



universität
wien

DIPLOMARBEIT / DIPLOMA THESIS

Titel der Diplomarbeit / Title of the Diploma Thesis

„Theory of Mind and empathy in moral judgments: a study
with autism and alexithymia“

verfasst von / submitted by

Marie-Christine Mayer

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Magistra der Naturwissenschaften (Mag. rer. nat.)

Wien, 2015 / Vienna, 2015

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

A 298

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Betreut von / Supervisor:

PhD Giorgia Silani

Table of Contents

Table of Contents	2
Abstract.....	3
Introduction.....	4
Morality.....	4
Autism and Alexithymia	5
Moral Dilemmas	6
Method	10
Participants.....	10
Materials and Procedure.....	13
Research Design.....	21
Results	23
Comparability of demographics	25
Questionnaire Results	26
Moral dilemma results	33
Correlations of dilemma results and questionnaire results.....	36
Correlations of emotional arousal with behaviour and judgment in dilemmas.....	40
Discussion.....	40
Primary hypothesis.....	41
Secondary hypotheses (prespecified).....	43
Secondary hypotheses (exploratory)	45
Conclusions	46
References.....	48
Appendix.....	52
German Abstract	52

Abstract

This study's aim was to further investigate autism spectrum disorders and distinguish the roles of Theory of Mind (ToM) and empathy in the repeatedly observed high frequency of utilitarian responses (i.e. actively sacrificing a person for a greater good) in moral tasks in this population. For exploratory purpose emotional arousal and emotion regulation tendencies were investigated additionally. Using moral dilemmas, alexithymia questionnaires (Toronto Alexithymia Scale – 26, Bermond Vorst Alexithymia Scale), an empathy and ToM questionnaire (Saarbrücker Persönlichkeitsfragebogen) and a computerized test (Multi-dimensionaler Empathietest Condensed and Revised), as well as a questionnaire about emotion regulation strategies (Emotion Regulation Questionnaire), 18 autistic and 18 non-autistic participants were tested. While no significant differences between autistic and healthy participants could be identified in moral decisions, empathy was generally lower in autistic participants, especially in alexithymic autistic participants, than in the control group and ToM deficits were detected in the experimental group as expected. Emotional arousal self-reports were far higher in the autistic than the non-autistic sample. Regarding emotion regulation the autistic group showed less of a tendency to apply cognitive reappraisal, which is an approach to deal with unpleasant or difficult situations by focussing on positive rather than negative aspects of it. This did seem to produce utilitarian answer behaviour in specific moral tasks. More research is needed to identify the roles of ToM and empathy in autistic persons and their behaviour in moral tasks since this study's small sample was not able to resolve this question.

Introduction

Morality

Moshman (2011) views morality as a "culture's standards regarding the proper way to live. (...) morality is cultural conformity; one learns to be moral by internalizing the values of one's culture" (pp. 66-67). The author adds however, that Piaget (1932/1965, p. 72, as cited in Moshman, 2011, p. 70) argued morality should not be imposed by culture, but that children learned through socially interacting that what was fair and logic was cooperation and social equality, as opposed to acting out of selfishness. According to Haidt (2007) morality deals with how people treat each other and how they participate in groups. Olatunji and Puncochar (2014) sum it all up by describing morality roughly as judging "whether a behaviour is right or wrong" (p. 186). However precise and detailed one may define morality, it is a concept that lies at the heart of our social lives and highly influences the decisions we make and the opinions we form.

In the literature reviewed for this paper, authors identified that the making of moral decisions is attributed to a cognitive approach, as well as an emotional approach of the information available to the individual (e.g. Moshman, 2011, pp. 93-96; Bargh, 1999, as cited in Haidt, 2007). In the cognitive approach the person making the decision is required to take perspective, i.e. putting him- or herself into other persons' situation and assuming other people's points of view in an attempt to understand how and why certain actions are taken and identify the other person's perception and reasoning (Moshman, 2011, pp. 93-96; Bargh, 1999, as cited in Haidt, 2007). In 2000, Hoffmann described the emotional approach as recognizing other people's feelings and sharing these emotions, alternatively as "psychological

processes that make a person have feelings that are more congruent with another's situation than with his [or her] own situation" (p.30, as cited in Moshman, 2011, p. 95). In short, moral decisions seem to be closely related to Theory of Mind (ToM) capabilities, as well as empathy (Bzdok et al., 2012); Bzdok et al. (2012) confirmed this assumption in their meta-analysis of neuroimaging studies.

Concerning the process of making moral decisions, studies have investigated the connection of ToM and moral judgment (e.g. Moran et al., 2011), as well as empathy and moral judgment (e.g. Crockett et al., 2010). To my knowledge, there have never been experiments that studied both ToM and empathy simultaneously.

Autism and Alexithymia

The American Psychiatric Association defines autism spectrum disorder as persistent deficits in social communication and social interaction across multiple contexts, including deficits in social reciprocity, nonverbal communicative behaviors used for social interaction, and skills in developing, maintaining, and understanding relationships. In addition to the social communication deficits, the diagnosis of autism spectrum disorder requires the presence of restricted, repetitive patterns of behavior, interests, or activities. (2013, p. 31)

Recent work in social neuroscience (for a review, see Bird & Cook, 2013) has shown that autism is associated with deficits in ToM but not empathy, while alexithymia is associated with impaired empathy but not ToM (Patil & Silani, 2014). However, there is a high prevalence rate of alexithymia in the autistic population (about 50%, Hill, Berthoz, & Frith, 2004). Alexithymia is an

emotion dysregulation [that] includes difficulties in identifying and describing feelings; difficulties in distinguishing feelings from the bodily sensations of emotional arousal; impaired symbolization, as evidenced by a paucity of fantasies and other imaginative activity; and a tendency to focus on external events rather than inner experiences (concrete thought; Sifneos, 1973, as cited in Hill et al., 2004).

Since autism is connected to both low ToM skills, as well as empathy deficits in part of the population, a sample of autistic participants compared to a non-clinical control group can hopefully shed light on both the roles of ToM skills, as well as levels of empathy in moral judgments when the influence of alexithymia is surveyed concurrently (Patil & Silani, 2014).

Moral Dilemmas

In moral dilemmas participants are confronted with stories that contain moral conflicts and involve risk of harm to at least one person (Christensen, Gomilla, 2012). In their review of dilemma tasks used in research the authors state that participants usually have to decide between two possible outcomes: whether to stay inactive and therefore not save a group of typically five people, or whether to get active and harm one person for the sake of rescuing the other group of people. In this thesis I will examine responses to non-moral, personal and impersonal dilemmas. Non-moral dilemmas are tasks where a decision between two alternatives has to be made, but where no moral context or possible harm of another person is included, i.e. a neutral dilemma where no ToM skills or empathy supposedly contribute to the responses given. These will be included to have emotionally neutral tasks, which in theory should be responded to in no significantly different way by autistic and non-

autistic participants, regardless of the presence of alexithymia as a comorbidity. Personal and impersonal dilemmas differ in their innate directness of harm, as well as their extent of personal distance between the participant and the potential victim (Christensen, Gomilla, 2012). Consider this as an example of an impersonal 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. (Greene et al., 2001)

In a personal version of this trolley dilemma the personal distance to the one individual that can be sacrificed for the greater good is decreased: instead of hitting a switch, the participant has to decide whether to push a person standing on a footbridge above the rails in front of the trolley in order to stop it from hitting the group of five people (Greene et al., 2001). According to the authors most people chose to get active in impersonal dilemmas, but to stay passive in personal dilemmas. Greene et al. (2001) claim that this dissimilarity in reaction is due to different levels of emotion evoked by the personal or impersonal demands of the dilemmas.

Responses that serve a greater good, in most dilemmas the rescue of five people at the cost of one person, are called utilitarian; refusing to make such a sacrifice is called a non-utilitarian response (Gleichgerricht & Young, 2013).

Gleichgerricht et al. (2013) studied autistic and non-autistic participants in a moral study and found that overall in personal dilemmas non-utilitarian responses were more likely in participants who empathized with the individual victim in question; in impersonal scenarios utilitarian responses are more likely than non-utilitarian decisions in non-autistic as well as autistic participants. Furthermore Miller, Hannikainen and Cushman (2014) found that cognitively putting oneself in the

position of doing harm, which refers to ToM abilities, also makes it less likely to respond with a utilitarian judgment. Thus both reduced empathy and deficits in ToM seem to lead to utilitarian judgments. However, in their study with autism Gleichgerrecht et al. (2013) argued that the high frequency of utilitarian judgments in the autistic experimental group was a result of low ToM capabilities, but since they did not control for alexithymia, utilitarian responses due to a lack of empathy cannot be ruled out. The study described in this thesis did control for alexithymia and is therefore able to distinguish between the roles of ToM skills as well as levels of empathy in moral judgments. These differences were assessed by comparing moral decisions of autistic to non-autistic participants and correlating the respective results with alexithymia-questionnaire-scores. Because of the association of autism and deficits in ToM (Bird & Cook, 2013) autistic participants were still expected to show more utilitarian responses, especially for personal dilemmas (Gleichgerrecht & Young, 2013), than the non-clinical control group, even when alexithymia was controlled for.

To test whether there actually are significant differences between autistic and alexithymic autistic moral judgments those two groups were compared concerning their frequency of utilitarian responses and results for empathy and ToM skill tests and questionnaires. Within the alexithymic autistic group more utilitarian responses were expected than within the non-alexithymic autistic group, since the first group per definition has both empathy and ToM deficits, which both make utilitarian responses more likely (Gleichgerrecht & Young, 2013; Miller et al., 2014). Non-alexithymic autists per definition only have ToM deficits, but no lack of empathy (Bird & Cook, 2013).

By comparing the effects of autism and alexithymia on moral judgments an estimate for the amount of influence of both ToM and empathy on moral judgments can be made.

Finally, participants were asked about their degree of arousal experienced after each dilemmatic situation. The levels of arousal of the experimental group and the control group were compared in the main analysis to find out if empathy and ToM skills have different effects on arousal and if on a more general basis autistic and non-autistic participants show different reports of arousal. According to the literature (Moretto & Làdavas, 2010; as cited in Gleichgerrcht & Young, 2013; Navarret, McDonald, Mott & Asher, 2011; as cited in Gleichgerrcht & Young, 2013) utilitarian judgments may be predicted by emotional arousal: the lower the level of arousal, the more likely utilitarian responses are.

Thus the primary hypothesis was concerned with whether there were differences in moral decision making (in relation to behaviour and moral judgment) between autistic and non-autistic participants. Secondary hypotheses attended to possible explanations of the expected differences in the 2 groups as to (1) whether there was a connection between ToM capabilities and moral decision making; (2) whether there was a connection between alexithymia, more general empathy levels, and moral decision making; and finally 2 exploratory theses, specifically (3) whether there were differences in reported emotional arousal between the 2 groups, and (4) whether there were connections between emotional regulation tendencies and moral decision making.

Method

Participants

To make sure that the instructions, texts and questions of the psychological tests described later in this chapter are fully understood by the participants, the experimental group consisted of autistic participants with average or higher than average intellectual abilities (i.e. high-functioning autism and Asperger's Syndrome). People with an autism spectrum disorder accompanied by intellectual impairments were excluded. The participants for both the experimental and the control group were adults from the age of 17 to 70 years. The choice of this age group was based on the following considerations: firstly, people below the age of 17 years are assumed not to be fully developed with regards to the relevant cognitive abilities and moral values; secondly, people above the age of 70 years may show increased risk of ongoing gradual intellectual decline and a lack of relevant computer skills. To achieve a sample that allows generalizations of results the participants were desired and chosen to be evenly distributed regarding their levels of education. Since autism is four times more often diagnosed amongst men than women (American Psychiatric Association, 2013, p. 57), the participants of the autistic group were expected but not necessarily sought to mirror this statistic. The search for the control group's members was matched to the clinical sample's demographics regarding sex, age, levels of education and intelligence parameters. The autistic participants were required to have valid clinical diagnoses of autism, whereas the healthy control group only consisted of subjects that scored below the cut off value of a screening questionnaire for autistic traits (AQ-k, Freitag et al., 2007). The experimental group's members were mostly found and contacted through the help of Austrian organisations for autists, such as

Autistenhilfe and *Rainman's Home*, as well as clinical institutions like hospitals and clinical psychologists. Advertisements via posters and internet platforms were utilized as well for the recruitment of both experimental and control group partakers. Additionally, acquaintances from the study leaders [supervisors] were recruited for the control group.

The considered candidates for the experimental group included 20 men and 3 women between the ages of 17 and 64 years ($M = 32.8$, $SD = 12.8$), all of whom identified themselves with the autism spectrum disorder, but not all attendees turned out to have a diagnosis of autism. All the individuals lived in or close to Vienna, only very few people that lived farther than a two hour trip of the testing site in Vienna were interested to take part in the study and eventually all potential candidates that would have faced a longer journey stated that the trip was too stressful and too long for them and thus dropped out.

Among the group of 23 self-appointed autists 18 had valid diagnoses and were therefore included in the study. Twelve provided state of the art diagnoses (i.e. were tested with the “Diagnostische Beobachtungsskala für Autistische Störungen”, in short ADOS, and the interview for parents or other attachment figures “Diagnostisches Interview für Autismus”, in short ADI), 2 only provided diagnoses based on the ADOS and 4 participants only offered diagnoses of Asperger's Syndrome issued by medical centres which sufficed for this study's purpose as well.

In the experimental group 15 participants were males and 3 were females between the age of 18 to 65 years ($M = 34.17$, $SD = 12.67$); a Chi-square test ruled out equal distribution of gender ($p = .005$). Levels of education ranged from (a) compulsory education, (b) apprenticeship, (c) higher education without a university entrance diploma, (d) higher education with a university entrance diploma, (e) a

Bachelor's degree to (f) a Master's degree and were equally distributed, $\chi^2(5, 18) = 3.33, p = .649$.

Seven of the autistic participants had a diagnosis of a mental disorder besides an autism spectrum disorder diagnosis, most of which were depression or anxiety disorder, and 8 reported to be on medication.

For the sample of the control group 39 people were tested for the variables to be matched to the autistic group (sex, age, levels of education and intelligence parameters). Of those 39 people approximately 46%, i.e. 18 participants, were included in the study. In accordance to the experimental group, 15 participants of the control group were male and 3 were female between the ages of 17 to 67 years ($M = 34.33, SD = 13.37$). Levels of education matched the levels of the experimental group, ranging from (a) compulsory education to (f) Master's degrees and were equally distributed within the control group, $\chi^2(4, 18) = 4.78, p = .311$.

One participant of the control group stated that she was on medication for a diagnosed depression.

Most of the participants were German native speakers. However, 2 of the autistic experimental group and 1 of the healthy control group were fluent but non-native German speakers. The frequency of non-native speakers was not significantly different for the two groups, $\chi^2(1, 36) = 0.36, p = .546$.

The experiments took place at the testing rooms of the psychological faculty of the University of Vienna. These testing rooms provided a quiet environment while being within short reach of public transport. The testing sessions were held for one person at a time, while there were up to 2 participants tested parallel in 1 room. Two sittings per participant of the experimental group were conducted, each lasting around 1 hour, while the members of the control group were tested on 3 separate occasions,

since they had to undergo a short pre-screening before being selected for the main part of the study, altogether likewise lasting around 2 hours.

Autistic subjects were paid 25€ and healthy subjects 20€ for the attendance of the main study and all partakers gave their written consent to be part of this study after being informed of the procedure in detail. The study was approved by the local Ethics Committee (University of Vienna) and conducted in accordance with the declaration of Helsinki.

Materials and Procedure

A data sheet covered the needed information of demographics like sex, age, level of education, clinical diagnoses and current medications.

Measurements of autistic spectrum characteristics to confirm a diagnosis of autism were acquired with the German translation of the short version of the **Autismus Spektrum Quotient-Kurzversion** (AQ-k; Freitag et al., 2007). The AQ-k assesses the 3 scales (a) *Social interaction and spontaneity*, (b) *Phantasy and imagination* and (c) *Communication and reciprocity*. Thirty-three items consisting of 1 statement each have to be rated on a 4-point Likert scale (Definitely Agree, Slightly Agree, Slightly Disagree and Definitely Disagree). To calculate the scoring the number of answers affirming typically autistic traits get summed up. Freitag et al. (2007) indicate the inner consistencies of the three scales vary between a Cronbach's Alpha of .65 and .87; the retest reliability for 20 subjects is numbered at .79. The authors state that the AQ-k's external validity is confirmed by comparison to various questionnaires collecting data of psychopathology characteristics; furthermore the AQ-k is able to distinguish between parents of children with autistic spectrum disorders and parents of mentally healthy children, the first showing mild signs of

autistic traits like their children, while the parents of clinically sane children do not (Freitag et al., 2007). Instructions and scoring of the AQ-k are standardized and thus objective.

Fluid intelligence data was assembled with the **Standard Progressive Matrices** (SPM; Raven, Kratzmeier & Horn, 1988) and crystallized intelligence data with the **Mehrfachwahl Wortschatz-Intelligenztest** (MWT-B; Lehrl, 1995). The SPM is a nonverbal and culture free cognitive ability test which consists of 5 sets of 12 graphic items each; these items show geometric patterns or figures that have to be logically completed with one out of 6 to 8 proposed figures in an unlimited amount of time (Heller, Kratzmeier & Lengfelder, 1998). The difficulty of the tasks increases over the course of the test, only 1 of the proposed choices is correct in each of the items (Heller et al., 1998). Correct answers are summed up, items 1 to 3 being sample problems for an illustrative instruction are not included in the final score; thus a maximum of 9 points can be achieved (Heller et al., 1998). Heller et al. (1998) state that the SPM's inner consistency for total scores is high (Cronbach's $\alpha = .75-.96$), while retest-reliability for several samples after 3 months is quantified as $\alpha = .72-.91$; the SPM's results correlate quite high with general intelligence tests such as the Hamburg-Wechsler Intelligenztest für Erwachsene and one-dimensional measurement across the sets of 12 items were confirmed by factor analyses; therefore validity is presumed to be fulfilled. Since instructions and scoring guidelines are specified in the manual, objectivity of the SPM's implementation is assumed. For the present study the paper-pencil version of the SPM was used.

While the SPM attempts to quantify a person's inherent cognitive ability (i.e. fluid intelligence, Heller et al., 1998) the MWT-B takes measurement of abilities that are learnt through experiences and education (i.e. crystallized intelligence), which

makes it a highly culture dependent multiple choice vocabulary test (Lehrl, 1995). The MWT-B consists of 37 lines, each line constituting 1 item; per line 5 words are given, of which only 1 word really exists; this 1 correct word has to be marked by the proband in order to get 1 point; 37 points can be accomplished in total, however the items' difficulty rises during the course of the test (Lehrl, 1995). In the test's manual Lehrl (1995) indicates the MWT's reliability can be assumed acceptable since test-retests do not show systematic variations in results and validity is relatively high when comparing the MWT to other global intelligence tests ($r = .71$). Instructions and scoring procedures are standardized, which makes the MWT-B an objective intelligence test.

Alexithymia-scores were attained with two questionnaires to heighten the assessment's accuracy and validity, first the **Toronto-Alexithymie-Skala-26** (TAS-26; Kupfer et al., 2001) and second the Bermond-Vorst Alexithymieskala (BVAQ; Bermond, et al. 1999). The TAS-26 assesses 3 scales: (a) *Difficulties in identifying emotions*, (b) *Difficulties in describing emotions* and (c) *Externally oriented cognitive style*. The questionnaire consists of 26 items that have to be answered on a 5-point Likert scale (Strongly Agree, Agree, Undecided, Disagree, Strongly Disagree). The evaluation of the scorings is done by adding up the scores of the respective items. The sums of the 3 scales are added up to a global alexithymia-scale. Evidence of reliability and validity is provided by the Manual (Kupfer et al., 2001); inner consistency of the three scales lies between Cronbach's Alpha of .67 and .84, inner consistency of the global alexithymia scale is stated to amount to a Cronbach's Alpha of .81; validity is reported to be verified by factor analyses and numerous correlations with various subscales of other sentiment questionnaires. The cut-off score for alexithymia is designated at 54 points, making results of 53 and below non-alexithymic (Kupfer et

al., 2001). Objectivity can be assumed since instructions and scoring are standardized and are specified in the manual.

The **Bermond-Vorst Alexithymieskala** (BVAQ; Bermond et al., 1999) comprises 40 items which have to be answered on a 5-point Likert scale (1 = certainly does not apply to me, 2-4 are not labelled, 5 = certainly does apply to me) and are structured evenly into 5 scales: (a) *Verbalizing*, (b) *Fantasizing*, (c) *Identifying*, (d) *Emotionalizing* and (e) *Analyzing*. The scales *Identifying*, *Analyzing* and *Verbalizing* make up factor 1 (*cognitive factor*); the scales *Emotionalizing* and *Fantasizing* make up factor 2 (*affective factor*; Vorst & Bermond, 2001). The results of the scales are built by adding up the degrees of agreement for each question and the sum of all scales results in a total alexithymia scale (Vorst & Bermond, 2001). Internal consistency for the German version of the BVAQ-40 in total is acceptable (Cronbach's $\alpha = .83$), the different subscales however range from .54 to .80, the two scales *Fantasizing* and *Emotionalizing* scoring below .60 (Müller, Bühner & Ellgring, 2004). Müller et al. (2004) determined the BVAQ's validity by correlating it with the standard measure of self-report questionnaires, the TAS-20 ($r = .62$); comparing the BVAQ to the widespread personality questionnaire NEO-FFI resulted in similar results to the comparison of the TAS-20 and the NEO-FFI (Bagby, Taylor et al., 1994, as cited in Müller et al., 2004), which the authors interpret as evidence for comparable validity of the BVAQ-40 and the TAS-20. While the BVAQ treats alexithymia as a continuous variable (Vorst & Bermond, 2001), Deborde et al. (2008) developed cut-off scores for the BVAQ-B 20 item short version, which consists of items 21 to 40 of the BVAQ-40; scores of 43 or lower are considered non-alexithymic, while scores of 53 or higher are considered alexithymic. Guidelines for

instructions and scoring are standardized and therefore the BVAQ is considered an objective measure.

The participants' empathy levels and ToM-capabilities were estimated with the help of the **Multi-dimensional Empathietest Condensed and Revised** (MET-CORE-2; Dziobek et al., 2008) and the German version of the IRI, the Saarbrücker Persönlichkeitsfragebogen (SPF; Paulus, 2009). The MET-CORE-2 is a 40 item photo-based computerized test which attempts to measure emotional and cognitive empathy, much like ToM-capabilities in relation to empathic functions, with both non-social stimuli as well as people-depicting materials; the photos have to be rated in respect to emotions they portray, as well as emotional reactions they stimulate (Dziobek et al, 2008). Each item consists of a pair of context and people stimulus; after each context picture participants are questioned about their initial level of arousal on a 10-point Likert scale (0 = very calm, 1-8 are not labelled and 9 = very aroused); in form of a multiple choice answering format (with 4 possible answers, only one being correct) for each people picture probands are asked about the displayed person's mental state which forms the subscale *cognitive empathy*; subsequently the correct answer is revealed and the participants are required to rate the degree of their empathic concern (*emotional empathy*; Dziobek et al., 2008). The authors claim that the use of photos rather than solely questions to evaluate ones' own empathy plus the use of implicit questions concerning emotional reactions to the photos strongly benefits ecological validity. Dziobek et al. (2008) found construct validity to be satisfactory when comparing the initial version of the MET to the Interpersonal Reactivity Index (IRI, i.e. the SPF in English), which is described in the following paragraph. Internal consistency is reported to be acceptably high (Cronbach's $\alpha = .71$ for the cognitive empathy scale, $\alpha = .91$ for the explicit emotional

scale and $\alpha = .92$ for the implicit emotional scale; Dziobek et al., 2008). Instructions and scoring are produced by the computer program, thus high objectivity is assumed.

Likewise the **Saarbrücker Persönlichkeitsfragebogen** (SPF; Paulus, 2009) assesses both an emotional and a cognitive scale of empathy. The questionnaire consists of 16 items that encompass 1 statement each, which have to be rated on a 5-point Likert scale (1 = never, 2 = rarely, 3 = sometimes, 4 = often, 5 = always; Paulus, 2009). The 16 items are evenly distributed among 4 subscales: (a) *perspective taking*, (b) *fantasy*, (c) *empathic concern* and (d) *personal distress*; (a) *perspective taking* means the tendency to consider another person's point of view, which is the only cognitive subscale of empathy in the SPF, (b) *fantasy* measures the intensity with which a person identifies with characters in movies or books, i.e. to imagine oneself in fictional situations (c) *empathic concern* entails questions about feelings evoked by orientating oneself on other people, for example compassion for an unfortunate person and (d) *personal distress* seeks to measure negative emotions such as unease or nervousness in a social interaction context; subscales b, c and d are emotional empathy constructs (Paulus, 2009). For scoring the values of the from 1 to 5 marked answers simply have to be added up. Concerning reliability Paulus (2009) arrives at a Cronbach's Alpha of .78, which he values satisfactory considering the small amount of items; according to the author a split-half analysis with Spearman-Brown-correction delivers a highly acceptable result as well ($r_{tt} = .80$). Paulus (2009) claims that internal and external validity are considered proven as a result of various analyses. Since instructions and scoring are standardized, objectivity is presumed.

Emotional regulation processes were measured with the German version of Gross' and John's **Emotion Regulation Questionnaire** (ERQ; Abler & Kessler, 2009). The ERQ consists of the 2 scales *cognitive reappraisal* and *expressive suppression*; while

cognitive reappraisal refers to changing one's thoughts about an emotion-triggering situation by focusing on positive rather than negative aspects in order to feel better, for example thinking about what the achievement of a difficult exam will enable after passing it, rather than focussing on the stressful part of undergoing the test, *expressive suppression* refers to inhibiting signs of emotional reactions visible to the outside world, e.g. suppressing the impulse to smile when dealt a great hand in a card game (Gross & John, 2003). Gross and John (2003) developed the ERQ as a 10-item paper-pencil-questionnaire, 6 items measuring *cognitive reappraisal* and 4 items assessing *expressive suppression*. Every item is made up of a statement which has to be rated on a 7-point Likert scale; for scoring the given ratings are simply summed up for each scale separately (Gross & John, 2003). Abler and Kessler (2009) examined reliability for the German version of the ERQ with a factor analysis and found inner consistencies of .76 for the *cognitive reappraisal*-scale and .74 for the *expressive suppression*-scale. Validity for the German version of the ERQ is assumed and supported by first comparisons to other established tests, but still has to be further explored (Abler & Kessler, 2009). Gross and John (2003) assessed validity of the original English version of the ERQ by comparing ERQ-results to a suppression index ($r = .53$) and a reappraisal index ($r = .24$) with moderately high results. Objectivity is assured due to standardized instructions and scoring rules.

Moral judgments were presented in a computerized form of **moral dilemmas** adapted from previous studies (Greene, Nystrom, Engell, Darley & Cohen, 2004; Moore, Clark & Kane, 2008). These dilemmas were phrased in third person and included 10 impersonal and 20 personal dilemmas, some stories incorporating inevitable harm and some scenarios incorporating avoidable harm. Non-moral dilemmas, i.e. situations where a decision has to be made, but without any moral

context or possible harm of another person, were included as well to measure differences in basic decision making. Additionally participants were asked if they would take action in the scenario and if this suggested action seemed morally acceptable to them (each of those two questions could be answered by clicking either *yes* or *no*). After each dilemma a question about the level of arousal was asked as well. Here a star had to be placed on a continuing scale ranging from *not at all aroused* to *very much aroused*. Scoring was done by the computer program and provided percentages of affirmative decisions and moral acceptance for each of the 3 types of dilemmas, the average extent of arousal per type of dilemma, as well as average reaction times of the two decisions concerning behaviour and judgment of the behaviour, separately for affirmative and negated decisions. According to Christensen and Gomilla (2012) moral dilemma tasks are widely used, but lack methodological homogeneity, which makes estimates concerning reliability and validity difficult to impossible. However the authors state in their effort to standardize moral tasks with regard to achieving quality criteria for psychological tests that ecological validity is supposedly highest when using dilemmas with extreme situations; in accordance to the authors hypothesis this present study included items that fit this description. Since instructions and scoring were computerized, objectivity can be presumed assured.

Parallel to the experiment described in this thesis, another moral judgments task differentiating between empathic and ToM abilities in moral decision making focussing on harm and intentions was performed by the same participants in the same setting. This task took about the same time as the dilemma-task and was incorporated into the research design. Questionnaires were shared for the two experiments.

Research Design

The study was conducted as a between- (autistic and non-autistic participants), as well as a within- (three types of dilemmas had to be dealt with by each participant) subjects design. Autistic participants were assigned to the experimental group, persons without an autism spectrum disorder were assigned to the control group. Both groups were asked to fill out 8 questionnaires, 7 of which were paper-pencil-tests plus 1 computerized test (MET-CORE-2), additionally to complete 2 moral tasks. For the experimental group these in total 10 tests were distributed to 2 testing sessions about 1 week apart from each other. Each first testing session started with a detailed description of the study and a declaration of consent, followed by a data sheet to fill out and subsequently questionnaires in randomized orders, 1 of those being 1 of the 2 intelligence tests in turns, and ended with one moral task, moral dilemmas and moral intention tasks alternately at the first or the second testing session for participants consecutively tested. In order to encourage participants to take part in the second session as well and to keep motivation as high as possible, the first part was completed with a computerized test including a lot of pictures which was hoped to conclude the testing with a positive feeling; on top of that the second testing sessions was designed to last a shorter amount of time than the first. Additionally to make the implementation as error free as possible, 2 groups of questionnaires were formed, the first containing 3 tests and the second containing 2 tests; this way randomization was facilitated. Table 1 shows the research design for the experimental group in detail.

Table 1

Research design experimental group

TESTING SESSION 1	TESTING SESSION 2
Description of the study and form of consent	
Data sheet	
Intelligence test: Standard Progressive Matrices (SPM; Raven, Kratzmeier & Horn, 1988) or Mehrfachwahl Wortschatz-Intelligenztest (MWT-B; Lehrl, 1995)	Intelligence test: Standard Progressive Matrices (SPM; Raven, Kratzmeier & Horn, 1988) or Mehrfachwahl Wortschatz-Intelligenztest (MWT-B; Lehrl, 1995)
Questionnaire group 1: a) Autismus Spektrum Quotient-Kurzversion (AQ-k; Freitag et al., 2007) b) Toronto-Alexithymie-Skala-26 (TAS-26; Kupfer et al., 2001) d) Saarbrücker Persönlichkeitsfragebogen (SPF; Paulus, 2009)	Questionnaire group 2: a) Bermond-Vorst Alexithymieskala (BVAQ; Bermond et al., 1999) c) Emotion Regulation Questionnaire (ERQ; Abler & Kessler, 2009)
Moral task: Dilemmas or Intentions	Moral task: Dilemmas or Intentions
Multi-dimensional Empathietest (MET-CORE-2; Dziobek et al., 2008)	

For the control group participants were invited thrice since a pre-screening was needed to determine whether potential candidates were comparable to the experimental group concerning sex, age, highest education and intelligence parameters, additionally that they showed no evidence of autistic traits (i.e. they scored below the AQ-k cut-off value). Apart from the short description and consent form, which was always presented first, the questionnaires' order was randomized. Testing sessions 1 and 2 of the main study were carried out as in the experimental group, with the exception of leaving out the already completed pre-screenings' questionnaires.

Table 2

Research design control group

PRE-SCREENING	
Short description of the study and form of consent for the pre-screening	
Data sheet	
Intelligence tests: Standard Progressive Matrices (SPM; Raven, Kratzmeier & Horn, 1988) and Mehrfachwahl Wortschatz-Intelligenztest (MWT-B; Lehrl, 1995)	
Autismus Spektrum Quotient-Kurzversion (AQ-k; Freitag et al., 2007)	
TESTING SESSION 1	TESTING SESSION 2
Detailed description of the study and form of consent for the main study	
Questionnaire group 1: a) Toronto-Alexithymie-Skala-26 (TAS-26; Kupfer et al., 2001) b) Saarbrücker Persönlichkeitsfragebogen (SPF; Paulus, 2009)	Questionnaire group 2: a) Bermond-Vorst Alexithymieskala (BVAQ; Bermond et al., 1999) c) Emotion Regulation Questionnaire (ERQ; Abler & Kessler, 2009)
Moral task: Dilemmas or Intentions	Moral task: Dilemmas or Intentions
Multi-dimensional Empathietest (MET-CORE-2; Dziobek et al., 2008)	

Results

Statistical analyses were performed with SPSS 22.0 with a significance level of $\alpha = .05$. For analyses of differences between the autistic and the non-autistic group in their response behaviour in the moral dilemmas, as well as their self-reported levels of arousal mixed model ANOVAS for repeated measures were calculated with the group affiliation (autistic or non-autistic) as the between-subject factor and the three types of dilemmas (non-moral, impersonal and personal) as the within-subject factor. Each statistical test was preceded by tests for required assumptions, in the case of

mixed model ANOVAS the dependent variable has to be measured at a continuous level, the within-subject factor has to consist of at least two categorical related groups (the three types of dilemmas), the between-subject factor (i.e. the independent variable) has to consist of at least two independent groups (autistic experimental group and non-autistic control group), there should be no significant outliers in any group of the within-subjects factor and the between-subjects factor, the dependent variable should be approximately normally distributed for each combination of the groups of the two factors, which was tested with the Kolmogorov-Smirnov-Test, there has to be homogeneity of variances for each combination of the groups of the two factors, which was tested with the Levene's Test, the sphericity has to be equal for all factors, which was tested with the Mauchly's Test. According to Zöfel (2003) mixed model ANOVAs are quite resistant to violations of normal distribution; following the author's instructions if normal distribution was not given an F -value was not yet considered critical when exceeding a $p = 0.05$, instead values of $p = 0.04$ or less were regarded as critical. Had the assumption of homogeneity of variances been violated, the respective level of significance would have been decreased to $p = 0.01$ to compensate for the inhomogeneity of variances (Zöfel, 2003), but in this analysis the homogeneity of variances was fulfilled in all cases. If the assumption of sphericity was violated, Greenhouse-Geisser-corrected degrees of freedom were reported. Partial eta-squared η^2 were calculated for results of the mixed model ANOVA; *partial* $\eta^2 = .01$ is considered a small, *partial* $\eta^2 = .06$ a moderate and *partial* $\eta^2 = .14$ a large effect (Cohen, 1988). Post-hoc tests were performed with a Bonferroni-correction to minimize the alpha error.

Considering the statistical power of a mixed model ANOVA the analysis was continued in spite of a few assumption violations, however due to the small sample sizes results have to be interpreted with caution.

To pinpoint specific relationships between dilemma results and various questionnaire results Pearson-correlations were performed if data was interval scaled and normally distributed, otherwise Spearman correlations were performed separately for the experimental and the control group. According to Cohen's rule of thumb (1988) correlations were interpreted equivalent to effect sizes and as such correlations of $r = .10$ were considered small, $r = .30$ moderate and $r = .50$ large associations.

For analysing whether 2 distributions were comparable to each other or whether data was equally distributed within a given group, chi-square tests were conducted, normal distribution was conducted with Kolmogorov-Smirnov-Tests outside of the mixed model ANOVA as well.

Comparisons of questionnaire results of the experimental and the control group were executed with t -tests for independent samples when the assumptions of normal distribution of data in each group, data was interval scaled and variances were homogenous, were fulfilled; if this was not the case, Mann-Whitney U-Tests were used instead.

Comparability of demographics

Control and experimental group had to meet the criterion of comparability concerning sex, age, highest education levels and intelligence parameters. All requirements were met, the intelligence parameters are displayed in detail in the section of questionnaire and test results.

While gender was unequally distributed within the groups, it was exactly equally distributed in the control and the experimental group with 15 males and 3 females each (Fisher's exact test $p = 1$).

An independent-samples t-test was conducted to compare the distribution of age between the experimental and the control group and showed no significant difference; $t(34) = 0.038, p = .970$.

The control group was not significantly different from the experimental group concerning the variable highest education, $\chi^2(5, 36) = 8.53, p = 0.129$.

Questionnaire Results

Fourteen of the 18 autistic participants scored above the cut off value of 17 of the autistic questionnaire **AQ-k** by Freitag et al. (2007) with a minimum of 5 points and a maximum of 30 points for the whole experimental group ($M = 21.33, SD = 7.14$); all the members of the control group scored below the cut off value of the autistic questionnaire with a minimum of 1 point and a maximum of 11 points ($M = 5.33, SD = 2.83$); the total scores of the 2 groups differed significantly ($t(34) = -8.84, p < .001$). Regarding the AQ-k's 3 subscales the participating 2 groups revealed the following results: for subscale (a) *Social interaction and spontaneity* with a maximum of 11 achievable points the experimental group came to a minimum of 0 points and a maximum of 11 points ($M = 7.33, SD = 3.58$), while the control group's scores only ranged from 0 to 4 points ($M = 1.00, SD = 1.28$); the results of the two groups were significantly different from each other ($U = -4.38, p < .001$); for subscale (b) *Phantasy and imagination* a maximum of 12 points is attainable; while the experimental group's results ranged from 1 to 11 points ($M = 7.22, SD = 2.76$), the control group's participants scored between 0 and 5 points ($M = 1.78, SD = 1.73$); the

Mann-Whitney-U-Test again confirmed a highly significant difference for the two groups ($U = -4.44, p < .001$); for subscale (c) *Communication and reciprocity* the autistic group scored between 2 and the highest possible 10 points ($M = 6.78, SD = 2.37$), the clinically healthy group attained 1 to 6 points in total ($M = 2.56, SD = 1.38$); once more the results differed significantly ($t(34) = -6.54, p < .001$).

Cognitive abilities were assessed in 2 parts: fluid and crystalline intelligence. For fluid intelligence measured with the **SPM** by Raven et al. (1988) a maximum of 9 points was attainable, the results of the experimental group ranged from 4 to 9 points with a means of 7.33 and a standard deviation of 1.5; the results of the control group ranged identically to the autistic group from 4 to 9 points, but with a mean of 7.28 and a standard deviation of 1.36. Fluid intelligence measured with the SPM was not significantly different from the autistic and the non-autistic group ($t(34) = -0.12, p = .908$).

Regarding crystalline intelligence measured with the **MWT-B** by Lehrl (1995) a maximum of 37 points could be achieved, here the experimental group's results ranged from 20 to 36 with a mean of 30.33 and a standard deviation of 4.04; in the control group results ranged from 23 to 35 with a mean of 29.78 and a standard deviation of 3.21. Crystalline intelligence results quantified with the MWT-B were comparable for both groups ($t(34) = -0.46, p = .651$).

Alexithymia-related results according to the **TAS-26** (Kupfer et al., 2001) came to the following results: with a total achievable sum of 90 points the experimental group's members scored between 35 and 69 points ($M = 50.39, SD = 8.58$) and the control group's results ranged from 23 to 49 points ($M = 34.22, SD = 6.24$); these results differed significantly from each other ($t(34) = -6.47, p < .001$); concerning the subscale (a) *Difficulties in identifying emotions* with a maximum of 35

points possible the autistic group attained between 10 and 26 points ($M = 19.00$, $SD = 4.90$), while the healthy control group's results varied from 7 to 17 points ($M = 10.00$, $SD = 2.47$); the two groups had significantly different results ($U = -4.59$, $p < .001$); concerning the subscale (b) *Difficulties in describing emotions* 25 points can be reached in total; the experimental group reached 10 to 24 points ($M = 18.78$, $SD = 3.81$) and the control group scored 5 to 18 points ($M = 11.22$, $SD = 3.10$); again the two groups showed significantly different response styles in this subscale ($t(34) = -6.53$, $p < .001$); similar results however were achieved in the subscale with a maximum of 30 attainable points (c) *Externally oriented cognitive style*: participants of the experimental group reached 9 to 22 points ($M = 12.61$, $SD = 3.27$) and the control group's members attained 9 to 17 points ($M = 13.00$, $SD = 2.57$); these results were not significantly different ($t(34) = 0.40$, $p = .694$). According to the questionnaire's cut-off value of 54 points (Kupfer et al., 2001) seven members of the autistic group can be considered alexithymic, however none of the control group exceeds the cut-off value.

Alexithymia-related results according to the **BVAQ** (Bermond et al., 1999) were as follows: 200 points can be scored in total and 40 points per subscale; the experimental group's results ranged from 80 to 150 points ($M = 113.52$, $SD = 21.10$), while the control group's scores varied between 59 and 116 points ($M = 90.39$, $SD = 16.08$); the total scores for the BVAQ were significantly different in the two groups ($t(34) = -3.70$, $p = .001$). Regarding the subscale (a) *Verbalizing* the autistic group described themselves with 15 to 38 points ($M = 29.78$, $SD = 6.25$), the non-autistic group with 11 to 30 points ($M = 18.83$, $SD = 5.44$); again these results were significantly different from each other ($t(34) = -5.60$, $p < .001$). Subscale (b) *Fantasizing* fluctuated between 8 and 31 points in the experimental group ($M =$

20.28, $SD = 6.65$) and from 8 to 34 points in the control group ($M = 20.56$, $SD = 7.88$); the two groups showed similar responses here ($t(34) = -0.11$, $p = .910$).

Concerning subscale (c) *Identifying* autistic participants responded with 10 to 36 points ($M = 23.47$, $SD = 6.81$), non-autistic participants with 8 to 23 points ($M = 13.61$, $SD = 4.80$); this subscale's outcomes were significantly different ($t(34) = -5.02$, $p < .001$). In subscale (d) *Emotionalizing* the experimental group's results ranged from 10 to 34 points ($M = 20.00$, $SD = 7.75$), the control group's from 9 to 33 ($M = 21.11$, $SD = 5.19$); there was no significant difference ($t(29,68) = 0.51$, $p = .617$). With regards to subscale (e) *Analyzing* experimental probands' answers ranged from 12 to 32 points ($M = 20.00$, $SD = 5.86$) and control probands' from 11 to 24 ($M = 16.28$, $SD = 3.53$); the results were significantly different ($t(34) = -2.31$, $p = .027$).

Factor 1, called the *cognitive factor*, is made up by the subscales *Identifying*, *Analyzing* and *Verbalizing*; the experimental group scored between 51 and 102 points ($M = 73.25$, $SD = 14.37$) and the control group between 33 and 65 points ($M = 48.72$, $SD = 10.40$); these results were significantly different ($t(34) = -5.87$, $p < .001$). Factor 2, called the *affective factor*, is composed of the subscales *Emotionalizing* and *Fantasizing*; autistic participants' scores ranged from 21 to 55 points ($M = 40.28$, $SD = 12.39$) and the control participants' from 25 to 54 ($M = 41.67$, $SD = 9.89$); for this factor the two groups presented similar results ($U = -0.13$, $p = .899$). Since Deborde et al. (2008) only developed a cut-off score for the BVAQ-B 20 item short version, according to which scores of 53 or higher are considered alexithymic, for this study the cut-off score will simply be doubled as the number of items are double as much in the used questionnaire, but the results have to be handled with extra caution. If for the BVAQ-40 item version a in total score of 106 or higher is considered the cut-off value, 11 autistic and 4 non-autistic participants are categorised as alexithymic; the

seven participants identified as alexithymic with the TAS-26 were all among the individuals detected by the BVAQ.

The **Multi-dimensional Empathietest Condensed and Revised** (MET-CORE-2; Dziobek et al., 2008) was 1 of 2 tests to estimate participants' trait empathy levels and ToM-capabilities. The 2 subscales *cognitive empathy* and *emotional empathy* are comprised of answer values and response times; moreover the subscale *emotional empathy* is split into empathy for positive emotions and empathy for negative emotions since Dziobek and colleagues (personal communication, 2014) found that these two concepts are not correlated with each other. Regarding subscale 1 *cognitive empathy* the experimental group's sum of correct answers in total ranged from 19 to 38 ($M = 29.94$, $SD = 5.08$) and the control group's from 21 to 38 ($M = 29.89$, $SD = 4.68$); these did not differ significantly from each other ($t(34) = -0.03$, $p = .973$); also part of subscale 1 the mean response times for all answers in milliseconds varied between 5348.95 and 14419.78 ($M = 8204.44$, $SD = 2950.48$) in the autistic group and between 2872.18 and 16184.63 ($M = 6675.00$, $SD = 3235.24$) in the non-autistic group; here the two groups showed significantly dissimilar results ($U = -2.12$, $p = .034$). When looking at the sum of correct answers and the mean response times of the question "what is this person feeling?" (i.e. *cognitive empathy*) for pictures depicting positive and negative emotions separately, the results appear as depicted in table 3.

Table 3

Cognitive empathy for positive and negative emotions according to the MET

		DESCRIPTIVE			DIFFERENCE BETWEEN EG & CG
MET Cognitive Scale			EG	CG	
Positive	Sum of Correct Answers	<i>M</i>	14.94	15.78	U = -1.10, p = .279
		<i>SD</i>	2.39	3.21	
	Mean Response Time	<i>M</i>	8437.11	6277.77	U = -2.75, p = .005
		<i>SD</i>	2991.92	2884.28	
Negative	Sum of Correct Answers	<i>M</i>	15.00	14.11	U = -1.01, p = .323
		<i>SD</i>	3.05	2.08	
	Mean Response Time	<i>M</i>	7971.77	7072.24	U = -1.55, p = .126
		<i>SD</i>	3025.21	3631.91	

The mean illustrates how many points on average were achieved for the specific task in the stated group, response times are presented in milliseconds. Significant p-values are displayed in bold.

Concerning subscale 2 *emotional empathy* the mean answer value for pictures showing positive emotions among the autistic sample ranged from 1.55 to 7.40 ($M = 4.63$, $SD = 1.59$), among the control sample from 2.20 to 7.80 ($M = 5.76$, $SD = 1.60$); those values were significantly different from each other ($t(34) = 2.13$, $p = .040$). The mean response time when watching and rating positive pictures fluctuated from 2143.65ms to 8931.35ms in the experimental group ($M = 4658.18$, $SD = 1967.49$) and from 1064.50ms to 5622.65ms in the control group ($M = 3041.07$, $SD = 1283.15$); these reactions were significantly different for the two samples ($t(34) = -2.92$, $p = .006$). While the mean answer value for pictures showing negative emotions varied from 1.90 to 8.75 within the autistic group ($M = 5.07$, $SD = 1.85$), it varied from 3.25 to 7.20 within the non-autistic group ($M = 5.52$, $SD = 0.98$); those values were not significantly dissimilar for the two samples ($t(25.85) = 0.92$, $p = .366$). The mean response time when watching and rating negative pictures ranged from 2353.60 to 9952.30 in the experimental group ($M = 4916.04$, $SD = 2328.00$) and from 1769.65 to

8588.50 in the control group ($M = 4178.84$, $SD = 1816.02$); these reactions were not significantly different for the two samples ($U = -1.01$, $p = .311$).

The second test to evaluate empathy and ToM attributes was the **Saarbrücker Persönlichkeitsfragebogen** (SPF; Paulus, 2009) which provides scores for 4 subscales, each with a possible maximum of 20 points: regarding (a) *perspective taking*, the experimental group's results ranged from 7 to 18 points ($M = 12.33$, $SD = 3.12$), the control group's from 9 to 18 points ($M = 14.78$, $SD = 2.34$); these results differed significantly from each other ($U = -2.54$, $p = .011$). In subscale (b) *fantasy*, the autistic participants scored between 6 and 18 points ($M = 11.61$, $SD = 3.43$), the non-autistic participants between 9 and 20 points ($M = 12.89$, $SD = 2.49$); here the two groups showed similar results ($U = -1.05$, $p = .293$). Regarding subscale (c) *empathic concern* the experimental participants had results varying between 22 and 48 points ($M = 37.39$, $SD = 6.72$), the control participants between 31 and 52 points ($M = 41.56$, $SD = 4.90$); the empathic concern scores were barely not significantly different from each other ($U = -1.94$, $p = .053$). Finally in subscale (d) *personal distress* probands with an autism spectrum disorder scored from 8 to 20 points ($M = 13.83$, $SD = 3.75$) and the healthy probands from 6 to 12 points ($M = 8.78$, $SD = 1.93$); in this scale the two groups showed significantly dissimilar results ($t(25.41) = -5.09$, $p < .001$).

With the help of the **Emotion Regulation Questionnaire** (ERQ; Abler & Kessler, 2009) the participants' tendencies to regulate emotions in one way or the other were tested. Concerning subscale *cognitive reappraisal* a total of 42 points is possible; the experimental group's points ranged from 7 to 34 ($M = 21.50$, $SD = 8.27$), the control group's from 16 to 38 ($M = 27.22$, $SD = 6.76$); the two groups did differ significantly in their responses ($t(34) = 2.27$, $p = .029$). On the subject of *expressive suppression* the

autistic group scored between 5 and 26 points ($M = 15.61$, $SD = 6.32$), the control group between 8 to 19 ($M = 12.33$, $SD = 3.34$); these results were not significantly unlike ($t(25.83) = -1.95$, $p = .063$).

Moral dilemma results

Results of the **moral dilemmas** and the mixed model ANOVA were as seen in table 4. Affirmative decisions in non-moral settings were equally distributed for both groups: concerning behaviour ($\chi^2(5, 36) = 6.91$, $p = .228$) and judgment ($\chi^2(4, 36) = 2.94$, $p = .569$), however emotional arousal was not equally distributed in the two samples ($\chi^2(31, 36) = 47.13$, $p = .032$). Behaviour and judgment decisions were moderately to highly correlated for each dilemma type: affirmative decisions in behaviour in non-moral dilemmas and believing in the moral rightfulness of the suggested action correlated moderately ($r_s = .32$, $p = .056$), affirmative behaviour and a positive judgment of the proposed action in impersonal dilemmas highly ($r_s = .54$, $p = .001$) and affirmative behaviour and judgment in personal dilemmas highly, too ($r_s = .68$, $p < .001$).

Table 4

Moral dilemma results

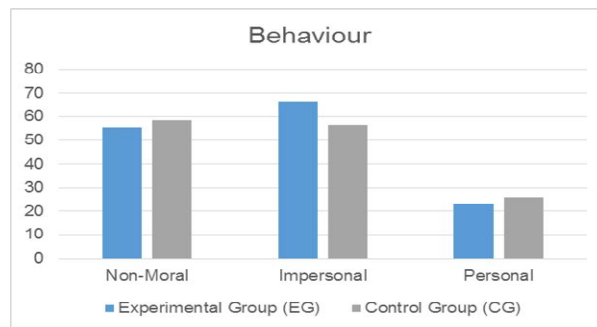
		DESCRIPTIVE DATA			ANOVA RESULTS	
Types of Decision and Type of Dilemma			Experimental Group (EG)	Control Group (CG)	EG vs. CG (Between factor)	Dilemma type (Within factor)
Behaviour	Non-Moral	<i>M</i>	55.56	58.33	$F(1, 36) = 0.08, p = .776$	$F(1.64, 36) = 34.46, p < .001$
		<i>SD</i>	14.00	15.39		
	Impersonal Utilitarian	<i>M</i>	66.67	56.48		
		<i>SD</i>	26.81	31.90		
	Personal Utilitarian	<i>M</i>	23.15	25.93	<i>Partial</i> $\eta^2 = .002$	<i>Partial</i> $\eta^2 = .503$
		<i>SD</i>	26.90	19.15		
Judgment	Non-Moral	<i>M</i>	67.59	75.93	$F(1, 36) = 0.07, p = .795$	$F(2, 36) = 28.55, p < .001$
		<i>SD</i>	17.59	21.56		
	Impersonal Utilitarian	<i>M</i>	60.19	55.56		
		<i>SD</i>	32.91	29.70		
	Personal Utilitarian	<i>M</i>	28.70	29.63	<i>Partial</i> $\eta^2 = .002$	<i>Partial</i> $\eta^2 = .456$
		<i>SD</i>	31.73	22.55		
Emotional Arousal	Non-Moral	<i>M</i>	4.27	1.48	$F(1, 36) = 14.04, p = .001$	$F(2, 36) = 133.49, p < .001$
		<i>SD</i>	2.95	1.82		
	Impersonal Utilitarian	<i>M</i>	15.30	11.56		
		<i>SD</i>	3.45	4.01		
	Personal Utilitarian	<i>M</i>	13.72	9.55	<i>Partial</i> $\eta^2 = .292$	<i>Partial</i> $\eta^2 = .797$
		<i>SD</i>	4.88	4.42		

Mean values are given in percentages, standard deviations in percentage points and emotional arousal values according to the self-reported scores on a continuing scale. Significant p-values are displayed in bold.

While there were no significant dissimilarities between the experimental and the control group regarding behaviour and judgment, the mean values of the two samples did differ slightly in their decisions as seen in figures 1 to 3.

Figure 1

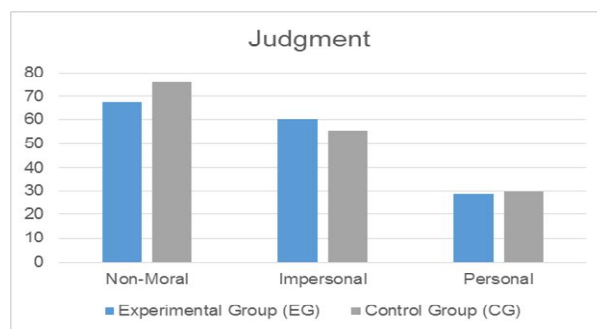
Behaviour decisions in the experimental and the control group



Plotted on the vertical axis is the percentage of agreement to the suggested behaviour (utilitarian actions in the cases of impersonal and personal dilemmas). The bars illustrate the group means per type of dilemma.

Figure 2

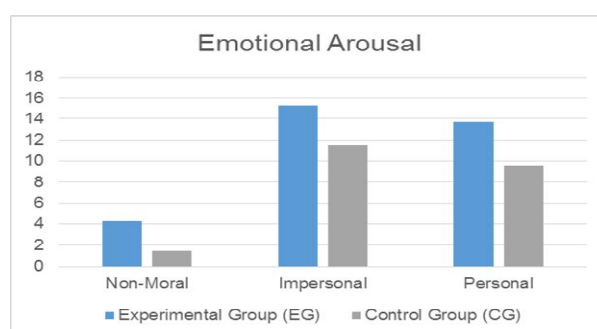
Judgment decisions in the experimental and the control group



Plotted on the vertical axis is the percentage of agreement to the rightfulness of the suggested actions (utilitarian actions in the cases of impersonal and personal dilemmas). The bars illustrate the group means per type of dilemma.

Figure 3

Reported levels of arousal in the experimental and the control group



Plotted on the vertical axis is the level of arousal. The bars illustrate the group means per type of dilemma.

The different types of dilemma were significantly different in all 3 decision categories (behaviour, judgment and emotional arousal). Post-hoc tests clarified that concerning *behaviour* decisions non-moral and impersonal dilemmas did not differ significantly across all participants ($p = 1$, CI [-0.19, 0.11]); decisions in non-moral and personal dilemmas were significantly dissimilar ($p < .001$, CI [0.20, 0.45]) and decisions in impersonal and personal dilemmas were significantly dissimilar, too ($p < .001$, CI [0.27, 0.47]).

Regarding *judgment* decisions the answers of all 36 test persons were not statistically different when comparing non-moral to impersonal dilemmas ($p = .090$, CI [-0.10, 0.30]); when comparing judgment decisions of non-moral and personal dilemmas a significant difference was observed ($p < .001$, CI [0.26, 0.59]) and when looking at judgment decisions of impersonal and personal dilemmas a significant difference was detected as well ($p < .001$, CI [0.17, 0.40]).

Finally *emotional arousal* reports across all partakers brought to light significant differences in all comparisons: non-moral and impersonal dilemmas ($p < .001$, CI [-12.36, -8.75]), non-moral and personal dilemmas ($p < .001$, CI [-10.74, -6.78]) and impersonal and personal dilemmas ($p = .007$, CI [0.42, 3.16]).

Correlations of dilemma results and questionnaire results

Apart from the alexithymia-related (TAS-26 and BVAQ), as well as ToM- and empathy-tests (MET-CORE-2 and SPF) analyses, which were prespecified, all correlations were exploratory (AQ-k, ERQ).

To label a participant as alexithymic, both the results of the **TAS-26** (Kupfer et al., 2001) and the **BVAQ** (Bermond et al., 1999) were taken in consideration; only

if a participant attained was the person categorised as alexithymic. This applied to 7 autists, everybody else was regarded as non-alexithymic. Correlating these labels (0 = non-alexithymic, 1 = alexithymic) with the dilemma results produced no significant correlations and could only be done for the experimental group as seen in table 5.

Table 5

Correlations of dilemma results and alexithymia in the control group

Types of Decision and Type of Dilemma		Behaviour	Judgment	Emotional Arousal
Non-Moral	<i>Spearman r</i> <i>Significance p</i>	$r = 0.07$ $p = .774$	$r = 0.12$ $p = .647$	$r = 0.41$ $p = .094$
Impersonal Utilitarian	<i>Spearman r</i> <i>Significance p</i>	$r = 0.07$ $p = .789$	$r = 0.06$ $p = .820$	$r = 0.01$ $p = .965$
Personal Utilitarian	<i>Spearman r</i> <i>Significance p</i>	$r = 0.035$ $p = .891$	$r = -0.03$ $p = .893$	$r = -0.25$ $p = .312$

However, when looking at the total and subscales' scores of the 2 questionnaires and dilemma results the following associations were found (in favour of legibility only significant correlations are reported at this point): The TAS-26's subscale *Difficulties in identifying emotions* and the self-reported level of emotional arousal for non-moral dilemmas correlated significantly ($r = .55, p = .017$) in the experimental group; as well as *Externally oriented cognitive style* and the self-reported level of emotional arousal for non-moral dilemmas ($r = .60, p = .009$) in the autistic sample and the TAS-26 total score and the self-reported level of emotional arousal for non-moral dilemmas ($r = .60, p = .008$), again this only applied to the experimental group. In the control group the TAS-26 subscale *Externally oriented cognitive style* and the judgments in personal dilemmas correlated positively ($r = .47, p = .047$). Moreover there was a correlation between the BVAQ total score and

affirmative behaviour in non-moral dilemmas in the control group ($r_s = -.72, p = .001$).

The **AQ-k** by Freitag et al. (2007) solely showed significant correlations with the experimental group (in favour of legibility insignificant correlations are not reported at this point): the analysis of the total score revealed associations with the behaviour decision in dilemmas for impersonal dilemmas ($r_s = -.61, p = .008$) and personal dilemmas ($r_s = -.58, p = .012$), as well as with the emotional arousal reported after impersonal dilemmas ($r = .54, p = .020$). The AQ-k's subscale *Social interaction and spontaneity* seems to be connected to behaviour decisions in impersonal ($r_s = -.77, p < .001$) and personal dilemmas ($r_s = -.50, p < .001$), judgments in personal dilemmas ($r_s = -.56, p = .016$) and emotional arousal in impersonal dilemmas ($r_s = .59, p = .010$). The score of the subscale *Phantasy and imagination* correlated with behaviour decisions in personal dilemmas ($r_s = -.47, p = .048$). Finally the correlation analysis displayed a relation between *Communication and reciprocity* and emotional arousal in impersonal dilemmas ($r = -.52, p = .029$).

The **MET-CORE-2**'s (Dziobek et al., 2008) scale *cognitive empathy* correlated in several subscales with dilemma results: the total sum of correct answers (cognitive empathy correct answers) was associated with the experimental group's emotional arousal in impersonal dilemmas ($r = .55, p = .017$) and with the control group's emotional arousal in non-moral dilemmas ($r_s = -.48, p = .045$). For the autistic sample a relation between the sum of correctly identified negative emotions (cognitive empathy negative correct answers) and behaviour in impersonal dilemmas ($r_s = -.48, p = .046$), as well as emotional arousal in impersonal dilemmas ($r = .60, p = .009$) was found. In the control sample correlations of correctly identified positive emotions (cognitive empathy positive correct answers) and judgments in non-moral

scenarios ($r = .53, p = .024$) and emotional arousal in non-moral dilemmas ($r_s = -.50, p = .035$), too, were discovered. The scale *emotional empathy* solely presented significant connections with the experimental group; specifically the average of emotional empathy for items depicting positive feelings (emotional empathy positive average) and behaviour in impersonal dilemmas ($r_s = .49, p = .037$). Again in favour of legibility only significant correlations are reported for this analysis.

Concerning the **SPF** (Paulus, 2009) only the control group featured relations with the dilemma results: *fantasy* showed correlations with behaviour in impersonal dilemmas ($r_s = .77, p < .001$) and personal dilemmas ($r_s = .59, p = .010$); in addition to judgments in personal dilemmas ($r_s = .52, p = .029$). *Empathic concern* and behaviour in personal dilemmas seemed to be connected ($r_s = .49, p = .038$), too. Lastly *personal distress* was associated with behaviour in impersonal dilemmas ($r_s = -.49, p = .038$). Once more in favour of legibility only significant correlations are reported for this analysis.

The 2 subscales of the **ERQ** (Abler & Kessler, 2009) each associated significantly with only 1 of the 2 groups: *cognitive reappraisal* was connected to behaviour decisions in impersonal dilemmas in the autistic group ($r_s = .59, p = .010$) and judgments in personal dilemmas in the same sample ($r_s = .47, p = .047$). Contrary the control group's behaviour decisions in non-moral dilemmas were related to *expressive suppression* ($r_s = -.48, p = .046$), as were judgments in impersonal dilemmas ($r = .52, p = .026$). Again in favour of legibility insignificant correlations are not reported for this analysis.

Correlations of emotional arousal with behaviour and judgment in dilemmas

Table 6 displays the results if the correlation analysis of emotional arousal and dilemma decisions.

Table 6

Correlations of emotional arousal and dilemma decisions

		Emotional Arousal					
		EG			CG		
	Dilemma	Non-Moral	Impersonal Utilitarian	Personal Utilitarian	Non-Moral	Impersonal Utilitarian	Personal Utilitarian
Behaviour	Non-Moral	$r_s = -.34$ $p = .164$	$r_s = -.10$ $p = .682$	$r_s = -.04$ $p = .871$	$r_s = -.01$ $p = .969$	$r_s = .27$ $p = .287$	$r_s = .18$ $p = .466$
	Impersonal Utilitarian	$r_s = .28$ $p = .256$	$r_s = -.71$ $p = .001$	$r_s = -.32$ $p = .197$	$r_s = .17$ $p = .493$	$r_s = .12$ $p = .650$	$r_s = .05$ $p = .834$
	Personal Utilitarian	$r_s = -.24$ $p = .331$	$r_s = -.72$ $p = .001$	$r_s = -.29$ $p = .244$	$r_s = .15$ $p = .541$	$r = -.68$ $p = .002$	$r = .37$ $p = .128$
Judgment	Non-Moral	$r_s = -.43$ $p = .077$	$r_s = .30$ $p = .223$	$r_s = .02$ $p = .932$	$r_s = -.34$ $p = .167$	$r = .23$ $p = .353$	$r = .13$ $p = .611$
	Impersonal Utilitarian	$r_s = -.03$ $p = .901$	$r_s = -.30$ $p = .235$	$r_s = -.14$ $p = .592$	$r_s = -.28$ $p = .254$	$r = .15$ $p = .553$	$r = .13$ $p = .611$
	Personal Utilitarian	$r_s = .17$ $p = .505$	$r_s = -.55$ $p = .018$	$r_s = -.23$ $p = .356$	$r_s = -.06$ $p = .804$	$r = .15$ $p = .561$	$r = .16$ $p = .531$

Significant p values are displayed in bold.

Discussion

The aim of this study was to distinguish the roles of cognitive abilities to put oneself into another person's shoes (ToM) and emotional abilities to feel what somebody else is assumed to feel like (empathy) in moral decision processes (see Moshman, 2011 for a review). The underlying objective was to demonstrate that an

autism spectrum disorder does not inevitably entail empathy deficits which seems to be commonly assumed, but rather that autism is associated with ToM deficits (Bird & Cook, 2013), which cause difficulties in adopting another person's perspective and thus may lead to uncommon behaviour in social situations. However about 50 percent of autists have comorbid alexithymia which is associated with low empathic concern for others (Hill, Berthoz, & Frith, 2004); this may explain the connection of low empathy and autism.

Primary hypothesis

This study's results could not support the claimed differences between autistic and non-autistic participants in moral decisions. There was no observable dissimilarity in moral behaviour or judgment between the two groups according to the statistical analysis with a mixed model ANOVA. Even when looking at autistic trait characteristics, the anticipated connection between those and utilitarian decision making was not existent, on the contrary high autistic characteristics measured with the AQ-k were associated with non-utilitarian response manners. However, when looking at the data in the form of graph one can observe a hint of a slight tendency to higher utilitarian behaviour in impersonal dilemmas in the experimental group. Statistically this was insignificant, though.

This unanticipated outcome has most likely to do with the small sample size, since previous research seems to be in complete agreement with each other that autistic samples produce considerably different results than healthy control samples (e.g. Gleichgerricht et al., 2013). Unfortunately autistic participants turned out to be more difficult to get in touch with and subsequently successfully recruited than assumed. On top of that most autists were recruited with the help of organisations that

work with autism spectrum disorder affected people and offer various therapy programs and social interaction opportunities, thus the in this study acquired sample might not be highly representative for the generally rather withdrawn population of autists. Naturally this has to be emphasised due to the use of incentives (money) to recruit people in the first place.

Furthermore the moral tasks in the form of dilemmas used in this study are no standardised tests, there are no reliability values available and validity is questionable, even though all actions proposed by Christensen and Gomilla (2012) in their meta-study to enhance test quality criteria were taken. Since the types of dilemma did have large effects in all conditions and produced significantly unlike behaviours in this study's participants for each kind of dilemma, their quality seems to be acceptable concerning internal validity, though. Non-moral dilemmas did produce similar results concerning behaviour and judgment in autistic and non-autistic participants as expected. Behaviour and judgment decisions were unlike for personal dilemmas compared to both other types, meaning if the personal distance to a story's character that can be sacrificed by killing him or her is slim, this evokes different conduct and general belief of rightfulness than having notable distance or making a decision without any moral content. Emotional arousal data was dissimilar in all comparisons: non-moral dilemmas prompted much lower statements of arousal than both impersonal and personal dilemmas, which was expected since in non-moral stories no harmful scenario is portrayed and no negative outcome is anticipated. Arousal rates reported after impersonal dilemmas surprisingly were higher than after personal dilemmas, though.

Behaviour and judgment decisions were moderately to strongly connected across all types of dilemmas, implicating that participants agreed to the suggested

action when they thought it was morally acceptable to do so and that they did decided against behaviours they did not deem morally tolerable.

Secondary hypotheses (prespecified)

A connection between low ToM skills and utilitarian judgment as Miller and colleagues reported 2014 could not be supported in this study either. While the experimental group did show deficits in cognitive perspective taking, no significant connections to dilemma decisions were observable. Interestingly the significantly positive correlation between the MET's subscores and dilemma results showed that higher ToM-values went hand in hand with higher emotional arousal in impersonal dilemmas in the autistic group, meaning that being able to adopt other people's perspective leads to more engagement in this study's uncomfortable tasks. Plus higher ToM capabilities seem to be connected to non-utilitarian behaviour in impersonal dilemmas, again for the experimental group.

Literature engaged in the research field of moral, empathy and ToM suggested that less empathy led to more utilitarian responses (Gleichgerricht et al., 2013); despite the fact autistic probands did report less empathic features, they did not show the correspondent answer behaviour in the moral tasks.

Despite the fact that previous research has repeatedly identified a connection between low empathy levels and high utilitarian behaviour occurrence (e.g. Crockett et al., 2010), this study could not replicate those results. Similar to the previous hypothesis concerning ToM, observed empathy levels in the autistic group were significantly lower than in the control group for various measurements and as expected there were autistic participants (39%) that showed strong evidence of an alexithymic comorbidity, but these did not have the assumed impact on moral

decision making in this sample. On the contrary high levels of emotional empathy for positive circumstances of others seemed to be quite strongly connected to utilitarian behaviour in impersonal dilemmas in the autistic group. Regarding the control group a high tendency to identify and empathize with fictional characters of books or movies was also associated with high rates of utilitarian behaviour in impersonal and personal dilemmas, as well as judgments in personal dilemmas. Furthermore the control group was observed to have a strong relationship between high empathic concern and utilitarian behaviour in personal dilemmas.

Certain Alexithymia-characteristics, specifically difficulties in identifying feelings and a strong tendency to focus on external events rather than inner experiences, were highly associated with high self-reported emotional arousal levels in the dilemma tasks regarding the autistic group. A connection between an externally oriented cognitive style and a strong inclination to utilitarian judgments in personal dilemmas could be observed, which is for once a result that is in agreement with the postulated connection of low empathy and high utilitarian decision rate, but surprisingly this was observed only in the healthy control group. Alexithymia as an aggregated category did not correlate significantly with moral decision making at all. A possible explanation may be the sole use of self-reported information concerning alexithymia measurement, however this is true for most if not all studies that measure alexithymia, since the widely used psychological tests in this research field all consist of self-report questionnaires. Moreover the non-alexithymia-specific empathy test MET did not deliver more hypothesis-confirming results. As a means to enhance validity of categorization of alexithymic and non-alexithymic test results of 2 questionnaires (TAS-26 and BVAQ) were used, compared and only if outcomes matched were participants classified as alexithymic.

Secondary hypotheses (exploratory)

A significant difference between the 2 groups and a large effect was evident in relation to emotional arousal, the autistic group on average reporting much higher scores across all types of dilemmas. Moretto and Ládavas (2010, as cited in Gleichgerrcht & Young, 2013), as well as Navarret and colleagues (2011, as cited in Gleichgerrcht & Young, 2013) assumed a relationship between low emotional arousal and utilitarian decisions. This theory is supported by this study's data. Concerning the autistic group the self-reported level of arousal in impersonal dilemmas showed strong negative connections to behaviour in impersonal dilemmas, to behaviour in personal dilemmas, as well as to judgments in personal dilemmas. Regarding the control group a large correlation was observed between the emotional arousal after impersonal dilemmas and behaviour decisions in personal dilemmas, again negative. Thus low emotional excitement after impersonal dilemmas did go hand in hand with utilitarian decisions in personal dilemmas for both groups; low emotional excitement after impersonal dilemmas were also accompanied by utilitarian responses in impersonal dilemmas and judgments in personal dilemmas for a bigger part of the autistic group. Contrary to expectation arousal levels in personal dilemmas did not show corresponding patterns. A possible explanation for these seemingly inconsistent results might be that the question of how aroused the participants were feeling was asked after dilemmas had been read, a behaviour had been set and a judgment had been given; thus the source of the grade of excitement is indistinctive. Interestingly the SPF's subscale *personal distress*, which measures the self-reported trait to get negatively emotionally aroused in social situations, was related to behaviour in impersonal dilemmas in the autistic sample: the higher a person described him- or herself as prone to personal distress, the lower the utilitarian behaviour percentage.

Conclusions

As a result of this study I would like to offer recommendations for future work on this subject. Given the structure of the interview, I would recommend to ask about the participant's emotional arousal before inquiring about behaviour and judgment. This might shed light on the initial condition the participant is in when dealing with the dilemma and its influence on utilitarian behaviour; or the usage of less subjective techniques, for example skin detectors for measuring electrical resistance, might give a more clear and objective answer to the matter in question. Possibly autistic individuals that did take part in this experiment were more nervous than the individuals recruited for the control group, which would offer an explanation for the much higher emotional arousal levels and this in turn might have an anticipated influence on the examined behaviour. Moreover an autism diagnosis implies "the presence of restricted, repetitive patterns of behavior, interests, or activities" (APA, 2013, p. 31), which indicates that autists might not feel comfortable in unknown surroundings and unknown situations; this could also be considered as a reason for high arousal rates and generally untypical answer behaviour, since external validity is susceptible here. Partially these unexpected results might also be explained by emotion regulation tendencies.

The exploratory search for differences in emotion regulation for the autistic versus the non-autistic sample did bring significant results to light: the autistic group showed less cognitive reappraisal tactics, i.e. described less effort to focus on positive thoughts when in an unpleasant situation. This might explain why this study's autistic participants reported such high emotional arousal in the moral tasks. Concerning the experimental group the more participants defined themselves as cognitively reappraising when regulating emotions, the more utilitarian behaviour decisions in

impersonal dilemmas, as well as utilitarian judgments in personal dilemmas were observed. Thus actively looking for positive aspects of a difficult situation made practical decisions that incorporated sacrifices more likely. Regarding the control group the tendency to suppress the expression of emotions was related to utilitarian judgments in impersonal dilemmas.

References

- Abler, B., & Kessler, H. (2009). Emotion Regulation Questionnaire – Eine deutschsprachige Fassung des ERQ von Gross und John. *Diagnostica*, 55(3), 144–152. <http://doi.org/10.1026/0012-1924.55.3.144>
- Bermond, B., & Vorst, H. (1999). The Amsterdam alexithymia scale: Its psychometric values and correlations with other personality traits.
- Bird, G., & Cook, R. (2013). Mixed emotions: the contribution of alexithymia to the emotional symptoms of autism. *Translational Psychiatry*, 3(7), e285. <http://doi.org/10.1038/tp.2013.61>
- Bird, G., Silani, G., Brindley, R., White, S., Frith, U., & Singer, T. (2010). Empathic brain responses in insula are modulated by levels of alexithymia but not autism. *Brain*, 133(5), 1515–1525. <http://doi.org/10.1093/brain/awq060>
- Bzdok, D., Schilbach, L., Vogeley, K., Schneider, K., Laird, A. R., Langner, R., & Eickhoff, S. B. (2012). Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. *Brain Structure and Function*, 217(4), 783–796. <http://doi.org/10.1007/s00429-012-0380-y>
- Center for History and New Media. (n.d.-a). Schnelleinstieg. Retrieved from http://zotero.org/support/quick_start_guide
- Center for History and New Media. (n.d.-b). Schnelleinstieg. Retrieved from http://zotero.org/support/quick_start_guide
- Center for History and New Media. (n.d.-c). Schnelleinstieg. Retrieved from http://zotero.org/support/quick_start_guide
- Center for History and New Media. (n.d.-d). Schnelleinstieg. Retrieved from http://zotero.org/support/quick_start_guide
- Christensen, J. F., & Gomila, a. (2012). Moral dilemmas in cognitive neuroscience of moral decision-making: A principled review. *Neuroscience and Biobehavioral Reviews*, 36(4), 1249–1264. <http://doi.org/10.1016/j.neubiorev.2012.02.008>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (Vol. 2nd). New Jersey: Lawrence Erlbaum. <http://doi.org/10.1234/12345678>
- Crockett, M. J., Clark, L., Hauser, M. D., & Robbins, T. W. (2010). Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proceedings of the National Academy of Sciences of the United States of America*, 107(40), 17433–17438. <http://doi.org/10.1073/pnas.1009396107>

- Deborde, A., Berthoz, S., Wallier, J., Fermanian, J., Falissard, B., Jeammet, P., & Corcos, M. (2008). The Bermond-Vorst Alexithymia Questionnaire cutoff scores: a study in eating-disordered and control subjects. *Psychopathology*, 41(1), 43–49. <http://doi.org/10.1159/000109955>
- Dziobek, I., Rogers, K., Fleck, S., Bahnemann, M., Heekeren, H. R., Wolf, O. T., & Convit, A. (2008). Dissociation of cognitive and emotional empathy in adults with Asperger syndrome using the multifaceted empathy test (MET). *Journal of Autism and Developmental Disorders*, 38, 464–473.
- Freitag, C. M., Retz-Junginger, P., Retz, W., Seitz, C., Palmason, H., Meyer, J., Rösler, M. & Von Gontard, A. (2007). Evaluation der deutschen Version des Autismus-Spektrum-Quotienten (AQ) - Die kurzversion AQ-k. *Zeitschrift Fur Klinische Psychologie Und Psychotherapie*, 36(4), 280–289. <http://doi.org/10.1026/1616-3443.36.4.280>
- Gleichgerrcht, E., Torralva, T., Rattazzi, A., Marengo, V., Roca, M., & Manes, F. (2013). Selective impairment of cognitive empathy for moral judgment in adults with high functioning autism. *Social Cognitive and Affective Neuroscience*, 8(7), 780–788. <http://doi.org/10.1093/scan/nss067>
- Gleichgerrcht, E., & Young, L. (2013). Low Levels of Empathic Concern Predict Utilitarian Moral Judgment. *PLoS ONE*, 8(4). <http://doi.org/10.1371/journal.pone.0060418>
- 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(2), 389–400. <http://doi.org/10.1016/j.neuron.2004.09.027>
- 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 (New York, N.Y.)*, 293(5537), 2105–2108. <http://doi.org/10.1126/science.1062872>
- Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, 85(2), 348–362.
- Haidt, J. (2007). The new synthesis in moral psychology. *Science (New York, N.Y.)*, 316(5827), 998–1002. <http://doi.org/10.1126/science.1137651>
- Hannikainen, I. A., Miller, R. M., & Cushman, F. A. (n.d.). Act versus Impact: Conservatives and liberals exhibit different structural emphases in moral judgment. *Cognition*.
- Heller, K. A., Kratzmeier, H., & Lengfelder, A. (1998). *Matrizen-Test-Manual, Band 1. Ein Handbuch mit deutschen Normen zu den Standard Progressive Matrices von J. C. Raven*. Göttingen: Beltz.

- Hill, E., Berthoz, S., & Frith, U. (2004). Brief Report: Cognitive Processing of Own Emotions in Individuals with Autistic Spectrum Disorder and in Their Relatives: EBSCOhost. *Journal of Autism & Developmental Disorders*, 34(2). Retrieved from <http://web.b.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=d15c4f23-1678-4036-9c4c-3fdf3b312312@sessionmgr115&vid=1&hid=119>
- Kupfer, J., Brosig, B., & Brähler, E. (2001). *TAS-26: Toronto-Alexithymie-Skala-26 (deutsche Version) . Manual*. Göttingen: Hogrefe.
- Lehrl, S. (1995). *Mehrfachwahl-Wortschatz-Intelligenztest (MWT-B)* (3rd ed.). Balingen: Perimed-Spitta.
- Miller, R. M., Hannikainen, I. a., & Cushman, F. a. (2014). Bad Actions or Bad Outcomes? Differentiating Affective Contributions to the Moral Condemnation of Harm. *Emotion (Washington, D.C.)*, 14(3), 1–39. <http://doi.org/10.1037/a0035361>
- Moran, J. M., Young, L. L., Saxe, R., Lee, S. M., O’Young, D., Mavros, P. L., & Gabrieli, J. D. (2011). Impaired theory of mind for moral judgment in high-functioning autism. *Proceedings of the National Academy of Sciences of the United States of America*, 108(7), 2688–2692. <http://doi.org/10.1073/pnas.1011734108>
- Moshman, D. (2013). *Adolescent rationality. Advances in Child Development and Behavior* (Vol. 45). <http://doi.org/10.1016/B978-0-12-397946-9.00007-5>
- Müller, J., Bühner, M., & Ellgring, H. (2004). The assessment of alexithymia: Psychometric properties and validity of the Bermond-Vorst alexithymia questionnaire. *Personality and Individual Differences*, 37(2), 373–391. <http://doi.org/10.1016/j.paid.2003.09.010>
- Olatunji, B. O., & Puncochar, B. D. (2014). Delineating the Influence of Emotion and Reason on Morality and Punishment, 18(3), 186–207.
- Patil, I., & Silani, G. (2014). *Double dissociation between theory of mind and empathy for moral judgments: a study with autism and alexithymia*. Unpublished project proposal, Department of Neuroscience, International School for Advanced Studies, Trieste, Italy.
- Paulus, C. (2009). *Der Saarbrucker Persönlichkeitsfragebogen SPF (IRI) zur Messung von Empathie: Psychometrische Evaluation der deutschen Version des Interpersonal Reactivity Index*. Saarbrücken: Universitäts- und Landesbibliothek.
- Raven, J., Kratzmeier, H., & Horn, R. (1988). *Standard progressive matrices: Sets A, B, C, D and E*. Weinheim: Beltz-Test-Ges.

Vorst, H. C. M., & Bermond, B. (2001). Validity and reliability of the Bermond ± Vorst Alexithymia Questionnaire. *Personality and Individual Differences*.
<http://doi.org/10.1037/0022-3514.85.2.348>

Zöfel, P. (2003). *Statistik für Psychologen: Im Klartext. Psychologie*. München: Pearson Studium.

Appendix

German Abstract

Diese Studie wurde zum Zweck der näheren Erforschung der Autismus-Spektrum-Störung durchgeführt, speziell zur Unterscheidung der Einflussnahme von Theory of Mind-Fähigkeiten (ToM) und Empathie auf moralische Entscheidungen. In vorhergegangenen Studien (beispielsweise Gleichgerrcht et al., 2013) zeigten autistische Personen verstärkte Tendenz zu utilitaristischem Verhalten. Konkret wurden in moralischen Dilemmata jene Antwortalternativen bevorzugt, bei welchen eine Person zum Wohle mehrerer geopfert wird. Explorativ wurden in der vorliegenden Arbeit außerdem emotionale Erregung und emotionale Regulationstendenzen erhoben. Durch die Anwendung moralischer Dilemmata, zweier Alexithymie-Fragebögen (Toronto Alexithymie Skala – 26, Bermond Vorst Alexithymie Skala), eines Papier-Bleistift-Verfahrens zur Erfassung von Empathie und ToM (Saarbrücker Persönlichkeitsfragebogen), sowie eines computerisierten Tests zur Messung derselben Konstrukte (Multi-dimensionaler Empathietest Condensed and Revised), als auch eines Fragebogens zu Emotionsregulierungs-Strategien (Emotion Regulation Questionnaire) wurden 18 autistische und 18 nicht-autistische ProbandInnen getestet. Während bei der Analyse der Häufigkeit utilitaristischer Verhaltensweisen keine signifikanten Unterschiede zwischen den autistischen Testpersonen und der Kontrollgruppe nachgewiesen werden konnten, zeigten sich die erwarteten niedrigeren Empathiewerte in der Versuchsgruppe im Vergleich zur Kontrollgruppe, insbesondere bei der alexithymen autistischen Gruppe.

Wie erwartet wurden auch ToM-Defizite bei den autistischen TeilnehmerInnen festgestellt. Selbstbeschriebene emotionale Erregungswerte fielen in der Versuchsgruppe weitaus höher aus, als in der Kontrollgruppe. Bezüglich Emotionsregulation schien die Strategie der kognitiven Neubeurteilung, welche das Fokussieren auf positive Aspekte einer schwierigen oder unangenehmen Situation beschreibt, unter den autistischen ProbandInnen weniger Anwendung zu finden als unter den Testpersonen der Kontrollgruppe. Zusammenhänge einer niedrigen Tendenz zu kognitiver Neubeurteilung und utilitaristischen Entscheidungen in spezifischen moralischen Prozessen konnten in dieser Studie beobachtet werden. Weiterführende Untersuchungen der Rollen von Empathie und ToM in moralischen Entscheidungsprozessen sind notwendig, da die kleine Stichprobe dieser Studie nicht zur vollständigen Beantwortung der Fragestellung ausreichte.