and Reasoning*, *9,* 203-244.

Carrington, S. J., & Bailey, A. J. (2009). Are there Theory of Mind regions in the brain? A review of the neuroimaging literature. *Human Brain Mapping*, *30*, 2313-2335.

Casebeer, W. D. (2003a). Moral cognition and its neural constituents. *Nature Reviews Neuroscience, 4,* 841-846.

Casebeer, W. D. (2003b). *Natural ethical facts: Evolution, connectionism, and moral cognition.* The MIT Press.

Casebeer, W. D., & Churchland P. S. (2003). The neural mechanisms of moral cognition: A multiple-aspect approach to moral judgment and decision-making. *Biology and Philosophy*, *18,* 169-194.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, *40,* 153-193.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54,* 1-22.

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

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

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

Churchland, P. M. (1998). Toward a cognitive neurobiology of the moral virtues. *Topoi, 17,* 83-96.

Ciaramelli, E., Muccioli, M., Làdavas, E., & Di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience*, *2*, 84-92.

Colom, R., García, L. F., Juan-Espinosa¸ M., & Abad, F. J. (2002). Null Sex Differences in General Intelligence: Evidence from the WAIS-III. *The Spanish Journal of Psychology, 5,* 29-35.

Colom, R., Rebollo, I., Palacios, A., Juan-Espinosa, M., & Kyllonen, P. C. (2004). Working memory is (almost) perfectly predicted by g. *Intelligence, 32,* 277-296.

Copeland, D. E., & Radvansky, G. A. (2004). Working memory and syllogistic reasoning. *The Quarterly Journal of Experimental Psychology*, 5*7*, 1437-1457.

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 effects of experience on performance in Wason's selection task. *Memory & Cognition, 10,* 496-502.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing. A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671-684.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*, 268-294.

Damasio, A. R. (2006). *Descartes' error: Emotion, reason and the human brain.* London: Vintage.

Declaration of Human Rights. Retrieved September 27, 2008, 10.28 am, from http://un.org/Overview/rights.html.

De Neys, W., Schaeken, W., & D'Ydewalle, G. (2005). Working memory and everyday conditional reasoning: Retrieval and inhibition of stored counterexamples. *Thinking & Reasoning*, *11*, 349-381.

Der, G., & Deary, I. J. (2006). Age and sex differences in reaction time in adulthood: Results from the United Kingdom Health and Lifestyle Survey. *Psychology and Aging, 21,* 62-73.

Dolan, R. J. (2002). Emotion, cognition, and behavior. *Science*, *298*, 1191-1194.

Eickhoff, S. B., Laird, A. R., Grefkes, C., Wang, L. E., Zilles, K., & Fox, P. T. (2009). Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty. *Human Brain Mapping*, *30*, 2907-2926.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11,* 19-23.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology*, *128,* 309-331.

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

Evans, J. St. B. T. (1989). *Bias in human reasoning.* Hove, UK: Lawrence Erlbaum Associates.

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

Evans, J. St. B. T., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition, 11*, 295-306.

Evans, J. St. B. T., Clibbens, J., & Rood, B. (1995). Bias in conditional inference: Implications for mental models and mental logic. *Quaterly Journal of Experimental Psychology, 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., & Feeney, A. (2004). Belief structures and reasoning. In R. J. Sternberg, & J. Leighton (Eds.), *The Nature of Reasoning* (pp. 78-102). New York: Cambridge University Press.

Evans, J. St. B. T., Handley, S. J., & Bacon, A. M. (2009). Reasoning under time pressure: A study of causal conditional inference. *Experimental Psychology*, *56,* 77-83.

Evans, J. St. B. T., Handley, S. J., Neilens, H., & Over, D. E. (2007). Thinking about conditionals: A study of individual differences. *Memory & Cognition, 35,* 1772-1784.

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

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

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

Fahrenberg, J., Hampel, R., & Selg, H. (2001). *FPI-R. Das Freiburger Persönlichkeitsinventar* (7., überarbeitete und neu normierte Auflage) [Freiburger personality inventory]. Göttingen, Bern, Toronto: Hogrefe Verlag.

Fangmeier, T., Knauff, M., Ruff, C. C., & Sloutsky, V. (2006). FMRI evidence for a three-stage model of deductive reasoning. *Journal of Cognitive Neuroscience, 18,* 320-334.

Ferré, P. (2003). Effects of level of processing on memory for affectively valenced words. *Cognition and Emotion, 17,* 859-880.

Fiddick, L., Spampinato, M. V., & Grafman, J. (2005). Social contracts and precautions activate different neurological systems: An fMRI investigation of deontic reasoning. *NeuroImage, 28,* 778-786.

Finger, E. C., Marsh, A. A., Kamel, N., Mitchell, D. G. V., & Blair, J. R. (2006). Caught in the act: The impact of audience on the neural response to morally and socially inappropriate behavior. *NeuroImage, 33,* 414-421.

Friston, K. J. (2005). Models of Brain Function in Neuroimaging. *Annual Reviews of Psychology, 56*, 57-87.

Frith, C. D., & Frith, U. (1999). Interacting minds – a biological basis. *Science, 286,* 1692-1695.

Fry, A. F., & Hale, S. (2000). Relationships among processing speed, working memory, and fluid intelligence in children. *Biological Psychology, 54,* 1-34.

Gangemi, A., Mancini, F., & Johnson-Laird, P. N. (2006, September). Reasoning in obsessive compulsive disorder. Paper presented at the annual conference of the British Psychological Society. Lancaster, England.

Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (Eds.) (2009). *Cognitive Neuroscience: the biology of the mind (3rd edition)*. New York, London: Norton & Company.

Gigerenzer, G., & Hug, K. (1992). Domain-specific reasoning: Social contracts, cheating, and perspective change. *Cognition, 43,* 127-171.

Gilhooly, K. J., Logie, R. H., Wetherick, N. E., & Wynn, V. (1993). Working memory and strategies in syllogistic-reasoning tasks. *Memory & Cognition, 81,* 115-124.

Gilhooly, K. J., Logie, R. H., & Wynn, V. E. (2002). Syllogistic reasoning tasks and working memory: Evidence from sequential presentation of premises. *Current Psychology: Developmental, Learning, Personality, Social, 21*, 111-120.

Gilligan, C. (1982). *In a different voice: Psychological theory and women's development*. Cambridge, MA: Harvard University Press.

Goel, V. (2003). Evidence for dual neural pathways for syllogistic reasoning. *Psychologica, 32,* 301-309.

Goel, V. (2007). Anatomy of deductive reasoning. *Trends in Cognitive Sciences, 11,* 435-441.

Goel, V, Büchel, C. Frith, C., & Dolan, R. J. (2000). Dissociation of mechanisms underlying syllogistic reasoning. *NeuroImage, 12,* 504-514.

Goel, V., & Dolan, R. J. (2003a). Reciprocal neural response within lateral and ventral medial prefrontal cortex during hot and cold reasoning. *NeuroImage, 20,* 2314-2321.

Goel, V., & Dolan, R. J. (2003b). Explaining modulation of reasoning by belief. *Cognition, 87,* B11-B22.

Goel, V., Shuren, J., Sheesley, L., & Grafman, J. (2004). Asymmetrical involvement of frontal lobes in social reasoning. *Brain: A Journal of Neurology, 127,* 783-790.

Goodenough, O. R., & Prehn, K. (2004). A neuroscientific approach to normative judgment in law and justice. *Philosophical transactions of the Royal Society of London. Series B: Biological Sciences*, *359,* 1709-1726.

Gough, P. B. (1966). The verification of sentences: The effects of delay of evidence and sentence length. *Journal of verbal Learning and verbal Behavior, 5,* 492-496.

Greene, J. D. (2003). From neural 'is' to moral 'ought': What are the moral implications of neuroscientific moral psychology? *Nature Reviews Neuroscience*, *4,* 847-850.

Greene, J. D. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences, 11,* 322-323.

Greene, J. D., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences*, *6,* 517-523.

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition, 107,* 1144-1154.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron, 44,* 389-400.

Greene, 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.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*, 1464-1480.

Griggs, R. A., & Cox, J. R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology, 73,* 407-420.

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

Hare, R. M. (1981). *Moral thinking: Its levels, methods and point.* New York: Oxford University Press.

Harenski, C. L., & Hamann, S. (2006). Neural correlates of regulating negative emotions related to moral violations. *NeuroImage*, *30,* 313-324.

Hauser, M. D. (2006). The liver and the moral organ. *Social Cognitive and Affective Neuroscience*, *1,* 214-220.

Hauser, M. D. (2007). *Moral minds: The nature of right and wrong.* New York: Harper Collins Publishers.

Heekeren, H. R., Wartenburger, I., Schmidt, H., Prehn, K., Schwintowski, H.-P., & Villringer, A. (2005). Influence of bodily harm on neural correlates of semantic and moral decision-making. *NeuroImage*, *24,* 887-897.

Heekeren, H., Wartenburger, I., Schmidt, H., Schwintowski, H.-P., & Villringer, A. (2003). An fMRI study of simple ethical decision-making. *NeuroReport, Cognitive Neuroscience and Neuropsychology*, *14,* 1215-1219.

Hossiep, R., Turck, D., & Hasella, M (1999). *BOMAT - advanced - Bochumer Matrizentest.* Göttingen: Hogrefe.

Hoyningen-Huene, P. (2006). *Formale Logik: Eine philosophische Einführung* [Formal logic: A philosophical introduction]. Ditzingen: Reclam.

Hume, D. (1748/2002). *Eine Untersuchung über den menschlichen Verstand* [An enquiry concerning human understanding] (H. Herring, Trans.). Ditzingen: Reclam.

Hume, D. (1751/2003). *Eine Untersuchung über die Prinzipien der Moral* [An enquiry concerning the principles of morals] (M. Kühn, Trans). Hamburg: Meiner Verlag.

Hutcheson, F. (1755/1968). *A system of moral philosophy.* New York: Kelley.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science*, *321,* 494-495.

Inaba, M., Nomura, M., & Ohira, H. (2005). Neural evidence of effects of emotional valence on word recognition. *International Journal of Psychophysiology, 57*, 165-173.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30,* 513-541.

Jacoby, L. L., Woloshyn, V., & Kelley, C. (1989). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. In D. Balota, A. David, & E. J. Marsh (Eds.), *Cognitive psychology: Key readings* (pp. 322-337). New York: Psychology Press.

Jäncke, L. (2005). *Methoden der Bildgebung in der Psychologie und den kognitiven Neurowissenschaften* [Imaging methods in psychology and cognitive neuroscience]. Stuttgart: Kohlhammer.

Janssen, J., & Laatz, W. (2005). *Statistische Datenanalyse mit SPSS für Windows* [Statistical data analysis with SPSS for Windows]. Heidelberg: Springer.

Jeffrey, R. N. (1981). *Formal Logic (2nd Ed)*. New York: McGraw-Hill.

Johnson-Laird, P. N. (1983). *Mental models. Towards a cognitive science of language, inference, and consciousness.* Cambridge, UK: University Press.

Johnson-Laird, P. N. (2006). How we reason. Oxford: Oxford University Press.

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

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

Johnson-Laird, P. N., Legrenzi, P., & Legrenzi, M. S. (1972). Reasoning and a sense of reality. *British Journal of Psychology, 63,* 395-400.

Just, M. A., & Carpenter, P. A. (1971). Comprehension of negation with quantification. *Journal of verbal Learning and verbal Behavior, 10,* 244-253.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122-149.

Kandl, M. (2009). *The impact of training on logical reasoning tasks with moral and neutral content.* Universität Gießen: Unveröffentlichte Diplomarbeit.

Kant, I. (1781/1998). *Kritik der reinen Vernunft* [Critique of pure reason]. Ditzingen: Reclam.

Kant, I. (1785/2000). *Grundlegung zur Metaphysik der Sitten* [Groundwork of the metaphysics of morals]. Ditzingen: Reclam.

Kant, I. (1797/2007). *Die Metaphysik der Sitten* [Metaphysics of morals]. Ditzingen: Reclam.

Kant, I. (1788/2008). *Kritik der praktischen Vernunft* [Critique of practical reason]. Ditzingen: Reclam.

Kaup, B., & Zwaan, R. A. (2003). Effects of negation and situational presence on the accessibility of text information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 439-446.

Kaup, B., Zwaan, R. A., & Ludtke, J. (2007). The experiential view of language comprehension: How is negation represented? In F. Schmalhofer, C. Perfetti, & A. Charles (Eds.), *Higher level language processes in the brain: Inference and comprehension processes.* (pp. 255-288). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Kensinger, E. A., & Corkin, S. (2003). Effect of negative emotional content on working memory and long-term memory. *Emotion, 3*, 378-393.

Klaczynski, P. A., & Laipple, J. S. (1993). Role of content domain, logic training, and IQ in rule acquisition and transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 653-672.

Klauer, K. C., & Meiser, T. (2007). Training effects in deductive reasoning: A theory-based review. In W. Schaeken, A. Vandierendonck, W. Schroyens, & G. D'Ydewalle (Eds.), *The mental models theory of reasoning: Refinements and extensions* (pp. 209-224). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Klauer, K. C., Meiser, T., & Naumer, B. (2000). Training propositional reasoning. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 53A,* 868-895.

Klauer, K. C., Stegmaier, R., & Meiser, T. (1997). Working memory involvement in propositional and spatial reasoning. *Thinking & Reasoning, 3*, 9-47.

Knauff, M. (2009a). Reasoning. In: M. D. Binder, N. Hirokawa, U. Windhorst, & M. C. Hirsch (Eds.), *Encyclopedia of Neuroscience* (pp. 3377-3382). Berlin: Springer.

Knauff, M. (2009b). A Neuro-Cognitive Theory of Deductive Relational Reasoning with Mental Models and Visual Images. *Spatial Cognition and Computation*, *9*, 109-137.

Knauff, M. (2007). How our brains reason logically. *Topoi*, *26*, 19-36.

Knauff, M., Fangmeier, T., Ruff, C. C., & Johnson-Laird, P. N. (2003). Reasoning, models, and images: Behavioral measures and cortical activity. *Journal of Cognitive Neuroscience, 15*, 559-573.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefontal cortex increases utilitarian moral judgements. *Nature, 446*, 908-911.

Kohlberg, L. (1969). Stage and sequence: The cognitive-development approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research.* Chicago: McNally.

Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., Kennedy, D. N., Hoppel, B. E., Cohen, M. S., Turner, R., Cheng, H., Brady, T. J., & Rosen, B. R. (1992). Dynamic Magnetic Resonance Imaging of Human Brain Activity During Primary Sensory Stimulation. *Proceedings of the National Academy of Sciences, 89*, 5675-5679.

Laird, A. R., Fox, M., Price, C. J., Glahn, D. C., Uecker, A. M., Lancaster, J. L., Turkeltaub, P. E., Kochunov, P., & Fox, P. T. (2005). ALE meta-analysis: Controlling the false discovery rate and performing statistical contrasts. H*uman Brain Mapping, 25*, 155-164.

Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., Kochunov, P. V., Nickerson, D., Mikiten, S. A., Fox, P. T. (2000). Automated Talairach Atlas labels for functional brain mapping. *Human Brain Mapping*, *10*, 120-131.

Lang, P. J., Bradley M. M., & Cuthbert, B. N. (1995). *International Affective Picture System (IAPS)*. Bethesda, MD: National Institute of Mental Health, Center for the Study of Emotion and Attention.

Lang, P. J., & Greenwald, M. K. (1985). *The International Affective Picture System Slides and Technical Report.* Gainesville, Florida: Center for Research in Psychophysiology.

LeDoux, J. (2006). *Das Netz der Gefühle: Wie Emotionen entstehen* [Emotional brain: The mysterious underpinnings of emotional life]. München: Deutscher Taschenbuch Verlag.

Lefford, A. (1946). The influence of emotional subject matter on logical reasoning. *Journal of General Psychology, 34,* 127-151.

Lieberman, M. D. (2003). Reflexive and reflective judgment processes: A social cognitive neuroscience approach. In J. P. Forgas, K. R. Williams, & W. von Hippel (Eds.), *Social Judgments: Implicit and Explicit Processes* (pp. 44-67). New York: Cambridge University Press.

Liepmann, D., Beauducel, A., Brocke, B., & Amthauer, R. (2001). *Intelligenz-Struktur-Test 2000 Revision (IST-2000-R) [Intelligence-Structure-Test]*. Göttingen: Hogrefe.

Lind, G. (1978). Wie mißt man moralisches Urteil? [How to measure moral judgment?] In G. Portele (Ed.), *Sozialisation und Moral: Neue Ansätze zur moralischen Entwicklung und Erziehung* [Socialization and morality: New approaches to moral development and education]. Weinheim: Beltz.

Lind, G. (2008). The meaning and measurement of moral judgment competence revisited – A dual-aspect model. In: D. Fasko, & W. Willis (Eds.), *Contemporary philosophical and* *psychological perspectives on moral development and education* (pp. 185-220). Cresskill, NJ: Hampton Press.

Lind, G., Grocholewska, K., & Langer, J. (1987). Haben Frauen eine andere Moral? Eine empirische Untersuchung von Studentinnen und Studenten in Österreich, Deutschland und Polen [Do women possess another morals? An empirical investigation of students from Austria, Germany, and Poland]. In L. Unterkirchner, & I. Wagner (Eds.), *Die andere Hälfte der Gesellschaft. Sozio-logische Befunde zu geschlechtsspezifischen Formen der Lebensbewältigung [The other half of society. Sociological findings on gender specifics of coping with life* (pp. 394-406). Wien: Verlag des Österr. Gewerkschaftsbundes.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature, 412,* 150-157.

Luo, Q., Nakic, M., Wheatly, T., Richell, R., Martin, A., & Blair, J. R. (2006). The neural basis of implicit moral attitude – An IAT study using event-related fMRI. *NeuroImage, 30,* 1449-1457.

Lynn, R., Backhoff, E., & Contreras-Niňo, L. A. (2004). Sex differences on g, reasoning and visualisation tested by the progressive matrices among 7-10 year olds: Some normative data for Mexico. *Personality and Individual Differences, 36,* 779-787.

Mandler, G. (1980). Recognizing: the judgment of previous occurrence. *Psychological Reviews, 87,* 252-271.

Manktelow, K. I. (2004). *Reasoning and thinking.* Hove, UK: Psychology Press.

Manktelow, K. I. & Evans, J. St. B. T. (1979). Faciliation of reasoning by realism: Effect or non-effect? *British Journal of Psychology, 70,* 477-488.

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

Manktelow, K. I., & Over, D. E. (1995). Deontic reasoning. In S. Newstead, & J. St. B. T. Evans (Eds.), *Perspectives on thinking and reasoning: Essays in honour of Peter Wason.* Hove, UK: Lawrence Erlbaum.

Markovits, H., Doyon, C., & Simoneau, M. (2002). Individual differences in working memory and conditional reasoning with concrete and abstract content. *Thinking and Reasoning, 8,* 97-107.

Meiser, T., Klauer, K. C., & Naumer, B. (2001). Propositional reasoning and working memory: The role of prior training and pragmatic content. *Acta Psychologica, 106,* 303-327.

Mendez, M. F., Chen, A. K., Shapira, J. S., & Miller, B. L. (2005). Acquired sociopathy and frontotemporal dementia. *Dementia and Geriatric Cognitive Disorders, 20,* 99-104.

Middendorf, J., & Kalish, A. (1996). The “Change-Up” in lectures. *The National Teaching and Learning Forum, 5,* 1-7.

Mikhail, J. M. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in Cognitive Sciences, 11,* 143-152.

Mill, J. S. (1863/1994). *Der Utilitarismus* [Utilitarianism] (D. Birnbacher, Trans.). Ditzingen: Reclam.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *The Psychological Review, 63,* 81-97.

Miller, P. (1993). *Theorien der Entwicklungspsychologie [Theories of developmental psychology]* (A. Hildebrandt-Essig, Trans.). Heidelberg, Berlin, Oxford: Spektrum Akademischer Verlag.

Moll, J., Eslinger, P. J., & De Oliveira-Souza, R. (2001). Frontopolar and anterior temporal cortex activation in a moral judgment task: Preliminary functional MRI results in normal subjects. *Arquivos de Neuropsiquiatria, 59,* 657-664.

Moll, J., De Oliveira-Souza, R., Bramati, I. E., & Grafman, J. (2002). Functional networks in emotional moral and nonmoral social judgments. *NeuroImage, 16,* 696-703.

Moll, J., De Oliveira-Souza, R., & Eslinger, P. J. (2003). Morals and the human brain: a working model. *NeuroReport*, *14*, 299-305.

Moll, J., De Oliveira-Souza, R., Eslinger, P. J., Bramati, I. E., Mourão-Miranda, J., Andreiuolo, P. A., & Pessoa, L. (2002). The neural correlates of moral sensitivity: A functional magnetic resonance imaging investigation of basic and moral emotions. *The Journal of Neuroscience, 22,* 2730-2736.

Moll, J., De Oliveira-Souza, R., Moll, F. T., Ignacio, F. A., Bramati, I. E., Caparelli-Daquer, E. M., & Eslinger, P. J. (2005). The moral affiliations of disgust. A functional MRI study. *Cognitive Behavioral Neurology*, *18,* 68-78.

Moll, J., Zahn, R., De Oliveira-Souza, R., Krueger, F., & Grafman, J. (2005). The neural basis of human moral cognition. *Nature Reviews Neuroscience*, *6,* 799-809.

Monin, B., Pizarro, D. A., & Beer, J. S. (2007). Deciding versus reacting: Conceptions of moral judgment and the reason-affect debate. *Review of General Psychology, 11,* 99-111.

Monti, M. M., Parsons, L. M., Martinez, M. J., & Osherson, D. N. (2007). Functional Neuroanatomy of Deductive Inference: A Language-Independent Distributed Network. *NeuroImage, 37*, 1005-1016.

Monti, M. M., Parsons, L. M., & Osherson, D. N. (2009). The Boundaries of Language and Thought in Deductive Inference. *Proceedings of the National Academy of Sciences*. [EPub ahead of print].

Moore, A. B., Clark, B. A., & Kane, M. J. (2008). Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science*, *19,* 549-557.

Mukamel, R., Gelbard, H., Arieli, A., Hasson, U., Fried, I., & Malach, R. (2005). Coupling Between Neuronal Firing, Field Potentials, and fMRI in Human Auditory Cortex. *Science, 309*, 951-954.

Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J.-B. (2005). Valid conjunction inference with the minimum statistic. *NeuroImage, 25*, 653-660.

Nichols, S., & Mallon, R. (2006). Moral dilemmas and moral rules. *Cognition*, *100,* 530-542.

Norman, D. A., & Shallice, T. (1986). In R. J. Davidson, G. E. Schwarts, & D. Shapiro (Eds.), *Consciousness and Selfregulation. Advances in Research and Theory* (pp. 1-18). New York: Plenum.

Noveck, I. A., Goel, V., & Smith, K. W. (2004). The neural basis of conditional reasoning with arbitrary content. *Cortex, 40,* 613-622.

Oatley, K., & Johnson-Laird, P. N. (1996). The communicative theory of emotions: Empirical tests, mental models, and implications for social interaction. In L. L. Martin, & A. Tesser (Eds.), *Striving and feeling: Interactions among goals, affect, and self-regulation*. (pp. 363-393). Mahwah, NJ: Erlbaum Associates.

O'Brien, D., & Overton, W. F. (1980). Conditional reasoning following contradictory evidence: A developmental analysis. *Journal of Experimental Child Psychology, 30,* 44-61.

O'Brien, D., & Overton, W. F. (1982). Conditional reasoning and the competence-performance issue: A developmental analysis of a training task. *Journal of Experimental Child Psychology, 34,* 274-290.

Ogawa, S., Lee, T. M., Nayak, A. S., & Glynn, P. (1990). Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magnetic Resonance in Medicine*, *14*, 68-78.

Ohira, H., Winton, W. M., & Oyama, M. (1998). Effects of stimulus valence on recognition memory and endogenous eyeblinks: Further evidence for positive-negative asymmetry. *Personality and Social Psychology Bulletin, 24,* 986-993.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia*, *9*, 97-113.

Overton, W., Byrnes, J. P., & O'Brien, D. P. (1985). Developmental and individual differences in conditional reasoning: The role of contradiction training and cognitive style. *Developmental Psychology, 21,* 692-701.

Pasupathi, M., & Staudinger, U. M. (2001). Do advanced moral reasoners also show wisdom? Linking moral reasoning and wisdom-related knowledge and judgement. *International Journal of Behavioral Development, 25,* 401-415.

Peterson, S. B., & Potts, G. R. (1982). Global and specific components of information integration. *Journal of Verbal Learning & Verbal Behavior, 21,* 403-420.

Phelps, E. A., LaBar, K. S., & Spencer, D. D. (1997). Memory for emotional words following unilateral temporal lobectomy. *Brain and Cognition, 35,* 85-109.

Piaget, J. (1932). *The moral judgment of the child.* New York: Free Press.

Pirolli, P. L., & Anderson, J. R. (1985). The role of practice in fact retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 136-153.

Pizarro, D. A., Uhlmann, E., & Bloom, P. (2003). Causal deviance and the attribution of moral responsibility. *Journal of Experimental Social Psychology*, *39,* 653-660.

Pizarro, D. A., Uhlmann, E., & Salovey, P. (2003). Asymmetry in judgments of moral blame and praise: The role of perceived metadesires. *Psychological Science, 14,* 267-272.

Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience, 139,* 23-38.

Prehn, K., Wartenburger, I., Mériau, K., Scheibe, C., Goodenough, O. R., Villringer, A., van der Meer, E., & Heekeren, H. R. (2008). Individual differences in moral judgment competence influence neural correlates of socio-normative judgments. *Social Cognitive and Affective Neuroscience, 3,* 33-46.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*, 515-526.

Raichle, M. E. (1994). Visualizing the Mind: Strategies of cognitive science and techniques of modern brain imaging open a window to the neural systems responsible for thought. *Scientific American*, *270,* 36-42.

Raichle, M. E., & Mintun, M. A. (2006). Brain Work and Brain Imaging. *Annual Reviews Neuroscience, 29*, 449-476.

Raine, A., & Yang, Y. (2006). Neural foundations to moral reasoning and antisocial behavior. *Social Cognitive and Affective Neuroscience*, *1,* 203-213.

Ramsey, N. F., Jansma, J. M., Jager, G., Van Raalten, T. R., & Kahn, R. S. (2004). Neurophysiological factors in human information processing capacity. *Brain, 127*, 517-525.

Reber, A. S. (1993). *Implicit learning and tacit knowledge*. Oxford, UK: Oxford University Press.

Rest, J. R. (1974). *Manual for the defining issue test: An objective test for moral judgment development.* Minneapolis: University of Minneapolis Press.

Rest, J. R. (1979). *Development in judging moral issues.* Minneapolis: University of Minnesota Press.

Rips, L. J. (1994). *The psychology of proof.* Cambridge, MA: MIT Press.

Rorden, C., & Brett, M. (2000). Stereotaxic display of brain lesions. *Behavioural Neurology, 12*, 191-200.

Ruff, C. C., Knauff, M., Fangmeier, T., & Spreer, J. (2003). Reasoning and working memory: Common and distinct neuronal processes. *Neuropsychologia, 41*, 1241-1253.

Salmaso, D., & Longoni, A. M. (1985). Problems in the assessment of hand preference. *Cortex, 21*, 533-549.

Samuels, R. (2006). The magical number two, plus or minus: some comments on dual-processing theories of cognition. *In Two Minds: Dual Process Theories of Reasoning and Rationality.* Intl. Conf., Cambridge, UK, July 5-7.

Sandkühler, H. J., (Ed.). (1999). *Enzyklopädie Philosophie* [Encyclopedia of Philosophy]. Hamburg: Meiner Verlag.

Schaich Borg, J., Hynes, C., Van Horn, J., Grafton, S., & Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments: An fMRI investigation. *Journal of Cognitive Neuroscience, 18,* 803-817.

Schmoranzer, A. (2009). *Welchen Einfluss hat Intelligenz auf moralisches Urteilen? Eine Untersuchung des Zusammenhangs anhand von Logikaufgaben mit moralischem und neutralem Inhalt.* Universität Gießen: Unveröffentlichte Diplomarbeit.

Shaftesbury, L. (1711/1963). *Characteristics of men, manners, opinions, times.* Gloucesters, MA: Smith.

Skoe, E. E. A., Eisenberg, N., & Cumberland, A. (2002). The role of reported emotion in real-life and hypothetical dilemmas. *Personality and Social Psychology Bulletin, 28,* 962-973.

Smith, A. (1759/ 2004). Theorie der ethischen Gefühle [The theory of moral sentiments] (W. Eckstein, Trans.). Hamburg: Meiner Verlag.

Squire, L. R., John T. Wixted, J. T., & Clark, R. E. Recognition memory and the medial temporal lobe: a new perspective. *Nature Reviews Neuroscience, 8,* 872-883.

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning.* Mawah, NJ: Lawrence Erlbaum Associates.

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General, 127,* 161-188.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences, 23,* 645-726.

Stojiljković, S. (1998). Personality characteristics and moral judgement. *Facta Universitatis, 1,* 507-514.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643-662.

Talairach, J., & Tournoux, P. (1988). *A Co-Planar Stereotaxic Atlas of the Human Brain*. New York: Thieme.

Thagard, P. (2006). *Hot thought: Mechanisms and applications of emotional cognition.* Cambridge, UK: The MIT Press.

Theysohn, J. M., Maderwald, S., Kraff, O., Moenninghoff, C., Ladd, M. E., & Ladd, S. C. (2008). Subjective acceptance of 7 Tesla MRI for human imaging. *Magnetic Resonance Materials in Physics, Biology and Medicine, 21,* 63-72.

Thomsen, K., Offenhauser, N., & Lauritzen, M. (2004). Principal neuron spiking: neither necessary nor sufficient for cerebral blood flow in rat cerebellum. *Journal of Physiology, 560,* 181-189.

Toms, M., Morris, N., & Ward, D. (1993). Working memory and conditional reasoning. *The Quarterly Journal of Experimental Psychology, 46*, 679-699.

Tugendhat, E., & Wolf, U. (2004). *Logisch-semantische Propädeutik* [Logical and semantic propaedeutics]. Ditzingen: Reclam.

Tulving, E. (1985). How many memory systems are there? *American Psychologist, 40,* 385-398.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology, 53,* 1-25.

Tun, P. A., & Lachman, M. E. (2008). Age differences in reaction time and attention in a national telephone sample of adults: Education, sex, and task complexity matter. *Developmental Psychology, 44,* 1421-1429.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, F., Crivello, F., Etard, O, Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage, 15,* 273-289.

Van Raalten, T. R., Ramsey, N. F., Duyn, J., & Jansma, J. M. (2008). Practice induces function-specific changes in brain activity. *PLoS ONE, 3*, e3270.

Verschueren, N., Schaeken, W., & D'Ydewalle, G. (2005a). A dual-process specification of causal conditional reasoning. *Thinking & Reasoning, 11*, 239-278.

Verschueren, N., Schaeken, W., & D'Ydewalle, G. (2005b). Everyday conditional reasoning: A working memory-dependent tradeoff between counterexample and likelihood use. *Memory & Cognition, 33,* 107-119.

Viskontas, I. V, Morrison, R. G, Holyoak, K. J, Hummel, J. E., & Knowlton, B. J. (2004). Relational integration, inhibition, and analogical reasoning in older adults. *Psychology and Aging, 4,* 581-591.

Waldmann, M. R., & Dieterich, J. H. (2007). Throwing a bomb on a person versus throwing a person on a bomb: Intervention myopia in moral intuitions. *Psychological Science, 18,* 247-253.

Walter, B., Blecker, C., Kirsch, P., Sammer, G., Schienle, A., Stark, R., & Vaitl, D. (2003). MARINA: An easy to use tool for the creation of MAsks for Region of INterest Analyses [abstract]. Presented at the 9th International Conference on Functional Mapping of the Human Brain, June 19-22, 2003, New York, NY. Available on CD-Rom in *NeuroImage, 19*, No. 2.

Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology I* (pp. 135-151). Harmondsworth: Penguin.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content*. Cambridge: Harvard University Press.

Wason, P. C., & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *The Quarterly Journal of Experimental Psychology, 23,* 63-71.

Weinkauf, W. (2001). *Die Philosophie der Stoa: Ausgewählte Texte* [The philosophy of the Stoa: Elected texts] (W. Weinkauf, Trans.). Ditzingen: Reclam.

Weiß, R. H. (2006). *Grundintelligenztest Skala 2 Revision (CFT 20-R) [culture-fair-test]*. Göttingen: Hogrefe.

Wiedenmann, P., & Knauff, M. (2008). Conditional reasoning with a moral content. *International Journal of Psychology, 43 (3/4)*, 773. (International Conference of Psychology, Berlin).

Wilkins, M. C. (1928). The effect of changed material on the ability to do formal syllogistic reasoning. *Archives of Psychology, 16*, 102.

Wolf, U. (1996). *Die Suche nach dem guten Leben – Platons Frühdialoge* [Searching for the good life – Platon´s early dialogues]. Reinbek: Rowohlt.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46,* 441-517.

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences, 104*, 8235-8240.

Young, L., & Koenigs, M. (2007). Investigating emotion in moral cognition: a review of evidence from functional neuroimaging and neuropsychology. *British Medical Bulletin, 84,* 69-79.

Ziegler, A. (1990). Deduktives Schließen mit mentalen Modellen [Deductive reasoning with mental models]. *Sprache & Kognition, 9*, 82-91.

Appendix

Contents of the Appendix

Appendix A: Deductive inference problems in original German language	167
Appendix B: Recognition items in original German language	171
Appendix C: Example questions of the Questionnaires for Pre-Studies I + II + III	173
Appendix D: Results of Pre-Study I	174
Appendix E: Results of Experiment I	175
Appendix F: Results of Pre-Study II	176
Appendix G: Results of Experiment II	177
Appendix H: Results of Pre-Study III	179
Deductive inference problems	179
Recognition items	180
Appendix I: Material of the logic training in original German language	181
Pretest – instruction and problems	181
Introduction to propositional logic	182
Training phase – problems	185
Posttest – instruction and problems	187
Appendix J: Material of the pseudo training in original German language	188
Introduction to matrices tasks	188
Training phase – examples of matrices tasks	193
Appendix K: Results of Experiment III	194
Appendix L: Results of Experiment IV	197
Post hoc Wilcoxon signed rank tests	197
Correlation tables with 6 participants	198
Correlation scatterplots with regression line	200
Appendix M: Results of Experiment V (behavioral data)	202
Appendix N: Results of Experiment V (brain imaging data)	204
Analysis of single effects	204
Conjunction analysis	205
Difference comparisons between conditions	206

Appendix A

Deductive inference problems in original German language

Deductive inference problems of all Pre-Studies and Experiments		
validity / conditions	valid	invalid
moral	Wenn ein Mensch behindert zur Welt kommt, dann hat er trotzdem ein Recht auf Leben. Ein Mensch kommt behindert zur Welt. Der Mensch hat ein Recht auf Leben.	Wenn ein Mensch behindert zur Welt kommt, dann muss er getötet werden. Ein Mensch kommt behindert zur Welt. Der Mensch darf nicht getötet werden.
	Wenn ein Mensch ein Kriegsgefangener ist, dann darf er nicht versklavt werden. Ein Mensch ist ein Kriegsgefangener. Der Mensch darf nicht versklavt werden.	Wenn ein Mensch ein Kriegsgefangener ist, dann darf er versklavt werden. Ein Mensch ist ein Kriegsgefangener. Der Mensch darf nicht versklavt werden.
	Wenn ein Verbrecher verhört wird, dann muss er menschenwürdig behandelt werden. Ein Verbrecher wird verhört. Der Verbrecher muss menschenwürdig behandelt werden.	Wenn ein Verbrecher verhört wird, dann darf er gefoltert werden. Ein Verbrecher wird verhört. Der Verbrecher darf nicht gefoltert werden.
	Wenn ein Mensch einer Minderheit angehört, dann muss er durch das Gesetz geschützt werden. Ein Mensch gehört einer Minderheit an. Der Mensch muss durch das Gesetz geschützt werden.	Wenn ein Mensch einer Minderheit angehört, dann ist er ohne Schutz durch das Gesetz. Ein Mensch gehört einer Minderheit an. Der Mensch muss trotzdem durch das Gesetz geschützt werden.
unmoral	Wenn ein Mensch behindert zur Welt kommt, dann muss er getötet werden. Ein Mensch kommt behindert zur Welt. Der Mensch muss getötet werden.	Wenn ein Mensch behindert zur Welt kommt, dann hat er trotzdem ein Recht auf Leben. Ein Mensch kommt behindert zur Welt. Der Mensch hat kein Recht auf Leben.
	Wenn ein Mensch ein Kriegsgefangener ist, dann darf er versklavt werden. Ein Mensch ist ein Kriegsgefangener. Der Mensch darf versklavt werden.	Wenn ein Mensch ein Kriegsgefangener ist, dann darf er nicht versklavt werden. Ein Mensch ist ein Kriegsgefangener. Der Mensch darf versklavt werden.
	Wenn ein Verbrecher verhört wird, dann darf er gefoltert werden. Ein Verbrecher wird verhört. Der Verbrecher darf gefoltert werden.	Wenn ein Verbrecher verhört wird, dann muss er menschenwürdig behandelt werden. Ein Verbrecher wird verhört. Der Verbrecher wird nicht menschenwürdig behandelt.
	Wenn ein Mensch einer Minderheit angehört, dann ist er ohne Schutz durch das Gesetz. Ein Mensch gehört einer Minderheit an. Der Mensch ist ohne Schutz durch das Gesetz.	Wenn ein Mensch einer Minderheit angehört, dann muss er durch das Gesetz geschützt werden. Ein Mensch gehört einer Minderheit an. Der Mensch ist ohne Schutz durch das Gesetz.

Continuation of Appendix A

Deductive inference problems of Pre-Study I and Experiments I + II + V		
validity / conditions	valid	invalid
abstract	Wenn A, dann B. A. B.	Wenn A, dann B. A. Nicht B.
	Wenn C, dann D. C. D.	Wenn C, dann D. C. Nicht D.
	Wenn X, dann Y. X. Y.	Wenn X, dann Y. X. Nicht Y.
	Wenn W, dann Z. W. Z.	Wenn W, dann Z. W. Nicht Z.
neutral	Wenn ein Mensch fremde Länder mag, dann verreißt er gerne. Ein Mensch mag fremde Länder. Der Mensch verreißt gerne.	Wenn ein Mensch fremde Länder mag, dann verreißt er gerne. Ein Mensch mag fremde Länder. Der Mensch verreißt nicht gerne.
	Wenn ein Mensch die Natur mag, dann mag er auch Blumen. Ein Mensch mag die Natur. Der Mensch mag auch Blumen.	Wenn ein Mensch die Natur mag, dann mag er auch Blumen. Ein Mensch mag die Natur. Der Mensch mag keine Blumen.
	Wenn ein Mensch starken Haarwuchs hat, dann muss er häufiger zum Friseur gehen. Ein Mensch hat starken Haarwuchs. Der Mensch muss häufiger zum Friseur gehen.	Wenn ein Mensch starken Haarwuchs hat, dann muss er häufiger zum Friseur gehen. Ein Mensch hat starken Haarwuchs. Der Mensch muss nicht häufiger zum Friseur gehen.
	Wenn ein Mensch älter wird, dann bekommt er Falten. Ein Mensch wird älter. Der Mensch bekommt Falten.	Wenn ein Mensch älter wird, dann bekommt er Falten. Ein Mensch wird älter. Der Mensch bekommt keine Falten.

Continuation of Appendix A

Deductive inference problems of Pre-Study II and Experiment II + V		
validity / conditions	valid	invalid
positive-emotional	Wenn ein Mensch glücklich verliebt ist, dann findet er sein Leben sinnvoll. Ein Mensch ist glücklich verliebt. Der Mensch findet sein Leben sinnvoll.	Wenn ein Mensch glücklich verliebt ist, dann findet er sein Leben nicht sinnvoll. Ein Mensch ist glücklich verliebt. Der Mensch findet sein Leben sinnvoll.
	Wenn ein Mensch Freunde hat, dann ist er in Gesellschaft. Ein Mensch hat Freunde. Der Mensch ist in Gesellschaft.	Wenn ein Mensch Freunde hat, dann ist er nicht in Gesellschaft. Ein Mensch hat Freunde. Der Mensch ist in Gesellschaft.
	Wenn ein Mensch an der Börse einen großen Gewinn macht, dann ist er glücklich. Ein Mensch macht an der Börse einen großen Gewinn. Der Mensch ist glücklich.	Wenn ein Mensch an der Börse einen großen Gewinn macht, dann ist er nicht glücklich. Ein Mensch macht an der Börse einen großen Gewinn. Der Mensch ist glücklich.
	Wenn ein Mensch ein Gewinner ist, dann ist er auf der Sonnenseite des Lebens. Ein Mensch ist ein Gewinner. Der Mensch ist auf der Sonnenseite des Lebens.	Wenn ein Mensch ein Gewinner ist, dann ist er nicht auf der Sonnenseite des Lebens. Ein Mensch ist ein Gewinner. Der Mensch ist auf der Sonnenseite des Lebens.
negative-emotional	Wenn ein Mensch unglücklich verliebt ist, dann findet er sein Leben sinnlos. Ein Mensch ist unglücklich verliebt. Der Mensch findet sein Leben sinnlos.	Wenn ein Mensch unglücklich verliebt ist, dann findet er sein Leben nicht sinnlos. Ein Mensch ist unglücklich verliebt. Der Mensch findet sein Leben sinnlos.
	Wenn ein Mensch ohne Freunde ist, dann ist er einsam. Ein Mensch ist ohne Freunde. Der Mensch ist einsam.	Wenn ein Mensch ohne Freunde ist, dann ist er nicht einsam. Ein Mensch ist ohne Freunde. Der Mensch ist einsam.
	Wenn ein Mensch an der Börse einen großen Verlust macht, dann ist er unglücklich. Ein Mensch macht an der Börse einen großen Verlust. Der Mensch ist unglücklich.	Wenn ein Mensch an der Börse einen großen Verlust macht, dann ist er nicht unglücklich. Ein Mensch macht an der Börse einen großen Verlust. Der Mensch ist unglücklich.
	Wenn ein Mensch ein Verlierer ist, dann ist er auf der Schattenseite des Lebens. Ein Mensch ist ein Verlierer. Der Mensch ist auf der Schattenseite des Lebens.	Wenn ein Mensch ein Verlierer ist, dann ist er nicht auf der Schattenseite des Lebens. Ein Mensch ist ein Verlierer. Der Mensch ist auf der Schattenseite des Lebens.

Continuation of Appendix A

Deductive inference problems of Pre-Study III and Experiments III + IV		
validity / conditions	valid	invalid
neutral1	Wenn ein Mensch abends ins Kino geht, dann sieht er sich einen Film an. Ein Mensch geht abends ins Kino. Der Mensch sieht sich einen Film an.	Wenn ein Mensch abends ins Kino geht, dann sieht er sich einen Film an. Ein Mensch geht abends ins Kino. Der Mensch sieht sich keinen Film an.
	Wenn ein Mensch im Wald spazieren geht, dann zieht er eine Jacke an. Ein Mensch geht im Wald spazieren. Der Mensch zieht eine Jacke an.	Wenn ein Mensch im Wald spazieren geht, dann zieht er eine Jacke an. Ein Mensch geht im Wald spazieren. Der Mensch zieht keine Jacke an.
	Wenn ein Mensch den Tierpark besucht, dann beobachtet er exotische Tiere. Ein Mensch besucht den Tierpark. Der Mensch beobachtet exotische Tiere.	Wenn ein Mensch den Tierpark besucht, dann beobachtet er exotische Tiere. Ein Mensch besucht den Tierpark. Der Mensch beobachtet keine exotischen Tiere.
	Wenn ein Mensch sportlich ist, dann trainiert er im Fitnessstudio. Ein Mensch ist sportlich. Der Mensch trainiert im Fitnessstudio.	Wenn ein Mensch sportlich ist, dann trainiert er im Fitnessstudio. Ein Mensch ist sportlich. Der Mensch trainiert nicht im Fitnessstudio.
neutral2	Wenn ein Mensch besonders starken Haarwuchs hat, dann muss er häufig zum Friseur gehen. Ein Mensch hat besonders starken Haarwuchs. Der Mensch muss häufig zum Friseur gehen.	Wenn ein Mensch besonders starken Haarwuchs hat, dann muss er häufig zum Friseur gehen. Ein Mensch hat besonders starken Haarwuchs. Der Mensch muss nicht häufig zum Friseur gehen.
	Wenn ein Mensch über die Straße geht, dann achtet er auf den Verkehr. Ein Mensch geht über die Straße. Der Mensch achtet auf den Verkehr.	Wenn ein Mensch über die Straße geht, dann achtet er auf den Verkehr. Ein Mensch geht über die Straße. Der Mensch achtet nicht auf den Verkehr.
	Wenn ein Mensch einen Bernhardiner hat, dann kauft er Hundefutter. Ein Mensch hat einen Bernhardiner. Der Mensch kauft Hundefutter.	Wenn ein Mensch einen Bernhardiner hat, dann kauft er Hundefutter. Ein Mensch hat einen Bernhardiner. Der Mensch kauft kein Hundefutter.
	Wenn ein Mensch einen Computer besitzt, dann benutzt er ihn zum Surfen im Internet. Ein Mensch besitzt einen Computer. Der Mensch benutzt ihn zum Surfen im Internet.	Wenn ein Mensch einen Computer besitzt, dann benutzt er ihn zum Surfen im Internet. Ein Mensch besitzt einen Computer. Der Mensch benutzt ihn nicht zum Surfen im Internet.

Appendix B

Recognition items in original German language

Recognition items of Pre-Study III and Experiments III + IV		
matching conditions	literal match	no literal match
moral	Wenn ein Mensch behindert zur Welt kommt, dann hat er trotzdem ein Recht auf Leben. Ein Mensch wird im Krankenhaus geboren. Dann hat er trotzdem ein Recht auf Leben.	Wenn ein Mensch behindert zur Welt kommt, dann muss er getötet werden. Ein Mensch wird im Krankenhaus geboren. Dann darf er nicht getötet werden.
	Wenn ein Mensch ein Kriegsgefangener ist, dann darf er nicht versklavt werden. Ein Mensch befindet sich im Krieg. Dann darf er nicht versklavt werden.	Wenn ein Mensch ein Kriegsgefangener ist, dann darf er versklavt werden. Ein Mensch befindet sich im Krieg. Dann darf er nicht versklavt werden.
	Wenn ein Verbrecher verhört wird, dann muss er menschenwürdig behandelt werden. Ein Verbrecher ist im Gefängnis. Dann muss er menschenwürdig behandelt werden.	Wenn ein Verbrecher verhört wird, dann darf er gefoltert werden. Ein Verbrecher ist im Gefängnis. Dann darf er nicht gefoltert werden.
	Wenn ein Mensch einer Minderheit angehört, dann muss er durch das Gesetz geschützt werden. Ein Mensch ist Leistungssportler. Dann muss er durch das Gesetz geschützt werden.	Wenn ein Mensch einer Minderheit angehört, dann ist er schutzlos vor dem Gesetz. Ein Mensch ist Leistungssportler. Dann ist er nicht schutzlos vor dem Gesetz.
unmoral	Wenn ein Mensch behindert zur Welt kommt, dann muss er getötet werden. Ein Mensch wird im Krankenhaus geboren. Dann muss er getötet werden.	Wenn ein Mensch behindert zur Welt kommt, dann hat er trotzdem ein Recht auf Leben. Ein Mensch wird im Krankenhaus geboren. Dann hat er kein Recht auf Leben.
	Wenn ein Mensch ein Kriegsgefangener ist, dann darf er versklavt werden. Ein Mensch befindet sich im Krieg. Dann darf er versklavt werden.	Wenn ein Mensch ein Kriegsgefangener ist, dann darf er nicht versklavt werden. Ein Mensch befindet sich im Krieg. Dann darf er versklavt werden.
	Wenn ein Verbrecher verhört wird, dann darf er gefoltert werden. Ein Verbrecher ist im Gefängnis. Dann darf er gefoltert werden.	Wenn ein Verbrecher verhört wird, dann muss er menschenwürdig behandelt werden. Ein Verbrecher ist im Gefängnis. Dann muss er nicht menschenwürdig behandelt werden.
	Wenn ein Mensch einer Minderheit angehört, dann ist er ohne Schutz durch das Gesetz. Ein Mensch ist Leistungssportler. Dann ist er ohne Schutz durch das Gesetz.	Wenn ein Mensch einer Minderheit angehört, dann muss er durch das Gesetz geschützt werden. Ein Mensch ist Leistungssportler. Dann muss er nicht durch das Gesetz geschützt werden.

Continuation of Appendix B

Recognition items of Pre-Study III and Experiments III + IV		
matching conditions	literal match	no literal match
neutral1	Wenn ein Mensch abends ins Kino geht, dann sieht er sich einen Film an. Ein Mensch geht abends ins Theater. Dann sieht er sich einen Film an.	Wenn ein Mensch abends ins Kino geht, dann sieht er sich einen Film an. Ein Mensch geht abends ins Theater. Dann sieht er sich keinen Film an.
	Wenn ein Mensch im Wald spazieren geht, dann zieht er eine Jacke an. Der Mensch geht nach draußen. Dann zieht er eine Jacke an.	Wenn ein Mensch im Wald spazieren geht, dann zieht er eine Jacke an. Der Mensch geht nach draußen. Dann zieht er keine Jacke an.
	Wenn ein Mensch den Tierpark besucht, dann beobachtet er exotische Tiere. Der Mensch besucht ein Museum. Dann beobachtet er exotische Tiere.	Wenn ein Mensch den Tierpark besucht, dann beobachtet er exotische Tiere. Der Mensch besucht ein Museum. Dann beobachtet er keine exotischen Tiere.
	Wenn ein Mensch sportlich ist, dann trainiert er im Fitnessstudio. Ein Mensch ist musikalisch. Dann trainiert er im Fitnessstudio.	Wenn ein Mensch sportlich ist, dann trainiert er im Fitnessstudio. Ein Mensch ist musikalisch. Dann trainiert er nicht im Fitnessstudio.
neutral2	Wenn ein Mensch besonders starken Haarwuchs hat, dann muss er häufig zum Friseur gehen. Ein Mensch hat einen Bart. Dann muss er häufig zum Friseur gehen.	Wenn ein Mensch besonders starken Haarwuchs hat, dann muss er häufig zum Friseur gehen. Ein Mensch hat einen Bart. Dann muss er nicht häufig zum Friseur gehen.
	Wenn ein Mensch über die Straße geht, dann achtet er auf den Verkehr. Ein Mensch geht durch die Stadt. Dann achtet er auf den Verkehr.	Wenn ein Mensch über die Straße geht, dann achtet er auf den Verkehr. Ein Mensch geht durch die Stadt. Dann achtet er nicht auf den Verkehr.
	Wenn ein Mensch einen Bernhardiner hat, dann kauft er Hundefutter. Der Mensch hat einen Hamster. Dann kauft er Hundefutter.	Wenn ein Mensch einen Berhardiner hat, dann kauft er Hundefutter. Der Mensch hat einen Hamster. Dann kauft er kein Hundefutter.
	Wenn ein Mensch einen Computer besitzt, dann benutzt er ihn zum Surfen im Internet. Ein Mensch besitzt einen Locher. Dann benutzt er ihn zum Surfen im Internet.	Wenn ein Mensch einen Computer besitzt, dann benutzt er ihn zum Surfen im Internet. Ein Mensch besitzt einen Locher. Dann benutzt er ihn nicht zum Surfen im Internet.

Appendix C

Example questions of the Questionnaires for Pre-Studies I + II + III

Example questions of Pre-Study I

Wenn X, dann Y.
X.

Nicht Y.

Wie viel hat diese Aufgabenstellung mit logischem Denken zu tun?
☐ nichts – ☐ etwas – ☐ viel – ☐ sehr viel

Wie viel hat diese Aufgabenstellung mit Moral zu tun?
☐ nichts – ☐ etwas – ☐ viel – ☐ sehr viel

Wie viel hat diese Aufgabenstellung mit Emotionen zu tun?
☐ nichts – ☐ etwas – ☐ viel – ☐ sehr viel

Example questions of Pre-Studies II + III

Wenn ein Mensch behindert zur Welt kommt, dann hat er trotzdem ein Recht auf Leben.
Ein Mensch kommt behindert zur Welt.

Der Mensch hat ein Recht auf Leben.

Wie viel hat diese Aufgabenstellung mit logischem Denken zu tun?
☐ nichts – ☐ etwas – ☐ viel – ☐ sehr viel

Wie viel hat diese Aufgabenstellung mit Moral zu tun?
☐ nichts – ☐ etwas – ☐ viel – ☐ sehr viel

Wie viel hat diese Aufgabenstellung mit positiven Emotionen zu tun?
☐ nichts – ☐ etwas – ☐ viel – ☐ sehr viel

Wie viel hat diese Aufgabenstellung mit negativen Emotionen zu tun?
☐ nichts – ☐ etwas – ☐ viel – ☐ sehr viel

Appendix D

Results of Pre-Study I

problem-evaluation I Wilcoxon signed rank test	logic dimension		morality dimension		emotionality dimension	
	z-values	p-values	z-values	p-values	z-values	p-values
moral-v vs. moral-iv	z = -3.418	p = .001	z = -1.904	p = .057	z = -1.667	p = .095
moral-v vs. unmoral-v	z = -3.052	p = .002	z = -.558	p = .577	z = -1.728	p = .084
moral-v vs. unmoral-iv	z = -3.149	p = .002	z = -.816	p = .415	z = -.347	p = .728
moral-v vs. neutral-v	z = -.039	p = .986	z = -3.735	p < .001	z = -3.466	p = .001
moral-v vs. neutral-iv	z = -3.366	p = .001	z = -3.735	p < .001	z = -3.521	p < .001
moral-v vs. abstract-v	z = -.797	p = .425	z = -3.736	p < .001	z = -3.527	p < .001
moral-v vs. abstract-iv	z = -1.802	p = .072	z = -3.736	p < .001	z = -3.531	p < .001
moral-iv vs. unmoral-v	z = -2.260	p = .024	z = -2.324	p = .020	z = .000	p = 1.000
moral-iv vs. unmoral-iv	z = -.093	p = .926	z = -2.737	p = .006	z = -1.574	p = .116
moral-iv vs. neutral-v	z = -3.129	p = .002	z = -3.839	p < .001	z = -3.535	p < .001
moral-iv vs. neutral-iv	z = -.295	p = .768	z = -3.839	p < .001	z = -3.523	p < .001
moral-iv vs. abstract-v	z = -3.105	p = .002	z = -3.843	p < .001	z = -3.530	p < .001
moral-iv vs. abstract-iv	z = -2.677	p = .007	z = -3.843	p < .001	z = -3.529	p < .001
unmoral-v vs. unmoral-iv	z = -2.138	p = .033	z = -.312	p = .755	z = -1.813	p = .070
unmoral-v vs. neutral-v	z = -3.185	p = .001	z = -3.638	p < .001	z = -3.670	p < .001
unmoral-v vs. neutral-iv	z = -2.399	p = .016	z = -3.638	p < .001	z = -3.729	p < .001
unmoral-v vs. abstract-v	z = -3.051	p = .002	z = -3.642	p < .001	z = -3.732	p < .001
unmoral-v vs. abstract-iv	z = -.378	p = .705	z = -3.642	p < .001	z = -3.736	p < .001
unmoral-iv vs. neutral-v	z = -2.938	p = .003	z = -3.734	p < .001	z = -3.445	p = .001
unmoral-iv vs. neutral-iv	z = -.090	p = .928	z = -3.734	p < .001	z = -3.559	p < .001
unmoral-iv vs. abstract-v	z = -3.107	p = .002	z = -3.736	p < .001	z = -3.578	p < .001
unmoral-iv vs. abstract-iv	z = -2.626	p = .009	z = -3.736	p < .001	z = -3.580	p < .001
neutral-v vs. neutral-iv	z = -3.302	p = .001	z = .000	p = 1.000	z = -.589	p = .556
neutral-v vs. abstract-v	z = -1.886	p = .059	z = -1.000	p = .317	z = -2.947	p = .003
neutral-v vs. abstract-iv	z = -2.049	p = .040	z = -1.000	p = .317	z = -2.956	p = .003
neutral-iv vs. abstract-v	z = -3.224	p = .001	z = -1.000	p = .317	z = -2.952	p = .003
neutral-iv vs. abstract-iv	z = -2.513	p = .012	z = -1.000	p = .317	z = -2.947	p = .003
abstract-v vs. abstract-iv	z = -2.484	p = .013	z = .000	p = 1.000	z = -1.000	p = .317

Appendix E

Results of Experiment I

post hoc paired-sample t-tests	due to significant IE of content × validity	
	error rates	
calculated comparisons	t-values (df)	p-values
moral-v vs. moral-iv	t(20) = -3.710	p = .001
moral-v vs. unmoral-v	t(20) = -2.646	p = .016
moral-v vs. neutral-v	--- (n.s.)	--- (n.s.)
moral-v vs. abstract-v	t(20) = -1.369	p = .186
moral-iv vs. unmoral-iv	t(20) = 3.202	p = .004
moral-iv vs. neutral-iv	t(20) = 3.532	p = .002
moral-iv vs. abstract-iv	t(20) = 3.325	p = .003
unmoral-v vs.unmoral-iv	t(20) = 1.921	p = .069
unmoral-v vs. neutral-v	t(20) = 2.646	p = .016
unmoral-v vs. abstract-v	t(20) = 2.329	p = .030
unmoral-iv vs. neutral-iv	t(20) = .000	p = 1.000
unmoral-iv vs. abstract-iv	t(20) = -.271	p = .789
neutral-v vs. neutral-iv	t(20) = -1.369	p = .186
neutral-v vs. abstract-v	t(20) = -1.369	p = .186
neutral-iv vs. abstract-iv	t(20) = -.295	p = .771
abstract-v vs. abstract-iv	t(20) = -1.000	p = .329

Appendix F

Results of Pre-Study II

problem-evaluation II Wilcoxon signed rank test	logic dimension		morality dimension		positive-emotionality dimension		negative-emotionality dimension	
	z-values	p-values	z-values	p-values	z-values	p-values	z-values	p-values
moral-v vs. moral-iv	z = -2.779	p = .005	z = -.141	p = .888	z = -3.785	p < .001	z = -2.220	p = .026
moral-v vs. unmoral-v	z = -2.772	p = .006	z = -2.573	p = .010	z = -3.935	p < .001	z = -3.927	p < .001
moral-v vs. unmoral-iv	z = -3.165	p = .002	z = -1.548	p = .122	z = -3.851	p < .001	z = -3.795	p < .001
moral-v vs. positive-emotional-v	z = -.172	p = .863	z = -3.925	p < .001	z = -3.391	p = .001	z = -3.480	p = .001
moral-v vs. positive-emotional-iv	z = -2.760	p = .006	z = -3.930	p < .001	z = -2.255	p = .024	z = -3.400	p = .001
moral-v vs. negative-emotional-v	z = -1.129	p = .259	z = -3.930	p < .001	z = -3.796	p < .001	z = -3.939	p < .001
moral-v vs. negative-emotional-iv	z = -2.894	p = .004	z = -3.923	p < .001	z = -2.808	p = .005	z = -3.403	p = .001
moral-iv vs. unmoral-v	z = -.685	p = .493	z = -2.639	p = .008	z = -3.069	p = .002	z = -3.735	p < .001
moral-iv vs. unmoral-iv	z = -1.603	p = .109	z = -.985	p = .325	z = -1.967	p = .049	z = -2.464	p = .014
moral-iv vs. positive-emotional-v	z = -2.511	p = .012	z = -3.257	p = .001	z = -3.889	p < .001	z = -3.439	p = .001
moral-iv vs. positive-emotional-iv	z = -.190	p = .849	z = -3.219	p = .001	z = -.524	p = .600	z = -.939	p = .348
moral-iv vs. negative-emotional-v	z = -2.014	p = .044	z = -3.299	p = .001	z = -2.926	p = .003	z = -3.539	p < .001
moral-iv vs. negative-emotional-iv	z = -.679	p = .497	z = -3.017	p = .003	z = -.026	p = .979	z = -1.516	p = .129
unmoral-v vs. unmoral-iv	z = -1.761	p = .078	z = -1.177	p = .239	z = -2.993	p = .003	z = -1.859	p = .063
unmoral-v vs. positive-emotional-v	z = -2.794	p = .005	z = -2.260	p = .024	z = -3.936	p < .001	z = -3.935	p < .001
unmoral-v vs. positive-emotional-iv	z = -.901	p = .368	z = -2.511	p = .012	z = -3.738	p < .001	z = -2.799	p = .005
unmoral-v vs. negative-emotional-v	z = -2.435	p = .015	z = -2.579	p = .010	z = -1.511	p = .131	z = -.022	p = .982
unmoral-v vs. negative-emotional-iv	z = -1.139	p = .255	z = -2.190	p = .029	z = -3.511	p < .001	z = -3.230	p = .001
unmoral-iv vs. positive-emotional-v	z = -3.031	p = .002	z = -3.121	p = .002	z = -3.931	p < .001	z = -3.830	p < .001
unmoral-iv vs. positive-emotional-iv	z = -2.379	p = .017	z = -3.566	p < .001	z = -3.560	p < .001	z = -1.946	p = .052
unmoral-iv vs. negative-emotional-v	z = -2.793	p = .005	z = -3.408	p = .001	z = -2.214	p = .027	z = -2.073	p = .038
unmoral-iv vs. negative-emotional-iv	z = -2.101	p = .036	z = -3.195	p = .001	z = -1.975	p = .048	z = -1.830	p = .067
positive-emotional-v vs. positive-emotional-iv	z = -2.537	p = .011	z = -.095	p = .924	z = -3.830	p < .001	z = -3.835	p < .001
positive-emotional-v vs. negative-emotional-v	z = -2.169	p = .030	z = -.090	p = .929	z = -3.934	p < .001	z = -3.929	p < .001
positive-emotional-v vs. negative-emotional-iv	z = -2.943	p = .003	z = -.703	p = .482	z = -3.855	p < .001	z = -3.688	p < .001
positive-emotional-iv vs. negative-emotional-v	z = -2.348	p = .019	z = -.606	p = .545	z = -3.735	p < .001	z = -3.487	p < .001
positive-emotional-iv vs. negative-emotional-iv	z = -.956	p = .339	z = -2.030	p = .042	z = -.357	p = .721	z = -.087	p = .930
negative-emotional-v vs. negative-emotional-iv	z = -2.659	p = .008	z = -.699	p = .485	z = -3.424	p = .001	z = -3.436	p = .001

Appendix G

Results of Experiment II

post hoc paired-sample t-tests	due to significant IE of content × validity			
	decision times		error rates	
calculated comparisons	t-values (df)	p-values	t-values (df)	p-values
moral-v vs. moral-iv	t(32) = -5.486	p < .001	t(32) = -4.572	p < .001
moral-v vs. unmoral-v	t(32) = -2.374	p =.024	t(32) = -2.478	p = .019
moral-v vs. neutral-v	t(32) = 3.024	p = .005	t(32) =.627	p = .535
moral-v vs. abstract-v	t(32) = 5.778	p < .001	t(32) = -.828	p = .414
moral-v vs. positive-emotional-v	t(32) = 2.264	p = .030	t(32) =.442	p = .662
moral-v vs. negative-emotional-v	t(32) = .423	p = .675	t(32) =.000	p = 1.000
moral-iv vs. unmoral-iv	t(32) = 2.749	p = .010	t(32) = 4.195	p < .001
moral-iv vs. neutral-iv	t(32) = 4.722	p < .001	t(32) = 4.804	p < .001
moral-iv vs. abstract-iv	t(32) = 6.174	p < .001	t(32) = 4.707	p < .001
moral-iv vs. positive-emotional-iv	t(32) = 4.434	p < .001	t(32) = 5.078	p < .001
moral-iv vs. negative-emotional-iv	t(32) = 4.096	p < .001	t(32) = 5.108	p < .001
unmoral-v vs. unmoral-iv	t(32) = -2.901	p = .007	t(32) = 2.232	p = .033
unmoral-v vs.neutral-v	t(32) = 4.308	p < .001	t(32) = 2.901	p = .007
unmoral-v vs. abstract-v	t(32) = 5.846	p < .001	t(32) = 2.213	p = .034
unmoral-v vs. positive-emotional-v	t(32) = 3.849	p = .001	t(32) = 2.640	p = .013
unmoral-v vs. negative-emotional-v	t(32) = 2.248	p = .032	t(32) = 2.378	p = .024
unmoral-iv vs.neutral-iv	t(32) = 3.539	p = .001	t(32) = 1.000	p = .325
unmoral-iv vs. abstract-iv	t(32) = 7.359	p < .001	t(32) =.620	p = .540
unmoral-iv vs. positive-emotional-iv	t(32) = 2.021	p = .052	t(32) =.770	p = .447
unmoral-iv vs. negative-emotional-iv	t(32) = 2.155	p = .039	t(32) =.226	p = .823
neutral-v vs.neutral-iv	t(32) = -2.784	p = .009	t(32) = -.702	p = .488

Continuation of Appendix G

post hoc paired-sample t-tests	due to significant IE of content × validity			
	decision times		error rates	
calculated comparisons	t-values (df)	p-values	t-values (df)	p-values
neutral-v vs. abstract-v	t(32) = 3.505	p = .001	t(32) = -1.305	p = .201
neutral-v vs. positive-emotional-v	t(32) = -.222	p = .826	t(32) = -.373	p = .712
neutral-v vs. negative-emotional-v	t(32) = -2.045	p = .049	t(32) = -.571	p = .572
neutral-iv vs. abstract-iv	t(32) = 3.682	p = .001	t(32) = -.273	p = .786
neutral-iv vs. positive-emotional-iv	t(32) = -.519	p = .608	t(32) = -.373	p = .712
neutral-iv vs. negative-emotional-iv	t(32) = -1.996	p = .055	t(32) = -1.000	p = .325
abstract-v vs. abstract-iv	t(32) = -1.343	p = .189	t(32) =.466	p = .645
abstract-v vs. positive-emotional-v	t(32) = -2.737	p = .010	t(32) =.941	p = .354
abstract-v vs. negative-emotional-v	t(32) = -3.765	p = .001	t(32) = 1.139	p = .263
abstract-iv vs. positive-emotional-iv	t(32) = -2.378	p = .024	t(32) =.000	p = 1.000
abstract-iv vs.negative-emotional-iv	t(32) = -4.535	p < .001	t(32) = -.627	p = .535
positive-emotional-v vs. positive-emotional-iv	t(32) = -1.806	p = .080	t(32) = -.571	p = .572
positive-emotional-v vs. negative-emotional-v	t(32) = -1.349	p = .187	t(32) = -.239	p = .813
positive-emotional-iv vs. negative-emotional-iv	t(32) = -.636	p = .529	t(32) = -1.000	p = .325
negative-emotional-v vs. negative-emotional-iv	t(32) = -2.733	p = .010	t(32) = -.649	p = .521

Appendix H

Results of Pre-Study III – Deductive reasoning problems

problem-evaluation III Wilcoxon signed ranks test	logic dimension		morality dimension		positive-emotionality dimension		negative-emotionality dimension	
	z-values	p-values	z-values	p-values	z-values	p-values	z-values	p-values
moral-v vs. moral-iv	z = -4.354	p < .001	z = -.524	p = .600	z = -3.767	p < .001	z = -3.686	p < .001
moral-v vs. unmoral-v	z = -3.220	p = .001	z = -.529	p = .597	z = -4.034	p < .001	z = -4.441	p < .001
moral-v vs. unmoral-iv	z = -4.535	p < .001	z = -1.680	p = .093	z = -3.977	p < .001	z = -4.116	p < .001
moral-v vs. neutral1-v	z = -1.809	p = .070	z = -4.724	p < .001	z = -3.971	p < .001	z = -4.043	p < .001
moral-v vs. neutral1-iv	z = -4.390	p < .001	z = -4.735	p < .001	z = -4.116	p < .001	z = -3.251	p = .001
moral-v vs. neutral2-v	z = -1.402	p = .161	z = -4.732	p < .001	z = -3.440	p = .001	z = -4.031	p < .001
moral-v vs. neutral2-iv	z = -4.449	p < .001	z = -4.730	p < .001	z = -4.118	p < .001	z = -3.775	p < .001
moral-iv vs. unmoral-v	z = -3.201	p = .001	z = -.238	p = .812	z = -3.003	p = .003	z = -4.136	p < .001
moral-iv vs. unmoral-iv	z = -1.881	p = .060	z = -1.238	p = .216	z = -1.425	p = .154	z = -1.763	p = .078
moral-iv vs. neutral1-v	z = -4.354	p < .001	z = -4.728	p < .001	z = -.219	p = .826	z = -4.336	p < .001
moral-iv vs. neutral1-iv	z = -.681	p = .496	z = -4.737	p < .001	z = -2.438	p = .015	z = -4.313	p < .001
moral-iv vs. neutral2-v	z = -4.427	p < .001	z = -4.661	p < .001	z = -.595	p = .552	z = -4.342	p < .001
moral-iv vs. neutral2-iv	z = -1.451	p = .147	z = -4.729	p < .001	z = -1.263	p = .207	z = -4.273	p < .001
unmoral-v vs. unmoral-iv	z = -4.115	p < .001	z = -1.407	p = .160	z = -1.523	p = .128	z = -3.059	p = .002
unmoral-v vs. neutral1-v	z = -2.497	p = .013	z = -4.723	p < .001	z = -2.952	p = .003	z = -4.548	p < .001
unmoral-v vs. neutral1-iv	z = -3.912	p < .001	z = -4.724	p < .001	z = -1.210	p = .226	z = -4.549	p < .001
unmoral-v vs. neutral2-v	z = -2.655	p = .008	z = -4.653	p < .001	z = -2.756	p = .006	z = -4.549	p < .001
unmoral-v vs. neutral2-iv	z = -3.799	p < .001	z = -4.722	p < .001	z = -1.898	p = .058	z = -4.551	p < .001
unmoral-iv vs. neutral1-v	z = -4.525	p < .001	z = -4.717	p < .001	z = -1.992	p = .046	z = -4.551	p < .001
unmoral-iv vs. neutral1-iv	z = -1.118	p = .264	z = -4.722	p < .001	z = -.862	p = .389	z = -4.465	p < .001
unmoral-iv vs. neutral2-v	z = -4.531	p < .001	z = -4.722	p < .001	z = -2.136	p = .033	z = -4.548	p < .001
unmoral-iv vs. neutral2-iv	z = -.318	p = .751	z = -4.718	p < .001	z = -.839	p = .402	z = -4.553	p < .001
neutral1-v vs. neutral1-iv	z = -4.382	p < .001	z = -.284	p = .776	z = -3.623	p < .001	z = -2.503	p = .012
neutral1-v vs. neutral2-v	z = -.486	p = .627	z = -.254	p = .799	z = -1.535	p = .125	z = -1.508	p = .132
neutral1-v vs. neutral2-iv	z = -4.440	p < .001	z = -1.511	p = .131	z = -1.845	p = .065	z = -2.292	p = .022
neutral1-iv vs. neutral2-v	z = -4.513	p < .001	z = -.618	p = .537	z = -3.446	p = .001	z = -3.093	p = .002
neutral1-iv vs. neutral2-iv	z = -.629	p = .529	z = -1.942	p = .052	z = -2.167	p = .030	z = -1.642	p = .101
neutral2-v vs. neutral2-iv	z = -4.431	p < .001	z = -1.705	p = .088	z = -2.387	p = .017	z = -3.020	p = .003

Continuation of Appendix H

Results of Pre-Study III – Recognition items

recognition items evaluation III Wilcoxon signed ranks test (m = match, nm = no match)	logic dimension		morality dimension		positive-emotionality dimension		negative-emotionality dimension	
	z-values	p-values	z-values	p-values	z-values	p-values	z-values	p-values
moral-m vs. moral-nm	z = -2.318	p = .020	z = -.175	p = .861	z = -2.671	p = .008	z = -3.676	p < .001
moral-m vs. unmoral-m	z = -2.094	p = .036	z = -.632	p = .527	z = -3.423	p = .001	z = -4.445	p < .001
moral-m vs. unmoral-nm	z = -2.886	p = .004	z = -.713	p = .476	z = -3.333	p = .001	z = -3.800	p < .001
moral-m vs. neutral1-m	z = -.462	p = .644	z = -4.690	p < .001	z = -2.656	p = .008	z = -4.141	p < .001
moral-m vs. neutral1-nm	z = -.369	p = .712	z = -4.711	p < .001	z = -3.536	p < .001	z = -3.601	p < .001
moral-m vs. neutral2-m	z = -2.393	p = .017	z = -4.712	p < .001	z = -2.514	p = .012	z = -4.134	p < .001
moral-m vs. neutral2-nm	z = -2.622	p = .009	z = -4.711	p < .001	z = -3.284	p = .001	z = -4.153	p < .001
moral-nm vs. unmoral-m	z = -.214	p = .831	z = -.936	p = .350	z = -1.459	p = .144	z = -2.955	p = .003
moral-nm vs. unmoral-nm	z = -.976	p = .329	z = -.185	p = .853	z = -.120	p = .904	z = -.014	p = .989
moral-nm vs. neutral1-m	z = -1.468	p = .142	z = -4.581	p < .001	z = -.595	p = .552	z = -4.295	p < .001
moral-nm vs. neutral1-nm	z = -2.068	p = .039	z = -4.553	p < .001	z = -.176	p = .860	z = -4.220	p < .001
moral-nm vs. neutral2-m	z = -.594	p = .553	z = -4.600	p < .001	z = -1.193	p = .233	z = -4.206	p < .001
moral-nm vs. neutral2-nm	z = -3.295	p = .001	z = -4.598	p < .001	z = -1.165	p = .244	z = -4.325	p < .001
unmoral-m vs. unmoral-nm	z = -.893	p = .372	z = -1.477	p = .140	z = -1.582	p = .114	z = -3.660	p < .001
unmoral-m vs. neutral1-m	z = -1.394	p = .163	z = -4.634	p < .001	z = -2.024	p = .043	z = -4.468	p < .001
unmoral-m vs. neutral1-nm	z = -1.950	p = .051	z = -4.634	p < .001	z = -1.705	p = .088	z = -4.468	p < .001
unmoral-m vs. neutral2-m	z = -.934	p = .350	z = -4.632	p < .001	z = -2.435	p = .015	z = -4.470	p < .001
unmoral-m vs. neutral2-nm	z = -3.330	p = .001	z = -4.637	p < .001	z = -2.414	p = .016	z = -4.519	p < .001
unmoral-nm vs. neutral1-m	z = -2.044	p = .041	z = -4.712	p < .001	z = -.791	p = .429	z = -4.468	p < .001
unmoral-nm vs. neutral1-nm	z = -2.821	p = .005	z = -4.711	p < .001	z = -.284	p = .776	z = -4.469	p < .001
unmoral-nm vs. neutral2-m	z = -.231	p = .818	z = -4.712	p < .001	z = -1.429	p = .153	z = -4.465	p < .001
unmoral-nm vs. neutral2-nm	z = -3.431	p = .001	z = -4.715	p < .001	z = -1.331	p = .183	z = -4.382	p < .001
neutral1-m vs. neutral1-nm	z = -.836	p = .403	z = -1.725	p = .084	z = -.975	p = .330	z = -1.186	p = .236
neutral1-m vs. neutral2-m	z = -2.150	p = .032	z = -.273	p = .785	z = -.743	p = .458	z = -.587	p = .557
neutral1-m vs. neutral2-nm	z = -1.975	p = .048	z = .000	p = 1.000	z = -.539	p = .590	z = -1.000	p = .317
neutral1-nm vs. neutral2-m	z = -2.717	p = .007	z = -1.994	p = .046	z = -1.360	p = .174	z = -1.100	p = .271
neutral1-nm vs. neutral2-nm	z = -2.172	p = .030	z = -1.725	p = .084	z = -1.691	p = .091	z = -1.611	p = .107
neutral2-m vs. neutral2-nm	z = -3.340	p = .001	z = -.447	p = .655	z = -.354	p = .723	z = -.333	p = .739

Appendix I

Material of the logic training in original German language

Pretest

Pretest - instruction
Lieber Teilnehmer, im folgenden Test geht es um logisches Denken. Sie sollen einige Aufgaben dieser Art bearbeiten: Wenn Hänschen in die Schule geht, dann lernt er schreiben. Hänschen geht in die Schule. ――――――――――――――――― Hänschen lernt schreiben. Ist dieser Schluss logisch gültig? Ja ☐ Nein ☐ So sehen die Aufgaben aus. Sie sollen hier beurteilen, ob der dritte Satz unter dem Strich logisch aus den Sätzen über dem Strich folgt. Dies bedeutet, dass der logische Schluss, der hier gezogen wird, gültig ist. Wenn Sie also finden, dass der Satz unter dem Strich aus den ersten Sätzen folgt, dann kreuzen sie bitte „Ja“ an, wenn nicht, dann „Nein“. Vielen Dank für Ihre Mitarbeit.

Pretest – problems	
valid	invalid
Wenn M, dann N. M. N.	Wenn W hinten ist, dann ist A seitlich. W ist hinten. A ist nicht seitlich.
Wenn ein Mensch viel Alkohol trinkt, dann schadet er damit seiner Gesundheit. Ein Mensch trinkt viel Alkohol. Der Mensch schadet damit seiner Gesundheit.	Wenn Tina sich einen Hund kauft, dann muss sie mit ihm spazieren gehen. Tina kauft sich einen Hund. Sie muss nicht mit ihm spazieren gehen.
Wenn ein Mensch Hunger hat, dann holt er sich etwas zu essen aus dem Kühlschrank. Ein Mensch hat Hunger. Er holt sich etwas zu essen aus dem Kühlschrank.	Wenn es Außerirdische gibt, dann liegt New York im Westen von Texas. Es gibt Außerirdische. New York liegt nicht im Westen von Texas.
Wenn Johanna abnehmen will, dann isst sie abends ganz viele Chips. Johanna will abnehmen. Sie isst abends ganz viele Chips.	Wenn man einen Freund nicht jeden Tag sieht, dann verliert man ihn. Man sieht den Freund nicht jeden Tag. Man verliert ihn nicht.
Wenn D links ist, dann ist Q rechts. D ist links. Q ist rechts.	Wenn es nicht schneit, dann regnet es. Es schneit nicht. Es regnet nicht.
Wenn man sich nicht auf die Prüfung vorbereitet, dann fällt man durch. Man bereitet sich nicht auf seine Prüfung vor. Man fällt durch.	Wenn die Möhre krumm ist, dann ist sie eine Banane. Die Möhre ist krumm. Sie ist keine Banane.
Wenn ein Mensch rote Ohren hat, dann kommt er vom Mars. Ein Mensch hat rote Ohren. Er kommt vom Mars.	Wenn ein Mensch vor die Hunde geht, dann wird er von ihnen gefressen. Ein Mensch geht vor die Hunde. Der Mensch wird nicht von ihnen gefressen.
Wenn es rot ist, dann ist es grün. Es ist rot. Es ist grün.	Wenn ein Mensch unglücklich ist, dann denkt er an Selbstmord. Ein Mensch ist unglücklich. Der Mensch denkt nicht an Selbstmord.
Wenn der Frühling kommt, dann wird es wärmer draußen. Der Frühling kommt. Es wird wärmer draußen.	Wenn C1, dann P3. C1. Nicht P3.
Wenn ein Tier im Meer lebt, dann muss es ein Meeresfisch sein. Ein Tier lebt im Meer. Das Tier muss ein Meeresfisch sein.	Wenn Stephan sein Zimmer putzt, dann ist die Vier ein Primzahl. Stephan putzt sein Zimmer. Die Vier ist keine Primzahl.

Continuation of Appendix I

Introduction to propositional logic

Aussagenlogik

Können Sie logisch Denken? Sicherlich werden Sie diese Frage ohne zu zögern bejahen. Aber was ist logisches Denken bzw. Logik überhaupt?
Im Alltag hört man oft den Satz: „Das ist doch logisch!“ Meistens meinen wir damit, dass ein Gedankengang folgerichtig ist bzw. den Gesetzen der Logik entspricht und ihn somit jeder leicht nachvollziehen kann. Doch ist das, was wir denken, wirklich immer so logisch und für alle verständlich? Bestimmt haben Sie schon Situationen erlebt, in denen jemand ihren gedanklichen Schlüssen nicht so ohne Weiteres folgen konnte, oder Sie konnten die Gedankengänge eines anderen mal nicht verstehen. Um dies zu vermeiden, begannen Philosophen im 4. Jahrhundert v. Chr., eine allgemein gültige Logik zu entwickeln, um eindeutige Urteile über logisches bzw. unlogisches Denken und Argumentieren fällen zu können.
Von einem fachlichen Standpunkt aus ist Logik die Wissenschaft von den Gesetzen und Formen des Denkens. Begründet wurde die formale Logik von Aristoteles. Den Mittelpunkt der aristotelischen Logik bildet die Lehre vom Schließen und der Beweisführung. Zuerst betrachtete er die Logik als Teil der Rhetorik, mit dem der Redner auf das Auditorium einwirken kann. Später sah er die Logik als Wissen, das den Weg zur Erlangung von Wahrheit weist.
Nach Aristoteles beschäftigten sich noch unzählige andere berühmte Philosophen und Mathematiker mit der Logik, die so auch stets weiterentwickelt wurde. Heute wird Logik vor allem in der Philosophie und Mathematik gelehrt.

Hier beschäftigen wir uns nun mit der Aussagenlogik. Dabei zieht man aus zwei oder mehr gegebenen Aussagen logische Schlüsse, indem man die Beziehungen der Aussagen zueinander bewertet. Um das nachvollziehen zu können, zunächst ein Beispiel:

Wenn es regnet, dann ist die Straße nass.
Es regnet.

Also ist die Straße nass.

Die Sätze über dem Strich nennt man Prämissen, den Satz unter dem Strich Konklusion. Die erste Prämisse besteht aus zwei Aussagen, nämlich „Es regnet“ und „Die Straße ist nass“.

Man stellt sich nun die Frage, ob die Konklusion aus den Prämissen folgen kann. Wenn dies der Fall ist, wie beim oben genannten Beispiel, dann bezeichnet man das Ganze als logisch gültigen Schluss. Statt gültig kann man auch valide sagen. Wenn man von einem logisch validen Schluss spricht, so meint man damit nicht nur die Konklusion (den Satz unter dem Strich), sondern das ganze System der drei Sätze, also die Beziehung zwischen den Prämissen und der Konklusion. Die Konklusion selbst stellt also keinen Schluss dar. Doch was macht denn nun einen logisch validen Schluss aus?

Drei Merkmale von logisch validen Schlüssen sind besonders wichtig, um diese Frage zu beantworten. Diese werden Ihnen nun vorgestellt:

1.

Wenn die Prämissen wahr sind, dann ist auch die Konklusion wahr.

Man geht also davon aus, dass die Konklusion auch wahr ist, falls die Prämissen wahr sind. Es ist hier sehr wichtig zu beachten, dass man nicht behauptet, dass die Prämissen immer der
Continuation of introduction to propositional logic

Wahrheit entsprechen. Aber wenn die Prämissen in einem logisch validen Schluss wahr sind, dann überträgt sich diese Wahrheit auf die Konklusion. Allerdings unterscheidet sich die Frage ob die Konklusion wahr ist wesentlich von der Frage, ob der Schluss valide ist:

Continuation of Appendix I

2.

Ob die Prämissen wahr sind, spielt für die Beurteilung der Korrektheit des Schlusses keine Rolle.

Das hört sich zunächst sehr seltsam an, erklärt sich aber durch ein Beispiel:

Wenn ein Mensch studiert, dann hat er die Eigenschaft A.
Ein Mensch studiert.
Der Mensch hat die Eigenschaft A.

Auch das ist ein logisch valider Schluss. Aber was ist die Eigenschaft A? Dieser Ausdruck ist so vage und unbestimmt, dass man nicht beurteilen kann, ob die Prämissen wahr oder falsch sind. Der Schluss ist aber trotzdem gültig.
Dieses Merkmal führt uns direkt weiter zum dritten Merkmal:

3.

Für die Korrektheit eines Schlusses sind die Bedeutungen der in ihm vorkommenden Begriffe unwesentlich.

Das kann man durch das obige Beispiel leicht nachvollziehen. Man weiß nicht, was „Eigenschaft A" inhaltlich heißen soll, es könnte praktisch alles sein. Wenn die „Eigenschaft A" dafür steht, dass man sich an der Uni eingeschrieben hat, dann wäre die Prämisse durchaus wahr. Es könnte aber auch sein, dass die „Eigenschaft A" dafür steht, dass man eine besonders große Nase hat, dann ist die Prämisse falsch. Es studieren schließlich auch Menschen mit kleinen Nasen. Trotzdem kann man einen logischen Schluss ziehen. Die inhaltlichen Bedeutungen der Prämissen sind für das logische Schließen also vollkommen unwichtig.

Man beurteilt die logische Gültigkeit nicht nach dem Inhalt der Prämissen bzw. deren Wahrheitsgehalt. Es kommt auf die logische Form der Aussagen an. Doch was ist die logische Form einer Aussage? Wie erhält man sie?

Die logische Form einer Aussage enthält nur noch die Bestandteile einer Aussage, die für die Beurteilung der Gültigkeit des Schlusses von Bedeutung sind. Dazu wird die Aussage abstrahiert. Man überträgt die einzelnen Teile der Aussagen in eine künstliche Sprache, die aus Symbolen besteht. So sind die Aussagen dann nur noch symbolisch dargestellt.
Elemente der künstlichen Sprache sind:

- Symbole die für Aussagen stehen: P, Q, R, S, T,....
- Symbole, die die Bildung komplizierterer Sätze erlauben, z.B.:

Symbol	Bedeutung
$\wedge$	und
$\vee$	oder
$\rightarrow$	wenn, ... dann

Die oben stehenden Beispiele können nun in die Sprache der Logik übersetzt werden, wobei man die zwei Aussagen zunächst durch Symbole ersetzt:

Wenn P, dann Q
P
Q

sowie „wenn, ... dann" durch $\rightarrow$:

$P \rightarrow Q$
P
Q

Continuation of Appendix I

Durch das Abstrahieren wird die Form der Aussage also vom Inhalt der Aussage getrennt. Doch wann genau ist ein Schluss nun logisch valide?

Es gibt Regeln, die verwendet werden, um die logische Gültigkeit einer Aussage zu bestimmen. Diese Regeln richten sich nur nach der Form einer Aussage.
Eine wichtige Regel der Aussagenlogik ist der Modus Ponens:

Aus der Bejahung der ersten Aussage folgt die Bejahung der zweiten Aussage.

Die Form des Modus Ponens sieht so aus:

Wenn P, dann Q
P
___
Q

Die erste Aussage in der ersten Prämisse (P) wird in der zweiten Prämisse genannt. Die zweite Aussage taucht dann in der Konklusion auf. Die zweite Aussage folgt also aus der Ersten. Alle Schlüsse, die diese Form haben, sind **gültig!**

Ein **ungültiger** Schluss taucht auf, wenn nach der Bejahung der ersten Aussage die zweite Aussage in der Konklusion verneint wird:

Wenn es regnet, dann ist die Straße nass.
Es regnet.
___
Die Straße ist **nicht** nass.

Ist dies denn nun ein logisch valider Schluss:

Wenn jemand religiös ist, dann ist er verheiratet.
Der Papst ist religiös.
___
Der Papst ist verheiratet.

Auch dieser Schluss folgt der Form des Modus Ponens und ist somit logisch valide. Wie oben erklärt, hat der Inhalt keinen Einfluss auf die Gültigkeit der Schlüsse. So kommt es auch dazu, dass inhaltlich falsche Schlüsse, wie dieser, als logisch valide gelten.

Falls Sie jetzt noch Fragen haben, oder Ihnen noch etwas unklar ist, haben Sie keine Scheu und wenden Sie sich bitte an Ihre Versuchsleitung. Sie wird Ihre Fragen gerne beantworten.

Ansonsten können Sie nun auf der nächsten Seite mit den Übungsaufgaben beginnen.

Continuation of Appendix I

Trainings phase – problems (conclusion had to be produced by participants)	
Wenn A, dann B. A.	Wenn es Rot ist, dann ist es auch Gelb. Es ist Rot.
Wenn die Sonne scheint, dann ist es hell draußen. Die Sonne scheint.	Wenn er ein Mechaniker ist, dann kann er einen Hubschrauber fliegen. Er ist ein Mechaniker.
Wenn ein Mensch eine tiefe Wunde hat, dann verliert er viel Blut. Ein Mensch hat eine tiefe Wunde.	Wenn es eine Überschwemmung gibt, dann treiben Leichen im Wasser. Es gibt eine Überschwemmung.
Wenn Peter abends in die Kneipe geht, dann trinkt er Bier. Peter geht abends in die Kneipe	Wenn Anna eine Eins in Mathe schreibt, dann ist der Mond aus grünem Käse. Anna schreibt eine Eins in Mathe.
Wenn ein Mensch ins Schwimmbad geht, dann genießt er die Sonne. Ein Mensch geht ins Schwimmbad.	Wenn Gerda die Pflanze nicht gießt, dann wächst sie trotzdem weiter. Gerda gießt die Pflanze nicht.
Wenn ein Kind sich verletzt, dann weint es. Ein Kind verletzt sich.	Wenn Tomaten reif sind, dann sind sie blau. Die Tomaten sind reif.
Wenn der Hund eine schlimme Verletzung hat, dann wird er eingeschläfert. Der Hund hat eine schlimme Verletzung.	Wenn Ein Mensch einen schwerwiegenden Fehler macht, dann muss er mit den Konsequenzen leben. Ein Mensch macht einen schwerwiegenden Fehler.
Wenn CX, dann CY. CX.	Wenn ein Tier Beine hat, dann ist es ein Säugetier. Ein Tier hat Beine.
Wenn ein Kind schwer erziehbar ist, dann kommt die Super Nanny. Ein Kind ist schwer erziehbar.	Wenn ein Mensch faul ist, dann schreibt er schlechte Noten. Ein Mensch ist faul.
Wenn ein Mensch alt ist, dann wird er bald sterben. Ein Mensch ist alt.	Wenn Max einen Ball wirft, dann fliegt er bis in den Himmel. Max wirft einen Ball.
Wenn C rechts ist, dann ist B links. C ist rechts.	Wenn RX, dann Z9. RX.
Wenn der Apfel grün ist, dann ist die Banane auch grün. Der Apfel ist grün.	Wenn Oliver die Aufgabe nicht löst, dann ist er ein Versager. Oliver löst die Aufgabe nicht.

Continuation of Appendix I

Trainings phase – problems (given conclusions)	
valid	invalid
Wenn GD, dann WQ. GD. WQ.	Wenn H, dann K. H. Nicht K.
Wenn die Ampel rot ist, dann darf man fahren. Die Ampel ist rot. Man darf fahren.	Wenn ein Mensch einen schweren Unfall hatte, dann ist er querschnittsgelähmt. Ein Mensch hatte einen schweren Unfall. Der Mensch ist nicht querschnittsgelähmt.
Wenn man lange in der Sonne sitzt, dann bekommt man einen roten Kopf. Man sitzt lange in der Sonne. Man bekommt einen roten Kopf.	Wenn ein Mensch Krebs hat, dann muss er viel leiden. Der Mensch hat Krebs. Der Mensch muss nicht viel leiden.
Wenn ein Tier im Wasser taucht, dann hat es Kiemen. Ein Tier taucht im Wasser. Das Tier hat Kiemen.	Wenn P links ist, dann ist G rechts. P ist links. G ist nicht rechts.
Wenn die Erde der „blaue Planet“ ist, dann sind ihre Bewohner auch alle blau. Die Erde ist der „blaue Planet“. Die Bewohner der Erde sind alle blau.	Wenn der Rollladen geschlossen ist, dann scheint die Sonne ins Zimmer. Der Rollladen ist geschlossen. Die Sonne scheint nicht ins Zimmer.
Wenn Albert ins Kino geht, dann kommt Klara auch mit. Albert geht ins Kino. Klara kommt auch mit.	Wenn es Grün ist, dann ist es Weiß. Es ist Grün. Es ist nicht Weiß.
Wenn ein Mensch lügt, dann wird seine Nase länger. Ein Mensch lügt. Seine Nase wird länger.	Wenn die Zeit reif ist, dann kann man sie ernten. Die Zeit ist reif. Man kann die Zeit nicht ernten.
Wenn C vorne ist, dann ist X rechts. C ist vorne. X ist rechts.	Wenn ein Tier Flügel hat, dann ist es ein Insekt. Das Tier hat Flügel. Das Tier ist kein Insekt.
Wenn es zu einer Naturkatastrophe kommt, dann sterben viele Menschen. Es kommt zu einer Naturkatastrophe. Viele Menschen sterben.	Wenn Ina den Job nicht kriegt, dann ist sie selbst dran schuld. Ina kriegt den Job nicht. Ina ist nicht selbst dran schuld.
Wenn es eine Hungersnot gibt, dann sterben viele Kinder. Es gibt eine Hungersnot. Es sterben viele Kinder.	Wenn ein Haus brennt, dann ersticken die Menschen darin qualvoll. Ein Haus brennt. Die Menschen ersticken nicht qualvoll.
Wenn jemand Theologie studiert, dann glaubt er an Gott. Jemand studiert Theologie. Er glaubt an Gott.	Wenn ein Mensch Blut spuckt, dann kommt der Krankenwagen. Ein Mensch spuckt Blut. Es kommt kein Krankenwagen.
Wenn Hans sich freut, dann springt er hoch in die Luft. Hans freut sich. Hans springt hoch in die Luft.	Wenn ein Mensch die Aufgabe nicht lösen kann, dann ist er dumm. Ein Mensch kann die Aufgabe nicht lösen. Der Mensch ist nicht dumm.

Continuation of Appendix I

Posttest

Posttest - instruction
Lieber Teilnehmer, im folgenden Test geht es noch mal um logisches Denken. Bitte bearbeiten Sie erneut einige Aufgaben dieser Art: Wenn Hänschen in die Schule geht, dann lernt er schreiben. Hänschen geht in die Schule. ――――――――――――――― Hänschen lernt schreiben. Ist dieser Schluss logisch valide? Ja ☐ Nein ☐ Beurteilen Sie, wie bei den Übungsaufgaben, ob die Konklusion valide ist. Wenn Sie dazu keine weiteren Fragen haben, dann können Sie jetzt anfangen. Vielen Dank für Ihre Mitarbeit.

Posttest – problems	
valid	invalid
Wenn B, dann V. B. V.	Wenn T1 hinten ist, dann ist T3 seitlich. T1 ist hinten. T3 ist nicht seitlich.
Wenn ein Mensch allein im Dunkeln ist, dann hat er große Angst. Ein Mensch ist allein im Dunkeln. Der Mensch hat große Angst.	Wenn Elke ein Mitglied von Greenpeace ist, dann ist sie sehr umweltbewusst. Elke ist ein Mitglied von Greenpeace. Sie ist nicht umweltbewusst.
Wenn ein Mensch Durst hat, dann kauft er sich im Getränkemarkt etwas zu trinken. Ein Mensch hat Durst. Der Mensch kauft sich im Getränkemarkt etwas zu trinken.	Wenn die Sonne untergegangen ist, dann schlafen alle nachtaktiven Tiere. Die Sonne ist untergegangen. Die nachtaktiven Tiere schlafen nicht.
Wenn die Uhr rückwärts läuft, dann vergeht die Zeit schneller. Die Uhr läuft rückwärts. Die Zeit vergeht schneller.	Wenn man unglücklich verliebt ist, dann muss man seinen Liebeskummer im Alkohol ertränken. Man ist unglücklich verliebt. Man muss seinen Liebeskummer nicht im Alkohol ertränken.
Wenn A links ist, dann ist B rechts. A ist links. B ist rechts.	Wenn Ilse die Fenster putzt, dann sind sie wieder sauber. Ilse putzt die Fenster. Sie sind nicht wieder sauber.
Wenn ein Mensch Zigaretten raucht, dann bringt er sich damit selbst um. Ein Mensch raucht Zigaretten. Der Mensch bringt sich damit selbst um.	Wenn eine Pflanze Blätter hat, dann ist es ein Baum. Eine Pflanze hat Blätter. Es ist kein Baum.
Wenn das Universum groß ist, dann ist die Erde eine Scheibe. Das Universum ist groß. Die Erde ist eine Scheibe.	Wenn ein Mensch übergewichtig ist, dann wird er von anderen gehänselt. Ein Mensch ist übergewichtig. Der Mensch wird nicht von anderen gehänselt.
Wenn es flüssig ist, dann ist es Eis. Es ist flüssig. Es ist Eis.	Wenn ein Mensch Bohnen isst, dann wird er grün. Ein Mensch isst Bohnen. Der Mensch wird nicht grün.
Wenn jemand ein Party gibt, dann lädt er seine Freunde ein. Jemand gibt eine Party. Er lädt seine Freunde ein.	Wenn OR, dann WU. OR. Nicht WU.
Wenn es Salz im Meer gibt, dann gibt es Zucker im See. Es gibt Salz im Meer. Es gibt Zucker im See.	Wenn das Telefon klingelt, dann geht der Telefonmann ran. Das Telefon klingelt. Der Telefonmann geht nicht ran.

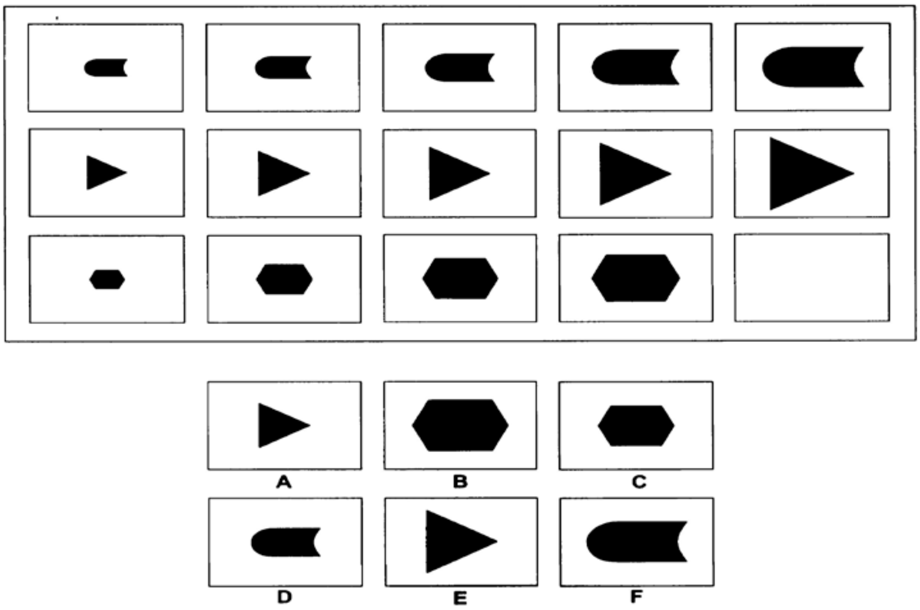
Appendix J

Material of the pseudo training in original German language

Introduction to matrice tasks

Matrizenaufgaben

Matrizenaufgaben finden besonders bei Intelligenz-Tests ihre Verwendung. Es sind Aufgaben, bei denen man mit Figuren und Formen arbeiten muss wie z.B. Kreisen, Rechtecken oder Punkten. Die Formen sind in mehreren Feldern angeordnet. Sie verändern sich nach einem logischen Prinzip von Feld zu Feld. Ein Feld ist freigelassen. Die Aufgabe besteht darin, das leere Feld logisch zu ergänzen. Dafür stehen einem mehrere Antwortalternativen zur Verfügung. Die richtige Antwort ergibt sich dadurch, dass man die Regel erkennt, nach der sich die Anordnung der Formen ändert. Hier ist ein Beispiel:



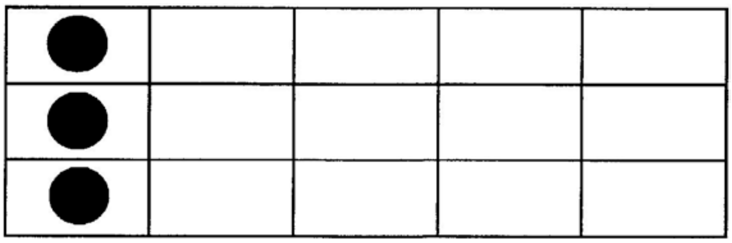
Das leere Feld im oberen Kasten muss ergänzt werden. Die richtige Lösung muss aus den unteren sechs Feldern ausgewählt werden. In diesem Fall ist „B“ die richtige Antwort.

Doch wozu braucht man solche Aufgaben überhaupt? Sie wurden entwickelt, um sprach- und rechenfreie Aufgaben bei Intelligenztests einsetzten zu können. Auf diese Weise lässt sich die Intelligenz einer Person erfassen, ohne den Einfluss ihres Vorwissens oder ihrer kulturellen Herkunft (z.B. Einwanderer die die Sprache nicht so gut beherrschen). Es kommt also nicht zu einer Benachteiligung von Menschen mit z. B. niedrigerem Bildungsniveau, die bei Tests, die Wissen im sprachlichen oder mathematischen Bereich voraussetzen, einen niedrigeren Intelligenzquotienten erzielen würden.

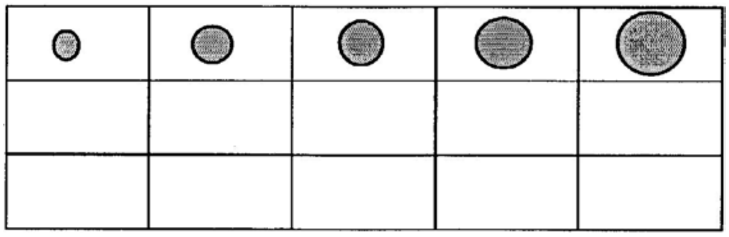
Die Aufgaben, die Sie hier bearbeiten sollen, sind dem Bochumer Matrizentest (BOMAT advanced) entnommen. Um alle Aufgaben dieses Tests lösen zu können muss man mehrere logische Regeln erkennen und anwenden, die zur korrekten Lösung führen.

Korrespondenz: Jede Zeile, Spalte oder Diagonale beinhaltet identische Figuren. Man muss also nur das passende Symbol ergänzen.

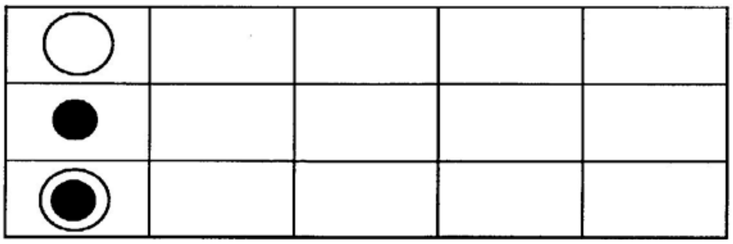
Continuation of Appendix J



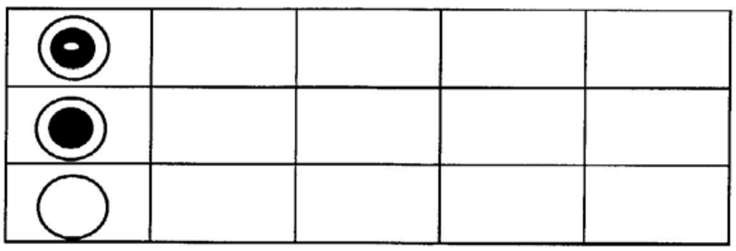
Variierende Korrespondenz: Jede Spalte, Zeile oder Diagonale beinhaltet ähnliche Grundformen, die in einer Teilkomponente (z.B. Form, Farbe, Größe) variieren. Die Grundformen verändern wie unten zu sehen, ihre Größe, manchmal aber z.B. auch ihre Farbe.



Addition: Die Figuren einer Zeile, Spalte oder Diagonalen werden übereinander gelegt. Die Symbole aus einer Spalte, Zeile oder Diagonalen werden zusammengefügt. Das letzte Feld ist die Summe aus den vorangegangenen Feldern. Alle Symbole sind in diesem Feld an der gleichen Stelle wie vorher. Hier enthält das unterste Feld sowohl den Kreis als auch den schwarzen Punkt, die einzeln in den oberen Feldern vorkommen:

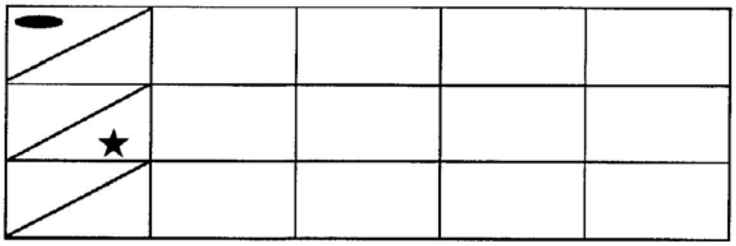


Subtraktion: Von den Elementen eines vollständigen Feldes einer Zeile, Spalte oder Diagonalen werden in den übrigen Feldern derselben Richtung Teile ausgeblendet Es werden also von Feld zu Feld Formen oder Symbole weggelassen. Hier werden von Innen nach außen Formen weggelassen:

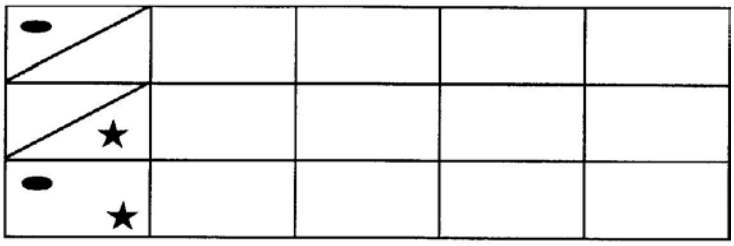


Schnittmengen: Nur die Elemente der Felder einer Zeile, Spalte oder Diagonalen werden übernommen, die auch in den anderen Feldern der Richtung enthalten sind. Hier ist es der Strich:

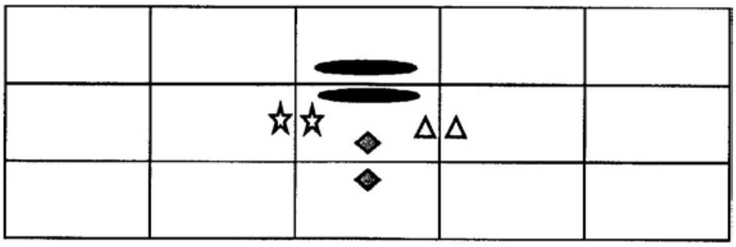
Continuation of Appendix J



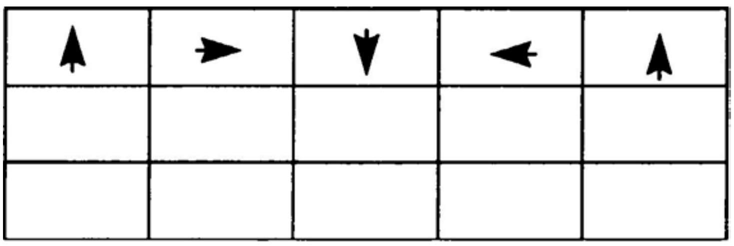
Einzelkomponentenaddition: aus den Feldern in derselben Zeile, Spalte oder Diagonalen werden nur diejenigen Teile übernommen, die nur in einem anderen Feld auftreten. Elemente, die in mehreren Feldern vorkommen, werden weggelassen oder verändert.



Nachbarschaftsprinzip/Puzzleprinzip: Hier steht jedes Feld in logischer Beziehung zu allen anderen Feldern. Man muss also alle Felder auf einmal in Augenschein nehmen, um den richtigen Lösungsweg zu finden. Hier grenzen immer die gleichen Symbole aneinander:



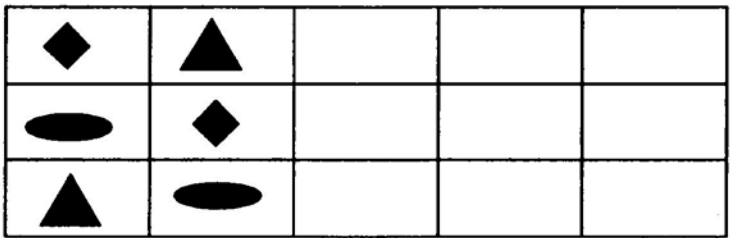
Reihung: Die Veränderungsregel (Addition, Subtraktion, Rotation etc.) setzt sich von Feld zu Feld fort. Diese Operation kann in jeder Zeile oder Spalte von vorne beginnen oder sich vertikal bzw. horizontal durch alle Felder hindurch fortsetzen. Hier dreht sich der Pfeil von Feld zu Feld um 45° weiter. Natürlich können sich auch andere Symbole drehen.



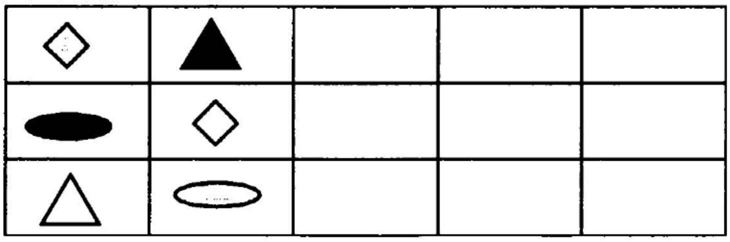
Vollständigkeit: In jeder Zeile, Spalte oder Diagonalen treten verschiedene Symbole auf, die auch innerhalb der anderen Zeilen, Spalten oder Diagonalen in unterschiedlichen Reihenfolgen vorkommen.

Continuation of Appendix J

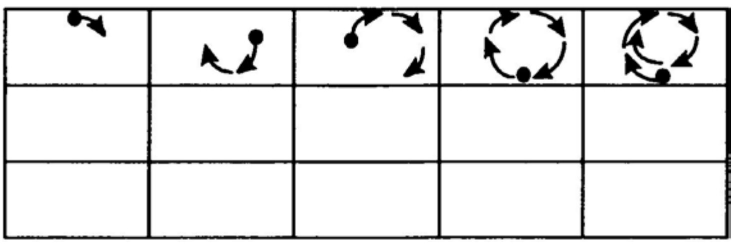
Dabei tritt jedes Symbol in der entsprechenden Richtung genau einmal auf. Wichtig ist also, dass jedes Symbol einmal vorkommt, so wie hier jeweils eine Raute, eine Ellipse und ein Dreieck.



Mehrdimensionale Vollständigkeit: In jeder Zeile, Spalte oder Diagonalen treten mehrdimensionale Symbole (variierend z.B. in Form, Farbe, Größe) auf. Jede Teilkomponente der Symbole tritt im entsprechenden Verlauf genau einmal auf, in unterschiedlicher Kombination. Hier muss man darauf achten, dass sowohl jede Eigenschaft eines Symbols jeweils einmal vorkommt, als auch jedes Symbol.

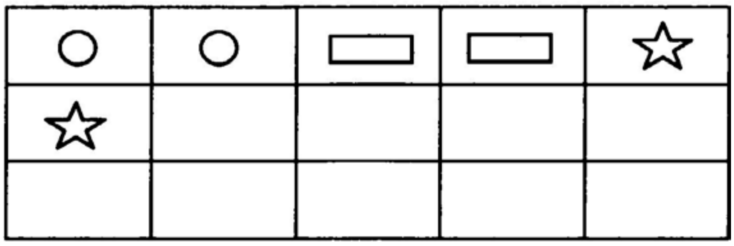


Sukkzessive Folgen: Die verschiedenen Operationen (z.B. Rotation) werden im Verlauf einer Zeile oder Spalte in kontinuierlich aufsteigender oder absteigender Anzahl durchgeführt. Die Veränderungsregel wird vom ersten zum zweiten Feld einmal angewandt, von Zweiten zum Dritten zweimal, dann dreimal usw. Der Punkt im Beispiel wird vom ersten zum zweiten Feld um 45° verschoben, dann vom Zweiten zum Dritten um 2x 45° usw. Die Veränderungsregel wird hier also in aufsteigender Reihenfolge gebraucht.

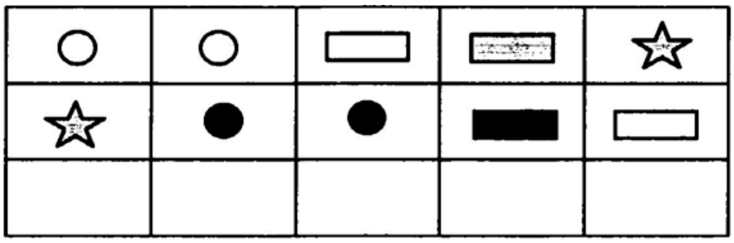


Eindimensionale Anzahlvariation: Verschiedene Symbole treten im horizontalen oder vertikalen Verlauf in einer bestimmten logisch vorgegebenen Anzahl auf, hier z.B. immer zweimal:

Continuation of Appendix J



Mehrdimensionale Anzahlvariation: Die verschiedenen Teilkomponenten mehrdimensionaler Symbole (variierend in Form, Farbe, Muster oder Größe) treten im horizontalen oder vertikalen Verlauf in einer bestimmten, logisch vorgegeben, Anzahl und/oder Reihenfolge auf. Es kommen also nicht nur die einzelnen Symbole mehrfach hintereinender vor. Auch die Eigenschaften, die ein Symbol haben kann (z.B. Form, Farbe, Größe) treten in bestimmter Anzahl mehrmals hintereinander auf. Hier gibt es drei Symbole und drei Farben, wobei die Symbole immer 2x hintereinander auftreten, die Farben 3x.



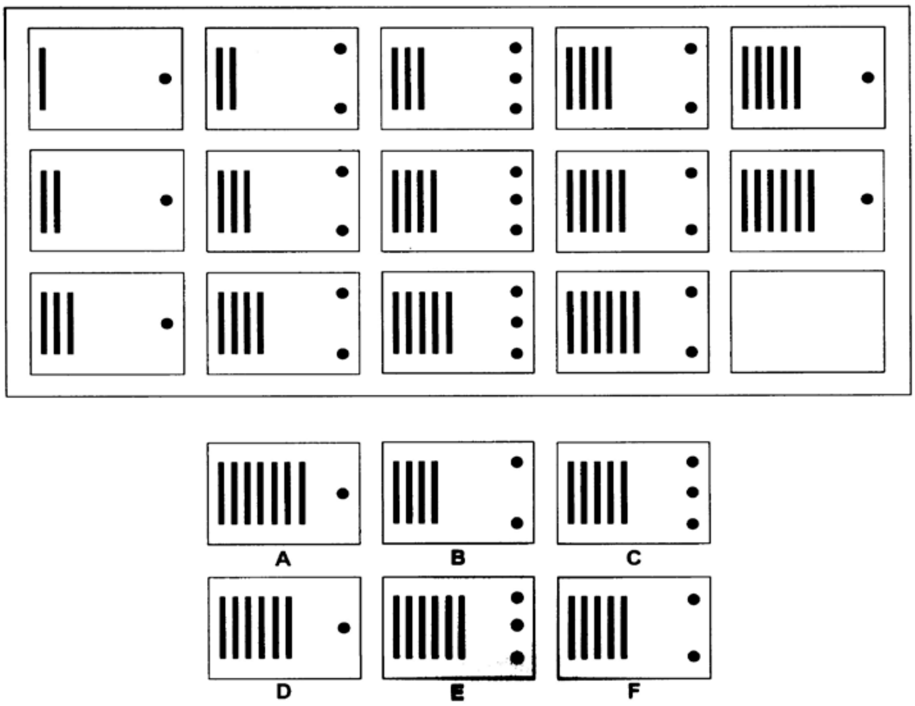
Mit diesen Regeln im Hinterkopf, nach denen sich die Felder verändern können, hat man eine gute Grundlage, um die richtige Antwort zu erschließen. Die Regeln sind eigentlich alle mathematische Prinzipien, wie Addition, Subtraktion, Schnittmengen etc., die graphisch umgesetzt werden.
Am wichtigsten ist es aber, dass man sich zunächst einen globalen Überblick über alle Felder verschafft. So kann man die Gemeinsamkeiten und Unterschiede der Felder oft schon auf den ersten Blick sehen. Sollte dies nicht der Fall sein, dann betrachtet man einzeln die Spalten, Zeilen und Diagonalen, da die regelhaften Veränderungen in jeder Richtung stattfinden können. Und manchmal kommt es ja auch nur darauf an, welche Symbole an das entsprechende Feld angrenzen.

Wenn Ihnen die Regeln alle klar sind, nach denen die Veränderungen auftreten können, und Sie keine weiteren Fragen mehr haben, beginnen Sie bitte auf der nächsten Seite mit den Aufgaben. Um die richtige Lösung zu kennzeichnen, kreuzen Sie immer den Buchstaben an, der für die entsprechende Lösung steht.

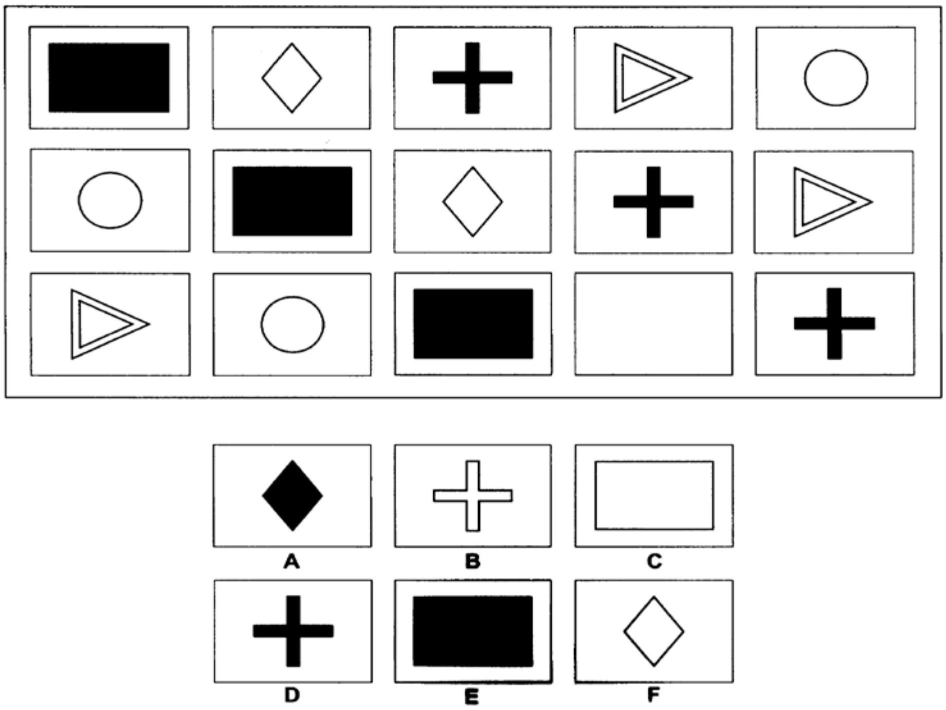
Continuation of Appendix J

Examples of the pseudo training tasks with matrices

1)



2)



Appendix K

Results of Experiment III

post hoc paired-sample t-tests	due to significant IE of content × validity				due to significant IE of content × validity of the group without training	
	decision times		error rates		error rates	
calculated comparisons	t-values (df)	p-values	t-values (df)	p-values	t-values (df)	p-values
moral-v vs. moral-iv	t(35) = -1.674	p = .103	t(35) = -4.353	p < .001	t(11) = -4.690	p = .001
moral-v vs. unmoral-v	t(35) = 3.199	p = .003	t(35) = -2.907	p = .006	t(11) = -2.345	p = .039
moral-v vs. neutral1-v	t(35) = 4.546	p < .001	t(35) = .620	p = .539	t(11) = -.616	p = .551
moral-v vs. neutral2-v	t(35) = 3.266	p = .002	t(35) = 2.707	p = .010	t(11) = 1.000	p = .339
moral-iv vs. unmoral-iv	t(35) = -.507	p = .616	t(35) = 3.820	p = .001	t(11) = 4.690	p = .001
moral-iv vs. neutral1-iv	t(35) = 3.132	p = .003	t(35) = 4.587	p < .001	t(11) = 5.000	p < .001
moral-iv vs. neutral2-iv	t(35) = 2.362	p = .024	t(35) = 5.686	p < .001	t(11) = 5.933	p < .001
unmoral-v vs.unmoral-iv	t(35) = -5.039	p < .001	t(35) = 2.201	p = .034	t(11) = 2.171	p = .053
unmoral-v vs. neutral1-v	t(35) = 1.460	p = .153	t(35) = 3.839	p < .001	t(11) = 2.419	p = .034
unmoral-v vs. neutral2-v	t(35) = -.088	p = .930	t(35) = 4.620	p < .001	t(11) = 3.189	p = .009
unmoral-iv vs. neutral1-iv	t(35) = 4.519	p < .001	t(35) = .723	p = .475	t(11) = -.432	p = .674
unmoral-iv vs. neutral2-iv	t(35) = 3.134	p = .003	t(35) = 2.092	p = .044	t(11) = .000	p = 1.000
neutral1-v vs. neutral1-iv	t(35) = -2.318	p = .026	t(35) = -.683	p = .499	t(11) = .432	p = .674
neutral1-v vs. neutral2-v	t(35) = -1.957	p = .058	t(35) = 1.784	p = .083	t(11) = 1.483	p = .166
neutral2-v vs. neutral2-iv	t(35) = -1.380	p = .176	t(35) = -1.673	p = .103	t(11) = -1.000	p = .339
neutral1-iv vs. neutral2-iv	t(35) = -.873	p = .389	t(35) = 1.405	p = .169	t(11) = .561	p = .586

Continuation of Appendix K

post hoc independent-sample t-tests	due to significant three-way IE of content × validity × group	
	error rates	
calculated comparisons	t-values (df)	p-values
moral-v groups no training vs. pseudo training	t(22) = -.405	p = .689
moral-v groups no training vs. logic training	t(22) = .364	p = .719
moral-v groups pseudo training vs. logic training	t(22) = .715	p = .482
moral-iv groups no training vs. pseudo training	t(18.319) = 1.164	p = .259
moral-iv groups no training vs. logic training	t(22) = 4.220	p < .001
moral-iv groups pseudo training vs. logic training	t(22) = 2.187	p = .040
unmoral-v groups no training vs. pseudo training	t = .482 (22)	p = .635
unmoral-v groups no training vs. logic training	t(13.792) = 2.378	p = .032
unmoral-v groups pseudo training vs. logic training	t(22) = 2.055	p = .052
unmoral-iv groups no training vs. pseudo training	t(15.572) = -1.239	p = .234
unmoral-iv groups no training vs. logic training	t(22) = .920	p = .368
unmoral-iv groups pseudo training vs. logic training	t(13.939) = 1.797	p = .094
neutral1-v groups no training vs. pseudo training	t(15.957) = 1.301	p = .212
neutral1-v groups no training vs. logic training	t(13.832) = 1.701	p = .111
neutral1-v groups pseudo training vs. logic training	t(22) = .596	p = .557
neutral1-iv groups no training vs. pseudo training	t(22) = .000	p = 1.000
neutral1-iv groups no training vs. logic training	t(22) = 1.119	p = .275
neutral1-iv groups pseudo training vs. logic training	t(22) = .810	p = .427
neutral2-v groups no training vs. pseudo training	t(22) = .596	p = .557
neutral2-v groups no training vs. logic training	t(11) = 1.483	p = .166
neutral2-v groups pseudo training vs. logic training	t(11) = 1.000	p = .339
neutral2-iv groups no training vs. pseudo training	t(22) = .432	p = .670
neutral2-iv groups no training vs. logic training	t(11) = 2.345	p = .039
neutral2-iv groups pseudo training vs. logic training	t(11) = 1.915	p = .082

Continuation of Appendix K

post hoc independent-sample t-tests	due to significant IE of content × group	
	decision times	
calculated comparisons	t-values (df)	p-values
moral groups no training vs. pseudo training	t(22) = -.403	p = .691
moral groups no training vs. logic training	t(22) = 1.096	p = .285
moral groups pseudo training vs. logic training	t(22) = 1.272	p = .221
unmoral groups no training vs. pseudo training	t(22) = -.282	p = .781
unmoral groups no training vs. logic training	t(22) = 2.083	p = .049
unmoral groups pseudo training vs. logic training	t(22) = 1.857	p = .077
neutral1 groups no training vs. pseudo training	t(22) = 1.154	p = .261
neutral1 groups no training vs. logic training	t(22) = 3.672	p = .001
neutral1 groups pseudo training vs. logic training	t(22) = 1.768	p = .091
neutral2 groups no training vs. pseudo training	t(22) = 1.242	p = .227
neutral2 groups no training vs. logic training	t(22) = 3.089	p = .005
neutral2 groups pseudo training vs. logic training	t(22) = 1.549	p = .136

Appendix L

Results of Experiment IV

post hoc Wilcoxon signed rank tests	due to significant IE content × validity		due to significant IE content × validity of average intelligence group	
	error rates		error rates	
calculated comparisons	z-values	p-values	z-values	p-values
moral-v vs. moral-iv	z = -1.983	p = .047	z = -1.983	p = .047
moral-v vs. unmoral-v	z = -2.041	p = .041	z = -1.633	p = .102
moral-v vs. neutral1-v	z = -1.342	p = .180	z = -1.342	p = .180
moral-v vs. neutral2-v	z = -1.342	p = .180	z = -1.342	p = .180
moral-iv vs. unmoral-iv	z = -1.222	p = .222	z = -1.466	p = .143
moral-iv vs. neutral1-iv	z = -1.403	p = .161	z = -1.403	p = .161
moral-iv vs. neutral2-iv	z = -2.060	p = .039	z = -2.060	p = .039
unmoral-v vs.unmoral-iv	z = -1.406	p = .160	z = -1.300	p = .194
unmoral-v vs. neutral1-v	z = -1.511	p = .131	z = -1.069	p = .285
unmoral-v vs. neutral2-v	z = -1.511	p = .131	z = -1.069	p = .285
unmoral-iv vs. neutral1-iv	z = .000	p = 1.000	z = -.378	p = .705
unmoral-iv vs. neutral2-iv	z = -.828	p = .408	z = -.557	p = .577
neutral1-v vs. neutral1-iv	z = -.378	p = .705	z = -.378	p = .705
neutral1-v vs. neutral2-v	z = .000	p = 1.000	z = .000	p = 1.000
neutral2-v vs. neutral2-iv	z = -1.414	p = .157	z = -1.414	p = .157
neutral1-iv vs. neutral2-iv	z = -1.134	p = .257	z = -1.134	p = .257

Continuation of Appendix L

Correlations with 6 participants

		Ø DT (un)moral	Ø DT neutral	Ø DT all	c-Score	CFT	IST
group	r_s	.386	.579*	.435	-.772**	-.878**	-.877**
	Sig. (2-tailed)	.215	.048	.158	.003	.000	.000
Ø DT (un)moral	r_s	1	.811**	.958**	-.552	-.636*	-.420
	Sig. (2-tailed)	.	.001	.000	.063	.026	.174
Ø DT neutral	r_s		1	.902**	-.755**	-.728**	-.480
	Sig. (2-tailed)		.	.000	.005	.007	.114
Ø DT all	r_s			1	-.650*	-.686*	-.434
	Sig. (2-tailed)				.	.014	.159

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

		Ø Er (un)moral	Ø Er neutral	Er all	c-Score	CFT	IST
group	r_s	.803*	.689*	.842**	-.772**	-.878**	-.877**
	Sig. (2-tailed)	.002	.013	.001	.003	.000	.000
Ø Er (un)moral	r_s	1	.577*	.946**	-.716**	-.797**	-.755**
	Sig. (2-tailed)	.	.049	.000	.009	.002	.005
Ø Er neutral	r_s		1	.789**	-.840**	-.635*	-0.453
	Sig. (2-tailed)		.	.002	.001	.027	.139
Ø Er all	r_s			1	-.839**	-.792**	-.709**
	Sig. (2-tailed)			.	.001	.002	.010

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

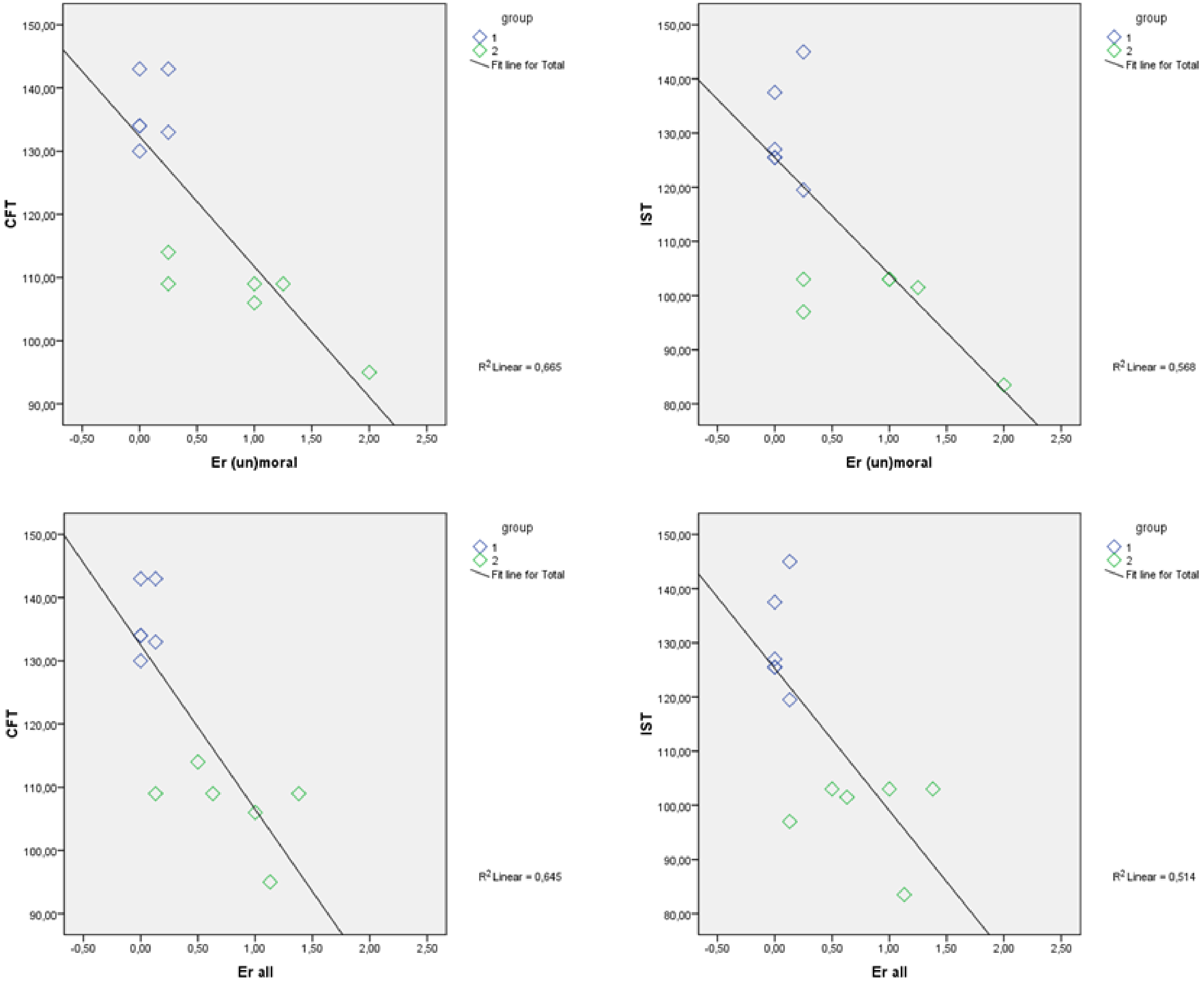
Continuation of Appendix L

		c-Score	CFT	IST
group	r_s	-.772**	-.878**	-.877**
	Sig. (2-tailed)	.003	.000	.000
C-Score	r_s	1	.827**	.621*
	Sig. (2-tailed)	.	.001	.031
CFT	r_s		1	.863**
	Sig. (2-tailed)		.	.000

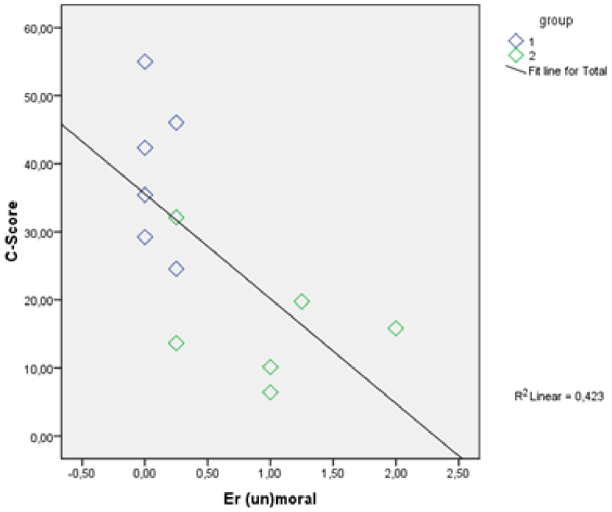
*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

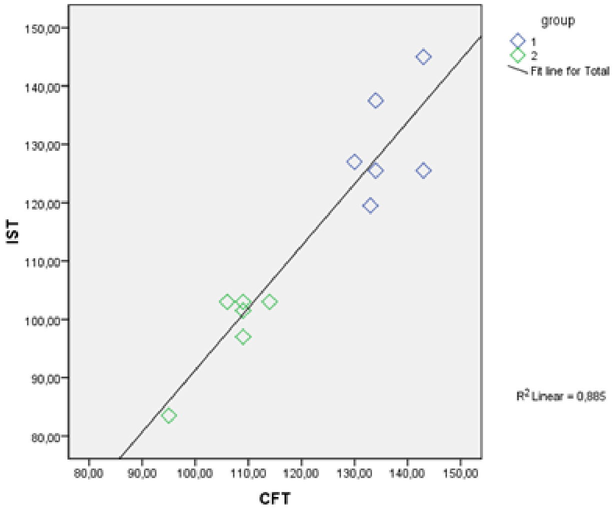
Continuation of Appendix L

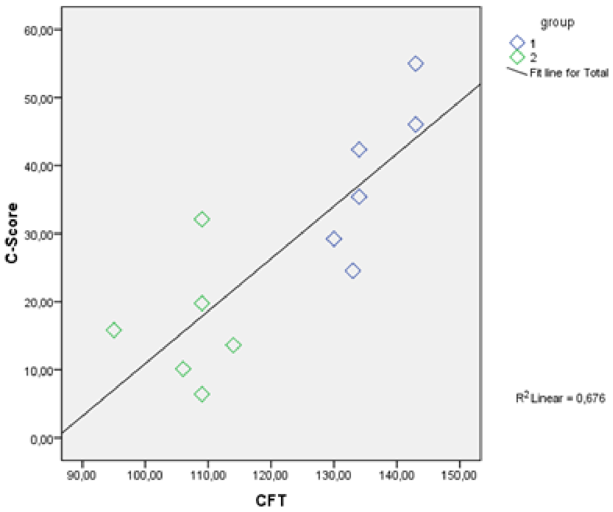
Correlation scatterplots with regression line based on 6 participants (1 = high intelligence group, 2 = average intelligence group)

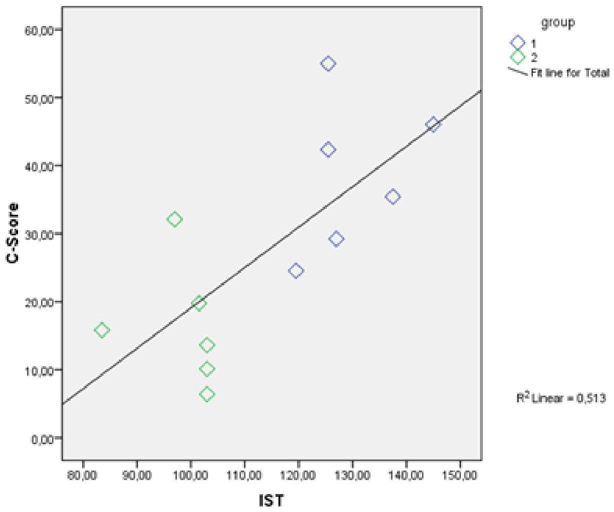


Continuation of Appendix L









Appendix M

Results of Experiment V (behavioral data)

post-hoc paired sample t-test	due to significant ME of content			
	error rates		decision times	
calculated comparisons	t-values (df)	p-values	t-values (df)	p-values
moral vs. unmoral	t = .779 (30)	p = .442	t = -.199 (30)	p = .844
moral vs. neutral	t = 3.906 (30)	p < .001	t = 4.445 (30)	p < .001
moral vs. abstract	t = 2.979 (30)	p = .006	t = 6.900 (30)	p < .001
moral vs. positive emotional	t = 2.843 (30)	p = .008	t = 3.765 (30)	p = .001
moral vs. negative emotional	t = 1.283 (30)	p = .209	t = 3.072 (30)	p = .004
unmoral vs. neutral	t = 4.119 (30)	p < .001	t = 4.882 (30)	p < .001
unmoral vs. abstract	t = 2.923 (30)	p = .007	t = 8.058 (30)	p < .001
unmoral vs. positive emotional	t = 2.695 (30)	p = .011	t = 4.262 (30)	p < .001
unmoral vs. negative emotional	t = 1.022 (30)	p = .315	t = 3.103 (30)	p = .004
neutral vs. abstract	t = -1.541 (30)	p = .134	t = 3.612 (30)	p = .001
neutral vs. positive emotional	t = -1.366 (30)	p = .182	t = -1.458 (30)	p = .155
neutral vs. negative emotional	t = -4.150 (30)	p < .001	t = -2.753 (30)	p = .010
abstract vs. positive emotional	t = -.441 (30)	p = .662	t = -5.417 (30)	p < .001
abstract vs. negative emotional	t = -2.794 (30)	p = .009	t = -4.780 (30)	p < .001
positive emotional vs. negative emotional	t = -2.306 (30)	p = .028	t = -1.244 (30)	p = .223

Continuation of Appendix M

post hoc paired-sample t-tests	due to significant IE of content × validity error rates				
calculated comparisons	t-values (df)	p-values	calculated comparisons	t-values (df)	p-values
moral-v vs. moral-iv	t(30) = -2.650	p = .013	unmoral-iv vs. abstract-iv	t(30) = .828	p = .414
moral-v vs. unmoral-v	t(30) = -3.069	p = .005	unmoral-iv vs. positive-emotional-iv	t(30) = -.465	p = .645
moral-v vs.neutral-v	t(30) = 1.976	p = .057	unmoral-iv vs. negative-emotional-iv	t(30) = -2.006	p = .054
moral-v vs. abstract-v	t(30) = .000	p = 1.000	neutral-v vs. abstract-v	t(30) = -1.976	p = .057
moral-v vs. positive-emotional-v	t(30) = 1.000	p = .325	neutral-v vs. positive-emotional-v	t(30) = -.626	p = .536
moral-v vs. negative-emotional-v	t(30) = -.528	p = .601	neutral-v vs. negative-emotional-v	t(30) = -2.528	p = .017
moral-iv vs. unmoral-iv	t(30) = 2.919	p = .007	neutral-iv vs. abstract-iv	t(30) = .000	p = 1.000
moral-iv vs.neutral-iv	t(30) = 3.702	p = .001	neutral-iv vs. positive-emotional-iv	t(30) = -1.409	p = .169
moral-iv vs. abstract-iv	t(30) = 3.503	p = .001	neutral-iv vs. negative-emotional-iv	t(30) = -3.276	p = .003
moral-iv vs. positive-emotional-iv	t(30) = 2.832	p = .008	abstract-v vs. positive-emotional-v	t(30) = .902	p = .374
moral-iv vs. negative-emotional-iv	t(30) = 1.938	p = .062	abstract-v vs. negative-emotional-v	t(30) = -.528	p = .601
unmoral-v vs.unmoral-iv	t(30) = 3.025	p = .005	abstract-iv vs. positive-emotional-iv	t(30) = -1.541	p = .134
unmoral-v vs.neutral-v	t(30) = 3.758	p = .001	abstract-iv vs.negative-emotional-iv	t(30) = -3.478	p = .002
unmoral-v vs. abstract-v	t(30) = 2.752	p = .010	positive-emotional-v vs. positive-emotional-iv	t(30) = -.571	p = .572
unmoral-v vs. positive-emotional-v	t(30) = 3.202	p = .003	positive-emotional-v vs. negative-emotional-v	t(30) = -1.153	p = .258
unmoral-v vs. negative-emotional-v	t(30) = 2.376	p = .024	positive-emotional-iv vs. negative-emotional-iv	t(30) = -2.065	p = .048
unmoral-iv vs.neutral-iv	t(30) = .828	p = .414	negative-emotional-v vs. negative-emotional-iv	t(30) = -1.030	p = .311

Appendix N

Results of Experiment V (brain imaging data)

Analysis of single effects

	abstract						**neutral**						**moral**					
ROIs	**cluster-size**	**t-value**	**x**	**y**	**z**	**BAs**	**cluster-size**	**t-value**	**x**	**y**	**z**	**BAs**	**cluster-size**	**t-value**	**x**	**y**	**z**	**BAs**
ACC	47	5.54	-3	18	27	BA 24	117	12.16	-6	24	27	BA 32	27	4.95	-6	24	27	BA 32
	115	6.34	6	27	30	BA 32	203	15.57	9	21	30	BA 32	80	5.74	9	24	30	BA 32
DLPFC	248	7.88	-33	-6	60	BA 6	400	20.85	-36	-6	57	BA 4	102	7.27	-36	-6	57	BA 4
	75	6.47	9	21	45	BA 6	280	16.52	9	21	45	BA 6	20	6.77	9	21	45	BA 6
mPFC	43	7.93	-3	21	45	BA 6	81	18.85	-3	21	45	BA 6	59	7.65	-3	21	45	BA 6
	193	8.69	3	24	45	BA 32	236	21.13	0	18	42	BA 6	168	8.02	0	18	42	BA 6
MTG	303	6.17	-54	-54	9	BA 39	530	14.42	-54	-33	0	/	285	6.83	-54	-30	-3	BA 21
	419	6.00	60	-51	-6	BA 37	344	9.66	48	-27	-3	BA 22	36	4.83	48	-27	-3	BA 22
OFC	27	6.45	-51	18	-6	BA 47	341	16.53	-33	24	-6	BA 13	118	7.74	-30	24	-6	/
	267	8.43	48	18	-6	BA 13	521	15.60	33	24	-6	/	120	6.62	48	21	-6	BA 47
Parietal	1244	11.26	-42	-48	54	BA 40	1426	19.27	-36	-51	48	BA 40	695	8.59	-36	-51	48	BA 40
	1122	10.06	36	-57	48	BA 7	1254	16.56	36	-57	51	BA 7	525	6.81	33	-57	48	BA 7
Precuneus	145	6.21	-9	-69	48	BA 7	332	11.96	-21	-66	42	BA 7	114	5.47	-21	-69	45	BA 7
	176	5.99	9	-66	48	BA 7	361	11.84	12	-72	45	BA 7	132	5.09	12	-72	48	BA 7
STS	1030	11.26	-42	-48	54	BA 40	1287	19.27	-36	-51	48	BA 40	543	8.59	-36	-51	48	BA 40
	961	10.06	36	-57	48	BA 7	842	16.56	36	-57	51	BA 7	359	6.81	33	-57	48	BA 7
Temporal Pole	40	6.57	-57	9	0	BA 22	142	15.49	-54	15	-6	BA 22	70	7.24	-51	21	-12	BA 47
	70	7.85	51	15	-3	BA 13	146	14.12	51	18	-9	/	63	5.99	48	21	-12	BA 47

Continuation of Appendix N

Conjunction analysis

ROIs	cluster-size	t-value	x	y	z	BAs
ACC	25	4.79	-3	24	27	BA 32
	76	5.73	9	24	30	BA 32
DLPFC	102	7.27	-36	-6	57	BA 4
	2	4.40	-15	3	63	BA 6
	1	4.12	-33	54	12	BA 10
	1	4.07	-27	0	45	BA 6
	18	6.47	9	21	45	BA 6
	1	4.08	9	18	57	BA 6
mPFC	42	7.65	-3	21	45	BA 6
	159	8.02	0	18	42	BA 6
MTG	145	5.89	-51	-51	9	BA 39
	36	4.83	48	-27	-3	BA 22
OFC	17	6.96	-33	24	-6	BA 13
	27	6.45	-51	18	-6	BA 47
	112	6.62	48	21	-6	BA 47
Parietal	691	8.59	-36	-51	48	BA 40
	514	6.81	33	-57	48	BA 7
Precuneus	92	5.47	-21	-69	45	BA 7
	95	5.09	12	-72	48	BA 7
	1	3.98	15	-69	36	BA 7
STS	536	8.59	-36	-51	48	BA 40
	1	4.25	-51	-30	-3	BA 21
	358	6.81	33	-57	48	BA 7
Temporal Pole	34	6.52	-51	15	-9	BA 22
	47	5.99	48	21	-12	BA 47

Continuation of Appendix N

Difference comparisons between conditions

abstract - moral

	cluster-size	t-value	x	y	z	BAs
ACC	10	4.36	-6	15	27	BA 24
	5	4.10	0	12	30	BA 24
	5	3.75	6	27	27	BA 32
DLPFC	99	5.28	-30	-12	63	BA 6
	1	3.92	-27	6	57	BA 6
	22	4.83	12	15	57	BA 6
	7	4.28	27	54	0	BA 10
mPFC	4	4.07	-3	21	45	BA 6
	43	4.85	3	24	45	BA 32
	3	4.07	12	24	57	BA 6
MTG	14	5.14	-54	-48	-9	BA 37
	34	4.74	-48	-69	3	BA 37
	1	4.13	-63	-54	18	BA 22
	351	6.46	60	-51	-6	BA 37
OFC	37	5.81	48	18	-6	BA 13
	19	4.50	42	51	-6	/
Parietal	970	9.15	-51	-45	51	BA 40
	915	7.81	36	-57	48	BA 7
Precuneus	17	4.22	-12	-66	48	BA 7
	1	3.91	-18	-72	48	BA 7
	22	4.66	6	-66	48	BA 7
STS	814	9.15	-51	-45	51	BA 40
	2	4.13	-63	-54	18	BA 22
	934	7.81	36	-57	48	BA 7
Temporal Pole	2	3.87	-57	9	0	BA 22
	34	5.72	51	15	-3	BA 13

abstract - neutral

	cluster-size	t-value	x	y	z	BAs
ACC	1	3.63	-3	39	27	BA 9
	1	3.85	0	36	30	BA 6
DLPFC	1	3.97	-27	12	54	BA 6
	15	4.33	9	24	57	BA 6
mPFC	7	4.13	9	27	60	BA 6
	3	4.01	-3	36	33	BA 6
	7	3.99	3	30	48	BA 8
MTG	73	4.97	-54	-48	-9	BA 37
	382	7.18	60	-51	-9	BA 37
	2	4.05	51	-27	-6	BA 22
OFC	53	5.42	48	18	-6	BA 13
	26	4.35	33	54	-12	BA 10
	1	4.07	24	24	-9	/
Parietal	845	7.88	-45	-45	51	BA 40
	9	4.63	-60	-21	24	BA 40
	904	7.18	36	-57	42	BA 39
Precuneus	24	4.23	-9	-66	48	BA 7
	2	4.08	-18	-72	48	BA 7
	20	4.45	6	-66	48	BA 7
STS	723	7.88	-45	-45	51	BA 40
	8	4.63	-60	-21	24	BA 40
	4	4.40	-63	-57	18	BA 22
	910	7.18	36	-57	42	BA 39
Temporal Pole	19	4.91	51	15	-3	BA 13

moral - neutral

	cluster-size	t-value	x	y	z	BAs
ACC	9	4.13	-3	51	18	BA 9
	3	4.01	0	51	15	BA 9
DLPFC	2	4.07	9	57	33	BA 8
mPFC	60	4.69	-6	54	33	BA 8
	84	4.82	6	57	33	BA 8
MTG	8	4.36	-63	-12	-21	BA 21
OFC	3	4.17	-42	24	-18	BA 47
Parietal	32	4.79	-48	-63	36	BA 39
STS	32	4.79	-48	-63	36	BA 39

Continuation of Appendix N

Difference comparisons between conditions

moral - abstract							neutral - abstract							neutral - moral
	cluster-size	t-value	x	y	z	BAs		cluster-size	t-value	x	y	z	BA's	
DLPFC	3	4.05	-15	54	33	BA 8	Temporal Pole	1	3.91	-54	15	-24	BA 38	no significances
mPFC	64	5.09	-9	54	36	BA 8		2	3.86	-60	6	-18	BA 21	
MTG	64	5.26	-57	-9	-15	BA 21								
Precuneus	6	4.09	-6	-51	24	BA 31								
STS	8	5.26	-57	-9	-15	BA 21								
Temporal Pole	40	5.09	-57	3	-18	BA 21								

(Statutatory declaration)

Eidesstattliche Erklärung

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

Gießen,

Patrick S. Wiedenmann