I see you smile, you must be happy!

Social-emotional gains and usability evaluation of the new training tool E.V.A.:

A pilot study

A thesis submitted in partial fulfillment of the requirements for the degree of
Master of Science in Cognitive Science - Embodied Cognition
At University of Potsdam, Faculty of Human Science

Franziska Blum, Bachelor of Science

Date of submission:
April 1, 2021

First supervisor:

Dr. Anne Weigand, Berlin School of Mind and Brain, Humboldt University of Berlin, Germany

Second supervisor:

Prof. Dr. Ralf Engbert, University of Potsdam, Germany

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-50550
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-505509

“

Perhaps the best term to describe living at the edge of our ability, thriving and flourishing, being challenged but not overwhelmed, is simply 'whelmed', which means taking on the challenges that really speak to you and that emerge from an awareness of your deepest values.

”

– Susan David, author of Emotional Agility

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne Hilfe Dritter und ohne Zuhilfenahme anderer als der angegebenen Quellen und Hilfsmittel angefertigt habe. Die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen sind als solche kenntlich gemacht.

Berlin, Deutschland, 01. April 2021

Franziska Blum

Abstract

Emotions are a complex concept and they are present in our everyday life. Persons on the autism spectrum are said to have difficulties in social interactions, showing deficits in emotion recognition in comparison to neurotypically developed persons. But social-emotional skills are believed to be positively augmented by training. A new adaptive social cognition training tool "E.V.A." is introduced which teaches emotion recognition from face, voice and body language. One cross-sectional and one longitudinal study with adult neurotypical and autistic participants were conducted. The aim of the cross-sectional study was to characterize the two groups and see if differences in their social-emotional skills exist. The longitudinal study, on the other hand, aimed for detecting possible training effects following training with the new training tool. In addition, in both studies usability assessments were conducted to investigate the perceived usability of the new tool for neurotypical as well as autistic participants.
Differences were found between autistic and neurotypical participants in their social-emotional and emotion recognition abilities. Training effects for neurotypical participants in an emotion recognition task were found after two weeks of home training. Similar perceived usability was found for the neurotypical and autistic participants. The current findings suggest that persons with ASC do not have a general deficit in emotion recognition, but are in need for more time to correctly recognize emotions. In addition, findings suggest that training emotion recognition abilities is possible. Further studies are needed to verify if the training effects found for neurotypical participants also manifest in a larger ASC sample.

Zusammenfassung

Emotionen sind ein komplexes Konzept und sie sind Teil unseres alltäglichen Lebens. Personen mit einer Autismus-Spektrum-Störung wird nachgesagt, dass sie Schwierigkeiten mit sozialen Interaktionen und Defizite in der Erkennung von Emotionen haben, im Vergleich zu neurotypischen Menschen. Allerdings glaubt man, dass sich sozio-emotionale Fähigkeiten mittels Training positiv beeinflussen lassen. Ein neues adaptives Trainingstool „E.V.A." wird vorgestellt, welches Emotionserkennung von Gesicht, Stimme und Körpersprache lehrt. Eine Querschnitts- und eine Längsschnittstudie mit erwachsenen neurotypischen und autistischen Teilnehmern wurden durchgeführt. Das Ziel der Querschnittsstudie war die Charakterisierung der zwei Stichproben, sowie die Aufdeckung von möglichen Unterschieden in deren sozio-emotionalen Fähigkeiten. Die Längsschnittstudie, zum anderen, zielte auf die Entdeckung von möglichen Trainingseffekten ab, die auf das Training mit dem neuen Tool folgen. Zusätzlich wurde in beiden Studien die wahrgenommene Benutzerfreundlichkeit von neurotypischen und autistischen Teilnehmern erfasst und untersucht.
Zwischen den neurotypischen und autistischen Teilnehmern wurden Unterschiede in deren sozio-emotionalen Fähigkeiten sowie deren Fähigkeit Emotionen zu erkennen gefunden. Neurotypische Teilnehmer zeigten Trainingseffekte nach einer zwei-wöchigen Nutzung des Trainingstools zu Hause. Die Benutzerfreundlichkeit wurde von den neurotypischen und den autistischen Teilnehmern ähnlich empfunden. Die vorliegenden Ergebnisse deuten darauf hin, dass Autisten kein generelles Defizit in der Erkennung von Emotionen haben, sie aber mehr Zeit dafür benötigen. Zusätzlich weisen die Ergebnisse auf die Möglichkeit des Trainings der Fähigkeit Emotionen zu erkennen hin. Weiterführende Studien sind notwendig um zu verifizieren ob sich die Trainingseffekte auch in einer größeren Stichprobe von Autisten zeigen.

Contents

List of Abbreviations

List of Figures

List of Tables

1 Introduction

1.1 Social Cognition and Emotions

Emotions are a part of our daily life, not only within social interactions such as reaching out to the cashier at the supermarket, your significant other, or your kids, but as well as, and maybe most strikingly, within ourselves. Scientists as Paul Ekman dedicated their research-life to study emotions in humans, and brought rich evidence to the state-of-the-art knowledge about the physiological basis of emotions and emotion recognition, especially through facial expressions. In addition, Ekman aimed to categorize emotions according to their specific characteristics, such as the unconscious contraction of particular sets of muscles, wrinkles arising during the expression of a certain emotion, etc. As such, Ekman writes in his book, that a specific characteristic of a sad facial expression is the eyes declining towards the ground (from German: Ekman, Kuhlmann-Krieg, and Reiss (2010) p. 145). Henceforth, a group of basic emotions emerged for sadness, anger, surprise, fear, disgust, contempt, and joy. But why these? According to Ekman, these seven emotions have one thing in common, namely, each of them has a characteristic and universal facial expression (from German: Ekman et al. (2010) p.82). It is a really difficult challenge to embed emotions into different categories such as basic and complex. Many researchers tried to develop different, possible to test, approaches on categorizing emotions (e.g. Prinz (2004)), but there is no concrete evidence, indicating how emotions can be categorized. Consequently, to this day, basic emotions are thought to be expressed and recognized by all cultures, while complex emotions are more sophisticated to the extent that they are influenced by the context and the culture someone has grown up and is currently living in (Fridenson-Hayo et al. (2016)). That is why the approach by Hepach and colleagues (Hepach et al. (2011)) seems to be more suitable, since they investigated the most frequently used emotions in everyday communication, resulting in a list of forty emotions. The recognition of emotions is part of our social cognition, that is key for social interactions. Cognition, on the other hand, summarizes different processes and mechanisms, such as attention, perception memory, and action planning, which help us humans to make sense of the world (Frith (2008)). When these cognitive processes are underlying social behavior (when interacting within a social group) it is then referred to as social cognition (Frith (2008), Kliemann et al. (2013)).

Furthermore, we have implicit and explicit social cognition. Implicit social cognition refers to the unconscious processing of social cues that includes the automatic and therefore more time and resource-efficient cognitive processes, but there is also not much flexibility. On the other hand, there is explicit social cognition which is more flexible. But it also occupies a lot of our cognitive resources and takes more time, since it includes the conscious and therefore

controlled processing of social cues (Kliemann et al. (2013)). Research suggests differences in social cognition in people with an autism spectrum condition (ASC) (e.g. Isaksson et al. (2019)).

1.2 Autism

The term autism spectrum condition (ASC), or in many resources also autism spectrum disorder (ASD), refers to a neurodevelopmental condition that is associated with deficits or impairments in social communication, restricted interests, and or repetitive behaviors (Rea, LaMotte, and Burrell (2018), Sharma, Gonda, and Tarazi (2018)). According to the International Classification of Diseases 11th revision (ICD-11)[1], the ASC is a neurodevelopmental disorder "characterised by persistent deficits in the ability to initiate and to sustain reciprocal social interaction and social communication, and by a range of restricted, repetitive, and inflexible patterns of behaviour, interests or activities that are clearly atypical or excessive for the individual's age and sociocultural context" (World Health Organization (2020)). Diagnosis for this group of pervasive developmental disorders (please refer to ICD-11) is not easy, due to great symptom and severity variability (Rea et al. (2018)). Prevalence of ASC across the whole world is highly variable, due to diagnostic differences and case definitions (Chiarotti and Venerosi (2020)). But longitudinal analysis of data of a wide time range all over the world suggests that the numbers are increasing. In part, this is due to better and more detailed diagnostic processes (Chiarotti and Venerosi (2020)). An overall "population prevalence of 1% across all ages" (Lai and Baron-Cohen (2015) p. 1013) is reported by Lai and colleagues as well as Happé and Frith (2020). The main three areas where impairments are found are cognitive and linguistic abilities and the capability of adaptation in everyday life (Rea et al. (2018)). The spectrum ranges from mild forms, where individuals learn coping strategies and can live a normal life, to most severe forms where life-long support is required, carried out in professional healthcare institutions in some cases (Alpert (2020)). The variability of impairments and deficits in cognitive and linguistic abilities are extremely high, thus a concrete cognitive profile cannot be defined. For the area of cognitive functioning it is possible to find persons with ASC facing intellectual impairments, but, on the other side, other individuals may be highly intelligent. Some people with ASC get lost in details whereas they forget about the overall picture at the same time, although the focus on details can be advantageous for "fields that value details, such as mathematics, engineering, or music" (Rea et al. (2018) p. 8). Difficulties in processing larger amounts of information, lower cognitive flexibility, and slower learning and processing abilities are present in some people with ASC (Rea et al. (2018)). Also, linguistic variability is high. Some individuals with ASC do not develop speech at all, some face deficits or delays in speech development and some have difficulties in receptive language (Rea et al. (2018)). Partly, as a consequence, many individuals with ASC face problems with everyday activities and their ability to adapt to them or changes in their daily routine (Rea et al. (2018)). An additional factor that makes it difficult to accurately diagnosing ASC is that many individuals also have other comorbid disorders,

[1] `https://icd.who.int/browse10/2019/en#/F84`, last seen 01/19/2021

such as attention deficit hyperactivity disorder (ADHD), depression, anxiety, bipolar disorder, Tourette syndrome or tic disorders, as well as childhood-onset schizophrenia (Sharma et al. (2018)). There are pharmacological and non-pharmacological therapies for ASC which include social-behavioral (SBT) and cognitive-behavioral therapy (CBT). For details about available therapies please refer to Sharma et al. (2018).

1.2.1 Emotion Recognition in Autism

Studies from Haviland and Lelwica (Haviland and Lelwica (1987)) as well as Leppänen (Leppänen, Moulson, Vogel-Farley, and Nelson (2007)) and colleagues showed that the ability to recognize and differentiate between different emotional expressions is present from early life. A study by Rump and colleagues (Rump, Giovannelli, Minshew, and Strauss (2009)) suggests that the ability of emotion recognition develops from early on throughout childhood into adulthood, in neurotypically developed (NT) children and adults, but differently in children and adults with ASC. In their study, ASC adults scored lower on the four emotions afraid, angry, disgust and surprise, compared to the NT adults. Also, Fridenson-Hayo and colleagues (Fridenson-Hayo et al. (2016)) found cross-cultural emotion recognition deficits in ASCs on the three modalities face, voice, and body, as well as the integration of them in context. When looking for publications in emotion recognition, the majority of studies focus on facial emotion recognition (e.g. Rump et al. (2009), Kuusikko et al. (2009), and see reviews Harms, Martin, and Wallace (2010) and Uljarević and Hamilton (2013)). Some studies also investigate other modalities (emotion recognition from prosody e.g. McCann and Peppé (2003), across different modalities e.g. Lindner and Rosén (2006), Fridenson-Hayo et al. (2016)). Contradicting research exists about the emotion recognition abilities of facial expressions in individuals with ASC compared to NTs. Some report no differences (Castelli (2005), Tracy, Robins, Schriber, and Solomon (2011), J. B. Grossman, Klin, Carter, and Volkmar (2000), Rosset et al. (2008), whereas other studies indeed find differences (R. B. Grossman and Tager-Flusberg (2008), Deruelle, Rondan, Gepner, and Tardif (2004); from facial, vocal and body movement stimuli Philip et al. (2010)). To date, no data can reflect a concrete theory, since according to Harms et al. (2010), it is difficult to compare the existing studies to each other. Especially in behavioral research, which is limited in the use of highly sensitive measures such as eye-tracking or neuroimaging studies, the mixed results can arise from demographics characteristics or task demands. "Demographic characteristics of the participants, task demands, and the variables measured [...] all account for the heterogeneity of findings regarding FER [Facial Emotion Recognition] in ASD." (Harms et al. (2010) p. 317). This statement from Harms and colleagues also reflects the heterogeneity within the ASC In addition to the above-mentioned influences, it also remains unclear if it is really helpful for persons with ASC to receive target cues from more than one modality. Existing research suggests that the integration of cues from different modalities (e.g. facial and vocal in a video) makes it more difficult for ASCs to correctly recognize an emotion (Pierce, Glad, and Schreibman (1997)). Another additional point is raised by Klin and colleagues (Klin, Jones, Schultz, Volkmar, and Cohen (2002)) and Dawson, Webb and MacPartland (Dawson, Webb, and McPartland (2005)), who report that individuals with ASC seem to

pay less attention to the face, especially to the eyes (Pelphrey et al. (2002), Baron-Cohen, Wheelwright, and Jolliffe (1997)), which increases the likelihood to miss important cues. This issue should also be addressed in future research and interventions. These findings seem to support that individuals with ASC are impaired in emotion recognition from facial expressions.

1.2.2 Effectiveness of Training Emotion Recognition

In the review by Kuou and Egel (Kouo and Egel (2016)), they conclude that the studies of training emotion recognition skills to children with Autism are promising, although when trying to compare existing research issues arise, making it hard to compare investigations. For example, some studies have participants with high functioning Autism, while others with low functioning Autism. They often differ in verbal abilities and in time exposed to the training material. Also, sample size varies greatly between different studies, and, especially, the materials are not comparable to each other, presenting their training material in varying ways and using different measurements across studies to assess the participants' ability in emotion recognition. Also, Downs and Strand (Downs and Strand (2008)) report significant gains in emotion recognition abilities in children with developmental delays, which suggests that there is a general possibility in training the ability of emotion recognition. Some authors who study emotion recognition even suggest that (e.g. Fridenson-Hayo et al. (2016)), although the development of emotion recognition improves with age in children with ASC and without, the deficits seem to persist; concluding that this "calls for interventions" (Fridenson-Hayo et al. (2016) p. 9). According to Berggren et al. (2018), some promising interventions and related studies for training emotion recognition exist, which report improvements in the ability of emotion recognition. However, the generalizability of these gains to real-life social interactions is still unknown. To account for the generalizability of emotion recognition improvements through emotion recognition training more studies are warranted, with longer duration, larger sample sizes, and follow-up assessment of social abilities. Given the variety of impairments and deficits in ASC (see section 1.2) there will probably never be the one intervention suiting all people with ASC, and there always need to be heterogeneous interventions that address the various needs of persons with ASC (Berggren et al. (2018)).

1.3 Usability

Usability plays a major role in the development of new technology. It addresses intuitiveness and easiness to use. Usability, usefulness, and aesthetics are imperative qualities when it comes to training-technologies, such as online platforms, applications for computers as well as mobile devices (e.g. a language learning app). If technology is perceived as useful and easy to use, the number of users will increase and the motivation to continue to use will be higher. Especially with learning-technologies, if the application is not perceived as useful, why would one want to use it? Existing research shows that the perceived usability of learning technology is positively related to learning outcomes (e.g. Meiselwitz and Sadera (2008)).

To achieve the goal of making an application fun to use, the concept of "gamification" is used. It means that the user has a particular learning goal that needs to be achieved, usually associated with a lot of effort. When using gamification, the effort can be turned into a fun way to achieve the goal, when giving the learner points, rewards, and different levels to achieve while engaging in learning (for more details on gamification please see Brull and Finlayson (2016)). Therefore, usability and how to make technology fun to use should not be neglected, rather focused when it comes to learning technologies.

1.4 Social Cognition Training Tools

Nowadays several training programs for teaching emotion recognition are available, especially for people with ASC. Some focus on teaching emotion recognition to children with ASC, since in general, early intervention is suggested to be of higher effect. Some interventions can be reviewed in Kouo and Egel (2016) and Lee, Lam, Tsang, Yuen, and Ng (2018).

In 2004 Baron-Cohen, Golan, Wheelwright and Hill developed an intervention called MindReading (Baron-Cohen, Golan, Wheelwright, and Hill (2004)). The MindReading intervention includes "412 emotions and mental states, grouped into 24 emotion groups, and six developmental levels (from age 4 into adulthood)" (Golan and Baron-Cohen (2006), p.594). Silent video clips of faces, voice recording as well as written descriptions of different situations which are likely to evoke this emotion are given for each emotion. For the face and voice parts, they made sure that a variety of genders, age range, and ethnicities were given. The intervention compromises three modules. "The learning center" is thought to be used to teach what one needs to know about the different emotions, hence lessons and quizzes are included. "The game zone" where the user learns about emotions while playing a game. The last module is "The emotion library", where all emotion entries are included and the user can explore freely. Two of the authors conducted a study on the intervention's effectiveness in teaching the recognition of complex emotions in adults with an ASC using the MindReading intervention (Golan and Baron-Cohen (2006)). In two experiments they found supporting findings that the use of the MindReading intervention led to significant improvement in the emotion recognition skills of adults with ASC. At the same time, they report that participants had difficulties to generalize the learned concepts and transfer them to real life. This seems to be a problem in general with such interventions, since the trained scenarios are always somehow artificial, and seemingly hard for ASCs to generalize (Rimland (1964)), especially if there is a social component to that learning scenario (Macleod (2016) p.45 ff.). At that time, no such intervention in Germany existed. The Social COgnition Training Tool (SCOTT), which is a game-like self-training tool for social cognition skills, was developed and evaluated at Humboldt University of Berlin (Rosenblau, O'Connell, Heekeren, and Dziobek (2019)). The SCOTT application was running on a desktop computer. SCOTT trains facial emotion recognition, vocal emotion recognition as well as emotion recognition from dynamic scenes (integration of face, voice, and body language from social interactions; Rosenblau et al. (2019)). Through Interviews with study participants from the target user group (Bölte, Golan, Goodwin, and Zwaigenbaum (2010)), some deficits in such training tools

were detected which were present in SCOTT too (Zoerner, Moebert, and Lucke (2017)). The user interface has various details which can be perceived as distracting. In addition, regarding difficulty, the individual's needs are not considered enough, since in SCOTT the difficulty of the tasks is increasing steadily. This was implemented by simply increasing the elements shown in a task or a more complex display of the task, which mainly results in a loss of motivation. Also, in SCOTT the user can to match two videos simply by matching the cutting edges of the presented videos, which shifts the focus away from the task itself. These issues needed to become addressed and the idea for a new social cognition training tool, based on SCOTT, but adaptive to the user, arose. Hence, SCOTT provided the primary basis for the E.V.A. app (please see section 2.1). First of all, the new training tool is desired to be more minimalistic and simpler in design to reduce distraction. A major aspect was the extension of the modules and especially the integration of a library of emotions with detailed descriptions and examples. Some general technical and conceptional aspects should be improved as well, for example, the employed videos should not be able to be matched by their edges. In addition, the new application should run on a tablet computer to increase flexibility for the user.
Another major aspect was the level of difficulty within the new tool. The difficulty should be able to adapt to the users' needs and his/ her skills. This means that the current level of difficulty should be influenced by the user's own learning progress as well as the difficulty of the material. The difficulty of the material can be determined by three aspects: 1) basic emotions are easier to recognize than complex ones; 2) the variability in expression of the different actors; 3) the similarities regarding the emotion's valence and arousal. To put this into an adaptive algorithm the materials need to be revised, evaluated, and rated regarding these three aspects. After that, these ratings can be fed in an adaptive algorithm to determine the appropriate level of difficulty for the current user. This would lead to a difficulty that can rise or lower, depending on the user's personal performance (Zoerner et al. (2017), for details on the implementation of adaptivity and determination of task difficulty please refer to Moebert, Schneider, Zoerner, Tscherejkina, and Lucke (2019)). All these reasons led to the idea of developing a new social cognition training tool.

In this thesis, the new tool called E.V.A. is introduced. Behavioral and usability results in a sample of neurotypically developed (NT) and autistic (ASC) adults (N= 44, NT= 31) trained with E.V.A. will be analyzed and presented. Following this introduction on the different aspects of this work is a detailed description of the two studies from which the data was obtained. This includes a detailed description of the new social cognition training tool E.V.A. In addition, the analysis of selected questionnaires and measurements of both studies is presented.

2 Methods

2.1 E.V.A. App

Figure 2.1: Welcome Screen

The application *Emotionen Vertstehen und Ausdrücken* (E.V.A.) ("understanding and expressing emotions") is a tool developed within the cooperative project *Emotionssensitive Systeme zum Training Sozialer Kognition* (*EMOTISK*[1]) ("emotion sensitive systems to train social cognition") by the Institute of Computational Science at the University of Potsdam and the Institute of Clinical Psychology at the Humboldt University of Berlin. The project has been funded by the German Federal Ministry of Education and Research (BMBF) (Schneider, Dziobek, and Weigand (2019)). This software is based on a previous game-like training software SCOTT (Zoerner et al. (2017)). The current and updated version consists of an adaptive mobile application, used with Android devices, to train the recognition and expression of complex social emotions (social cognition training) by using several different training tasks that combine pictures, video, and audio materials. In total, the app consists of the following modules: Face Puzzles, Voice Puzzles and Film Puzzles. The app also contains an extra feature titled the Library of Emotions (see section 2.1.1). Over fifty trained actors from different cultural backgrounds participated in the recording of the material, thus ensuring social, ethnic, and age diversity. In total, the actors expressed a compendium of forty of the most frequent complex emotions in everyday communication settings, as found

[1] `https://www.technik-zum-menschen-bringen.de/projekte/emotisk`, last seen 01/04/2021

	High arousal	Low arousal
Negative valence	Contemptuous (3.13)	Melancholic (4.2)
	Guilty (3.45)	Compassionate (4.45)
	Disgusted (3.62)	Embarrassed (4.46)
	Wistful (3.63)	Bored (4.55)
	Desperate (3.85)	
	Confused (3.98)	
	Aghast (4.07)	
	Envious (4.23)	
	Aggrieved (4.32)	
	Jealous (4.49)	
	Offended (4.595)	
	Anxious (4.635)	
	Frustrated (4.805)	
	Concerned (4.95)	
	Angry (4.98)	
	Disappointed (4.99)	
	Sad (5)	
	Troubled (5.14)	
	Doubtful (5.15)	
	Cross (5.57)	
Positive valence	Surprised (4.94)	Apologetic (4.38)
	Enthusiastic (4.98)	Pardoning (4.65)
	Lyrical (5.12)	Relieved (5.2)
	In love (5.5)	Proud (5.41)
	Expectant (5.69)	Grateful (5.875)
	Curious (6.49)	Content (5.955)
		Confident (6.04)
		Amused (6.25)
		Joyful (6.48)
		Interested (6.78)

The most frequent word is listed first. The German translations were used in the actual questionnaire.

Figure 2.2: Original list of the 40 most frequent emotions in everyday communication from Hepach et al. (2011)

by Hepach et al. (2011) (see figure 2.2). All materials included in E.V.A. had already been constructed and employed for the previous training software SCOTT. For the E.V.A. app, all these materials were again revised and reevaluated, by the Psychologists of the Humboldt University of Berlin, to ensure high quality of the training materials.

The game is structured in three phases. In the beginning, the early game, the user gets introduced to the topic of the game and how the app functions. During this first phase, the level of difficulty increases gradually. During the mid game phase, new content is continuously accessible as the user plays. The level of difficulty continues to increase gradually. When the user reaches the end game, most of the available content is accessible. In this end game phase, the level of difficulty is dependent on the personal progress of the user. Resulting in a rise or fall of difficulty, depending on the user performance (Zoerner et al. (2017), for more details on the adaptive manner of the E.V.A. app please see Moebert et al. (2019)). The more a user plays or trains with E.V.A. and the better his content-related results (matching and labeling emotions) are, the more tasks, even new modules become available. E.V.A. provides two different scores, the level of emotion recognition ("Emotionserkennungswert"), which is dependent on the user performance and the experience level ("Erfahrungs-Level"), which is only dependent on the time spent playing with E.V.A.

2.1.1 Modules and Additional Features of E.V.A.

2.1.1.1 The Face Puzzle

This module is designed to train the recognition and understanding of emotions which are visible through facial expressions. The explicit version shows a short video of the whole face of a person expressing a random emotion in addition to five emotional labels. The user needs to choose the emotional label that, in his understanding, matches best the expressed emotion (see figure 2.3 top). In the implicit version the user sees a short video of a person expressing a random emotion, but only the upper half of the face (mostly eyes). In addition, there are three videos at the bottom of the screen where only the lower half of the face is shown. The player watches all four videos and tries to choose the right match between the bottom half of that specific emotion and upper half of the face (see figure 2.3 bottom). For doing so, the user has to drag the chosen emotion video into the screen center.

An additional feature of the Face Puzzle is the so called **Emoblitz**. The task begins with a covered picture of an emotion-expressing face and three visible emotion labels on the right. The image starts to uncover successively whereas the user has to select the label describing the depictured emotion as rapidly as possible (see figure 2.4).

2.1.1.2 The Voice Puzzle

Within the Voice Puzzle users train emotional prosody recognition. In the explicit version of the task the user listens to a sentence expressed with a congruent or incongruent prosody

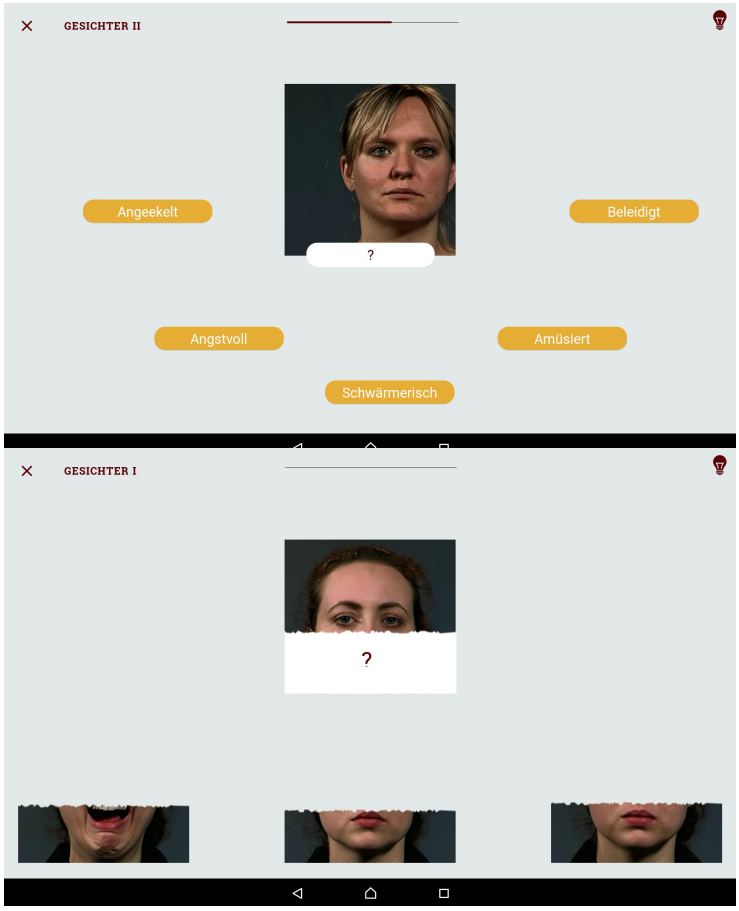


Figure 2.3: Face Puzzle
top: explicit; bottom: implicit

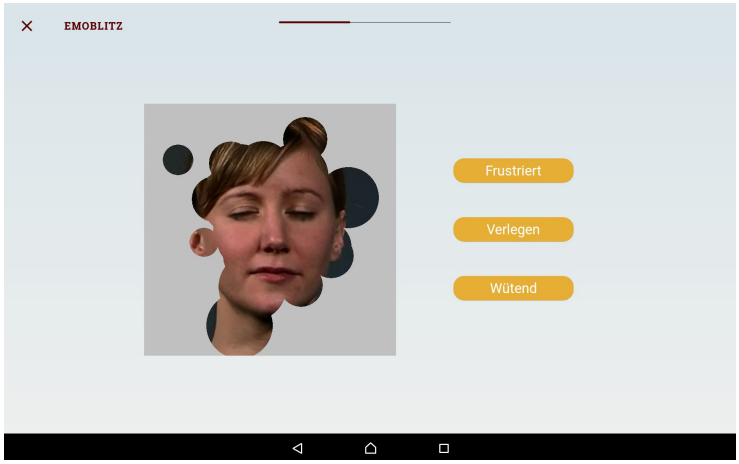


Figure 2.4: "Emoblitz": the face gets visible more and more, the participant has to choose the depicted emotion as rapidly as possible from three emotion labels

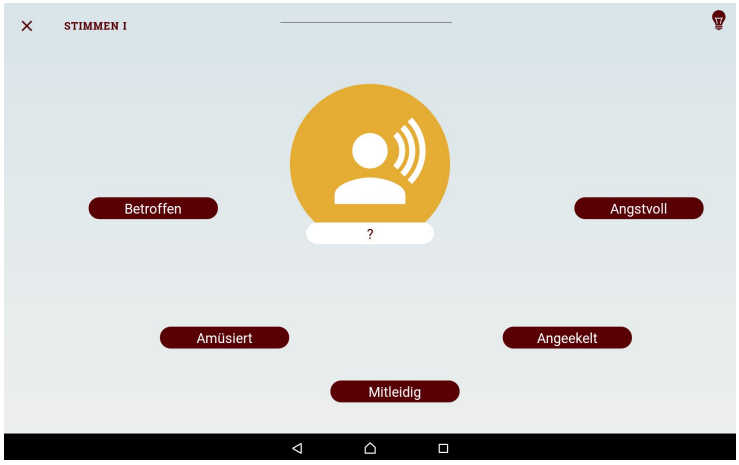


Figure 2.5: Voice Puzzle
explicit

in relation to the content and has to select from five emotion labels which describes the searched emotion (see figure 2.5). An implicit version of this task is planned, but not yet implemented.

2.1.1.3 The Film Puzzle

This module combines facial and vocal expressions of emotions in a short film sequence, implying that also the body posture plays a role in the emotion recognition. The user sees a few given video sequences which need to be ordered correctly. One of them is already placed in the right position, but then the other two (or more) need to be ordered correctly, for the right sequence of the situation, as shown in the video, to be completed (see figure 2.6). The number of video sequences which need to be ordered can differ from task to task.

2.1.1.4 Additional Features of the E.V.A. app

In **the Library of Emotions ("Emotionsschatz")** all forty emotions employed in E.V.A. are listed for detailed description and inspection. A particular emotion can be inspected in detail by clicking the selected emotion card, which will lead the user to a detailed description of this particular emotion (see figure 2.7, p. 13). The top left screen in figure 2.7 shows the overview with a description, synonyms, categorization in regard of valence and arousal level as well as to which group of emotions the selected construct belongs to. There are also pictures of actors expressing this emotion and detailed description of changes in facial expression when experiencing this emotion (see figure 2.7 top right). The bottom left of figure 2.7 shows how voice examples and a description of changes in the voice are given. Examples of situations one could experience this emotion and changes in body posture are given too, which can be seen in figure 2.7 bottom right. For each emotion an individual

Figure 2.6: Film Puzzle: first video sequence is given, the other three video samples have to get ordered correctly

entry exists in the app, displayed by six actors (three female, three male). The included emotions can be filtered in regard of their valence and/or arousal level. Each entry can get bookmarked by the user (see figure 2.8).

2.2 Study Protocol - Testing Procedure

2.2.1 Preliminary Work

Experimental data from the pilot-usability study were obtained from two research studies. In study A (laboratory only, cross-sectional) participants (both NT subjects as well as persons with ASC) attended one single lab session for app-training, behavioral tasks and usability evaluation (about 90 minutes), whereas in study B (longitudinal) participants attended a lab session (T1) for app explanation and a first trial of the two behavioral tasks. Next they received an android-system tablet with the E.V.A. app for active training at home (minimum 3 hours/ week) for a time period of two weeks. Afterwards, the second lab session (T2) took place for usability evaluation and a second trial of the two behavioral tasks (see section 2.4) on emotion recognition. Further descriptions can be found in sections 2.2.2 and 2.2.3.
Both studies were conducted at the Berlin School of Mind and Brain, a Graduate School of the Humboldt University in Berlin. NT participants were recruited via public advertisement on the internet, thus ensuring a wide spectrum of age and educational levels. The NT participants then contacted the research team (one member of the team was specifically trained to attend participants) via electronic mail through the provided contact information on the public notices. The experimenter team then sent back an online link to a demographic questionnaire to conclude if the participant fits inclusion criteria. Inclusion criteria included native German language and a participant age between 18 and 80 years.
Besides filling out demographic data on the questionnaire, participants could select if they wanted to take part in study A (laboratory only) or study B (longitudinal). Brief descriptions

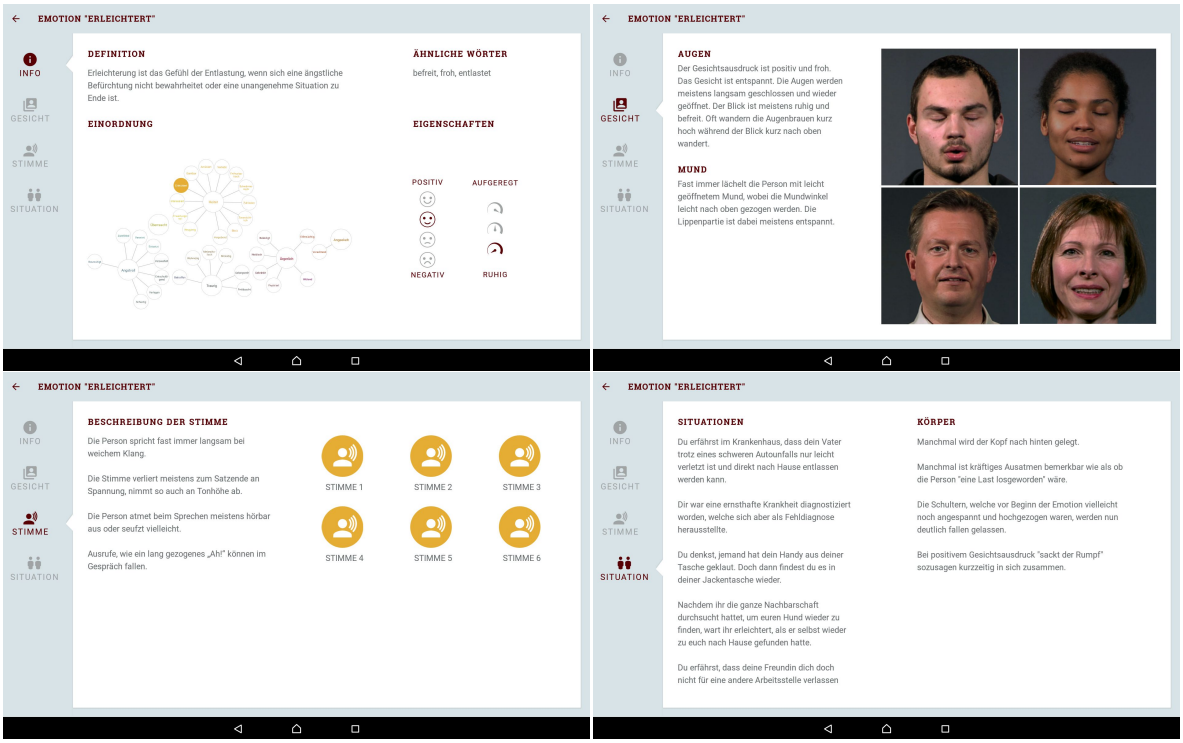


Figure 2.7: Details and examples for a particular emotion, chosen to be inspected in the Library of Emotions (for bigger representation please see figures A.2 and A.3 in Appendix A on pp. 54 ff.)

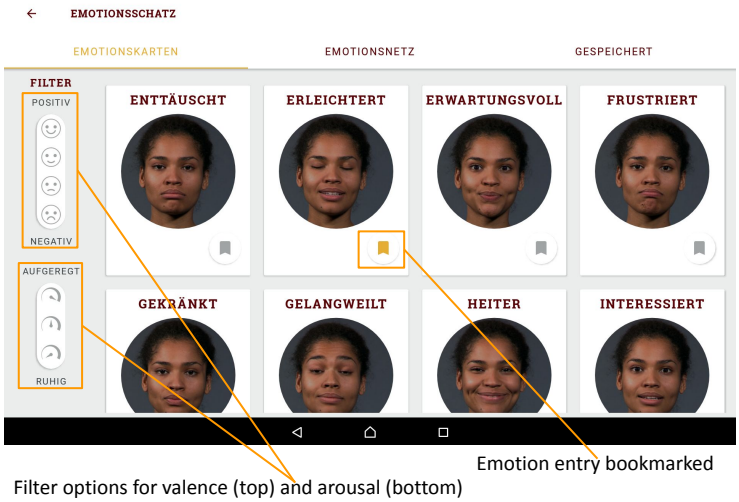


Figure 2.8: Library of Emotions where entries can be filtered in regard of the valence or arousal level as well as bookmarked

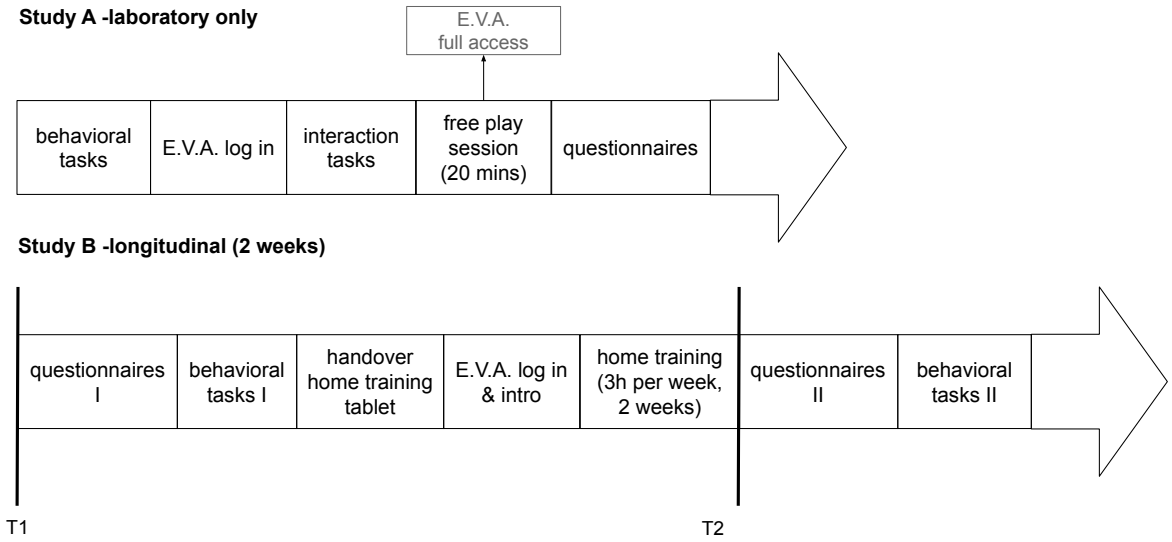


Figure 2.9: Overview on the study procedure, study A laboratory only and study B longitudinal (T1: prior to E.V.A. training, T2: after E.V.A. training)

of both designs were included in the email sent previously to the participants. If participants met inclusion criteria and selected a specific study design (A or B), they were invited to assist on a set date to the research lab. Two days before arrival to the lab, participants received an online link with a series of behavioral measures (see figure 2.9 and section 2.3).
On the other hand, participants with ASC were recruited two ways. One way was through the *Hochschulambulanz für Psychotherapie und Psychodiagnostik* ("university ambulance for psychotherapy and psychodiagnostics") of the Humboldt University of Berlin. The second way was through a research participant list, previously designed in a former study during which informed consent to be contacted again was obtained. Participants with ASC were contacted by the experimenter team individually and invited to take part in one of the two studies (A or B). All other reclusion steps were the same as described previously. All participants gave written informed consent before participation and received economic compensation in exchange. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the ethics committee of the department of Psychology at Humboldt University of Berlin.

2.2.2 Testing: Study A - Laboratory Only

After greeting the participants at the research lab and explaining the study design (app usability study, either cross-sectional or longitudinal), a behavioral evaluation on a lab computer started with one of two possible tasks, Face Puzzle Explicit or Face Morphing Task (see section 2.4). The tests were presented to the participants in a randomized order. After completing both tasks, participants were given an android-system tablet with the installed E.V.A. app. After receiving user and password codes, participants were instructed to log in "as they would do when at home". After log-in, the participant went through the on-

boarding process to receive an overview on the operation of the app. After a couple minutes of free interaction, the participants were asked to find the answer to three questions about the application (interaction tasks), which was used to assess how user friendly and intuitive the application was designed. In the next step, participants were given another tablet with full access to all E.V.A modules, for a 20 minutes free play session. Afterwards, the participants answered some additional questionnaires (see section 2.3) which were provided on the lab computer (figure 2.9). Then participants were compensated economically and dismissed from the study. One day after the lab session, the participants were contacted via email to give voluntary feedback on the lab session and the E.V.A. app.

2.2.3 Testing: Study B - Longitudinal

For study B participants attended the lab twice with 14 days home-training in between (T1-T2). A few days prior to their first attendance the participants received a link via email for some questionnaires they were asked to fill out. At their first attendance (T1) participants were greeted at the lab and the study design was explained to them. They started with one of the behavioral tasks on the lab computer (see section 2.4) in randomized order. After those tasks the participant received the android-system tablet with the installed E.V.A. for home-training (minimum 3 hours/ week, for 2 weeks). Thereafter the participants received another tablet (also with installed E.V.A.) where they were asked to log in (with previously received user and password codes) as they would do when at home. They went through the on-boarding process and reached the home screen. The experimenter gave an overview of the functionality of the E.V.A. app, the three modules, the Library of Emotions and the two scores (for details please see section 2.1). Subsequent to the first lab session the 2 week home-training started where the participants should play at least three hours per week with the E.V.A app. Two days prior to the second lab session (T2) the participants received an email with a link for some questionnaires they were asked to fill out (see section 2.3). The second lab session (T2) commenced similarly to T1 with one of the behavioral tasks on the lab computer (see section 2.4) in randomized order (figure 2.9). Participants were asked for subjective feedback about training with the E.V.A. app and returned the tablet. As in study A, the participants were contacted once again via email, one day after the second lab session, where they could give voluntary feedback on the testing procedure and the E.V.A. app.

2.3 Questionnaires - Assessment

In both studies (A and B) an extensive amount of behavioral data has been collected from the participants to cover a wide range of possible hypotheses to be tested. Since the prior intention of both studies had been the usability aspects of the social cognition training tool E.V.A., measures for social-emotional, as well as usability assessments, were chosen to be performed. Analysis of all these behavioral data would go beyond the scope of this thesis. That is why five specific questionnaires were chosen. The Cognitive and Emotional Empathy Questionnaire (CEEQ), the Self Report Emotional Intelligence Test (SREIT), and the German

version of the Interpersonal Reactivity Index (SPF) for social-emotional assessment. For usability assessment, the theory-based questionnaire for modular evaluation of technologies (meCUE) and the System Usability Scale (SUS), were analyzed.

2.3.1 Social-Emotional Assessment

2.3.1.1 Cognitive and Emotional Empathy Questionnaire (CEEQ)

The cognitive and emotional empathy questionnaire (CEEQ) is considered a multidimensional tool since it assesses both, cognitive and emotional empathy separately. The questionnaire consists of four subscales: *empathic concern (EC), perspective taking (PT), emotion mirroring (MIR)*, and *mental state perception (MSP)*. The two rather traditional empathy subscales, *empathic concern* (assessing one's ability to feel sympathy for others or being worried about them) and *perspective taking* (assessing one's ability to take someone else's point of view), are complemented by two new subscales, the *emotion mirroring* (assessing one's tendency to spontaneously experience emotions observed to be experienced by someone else), and the *mental state perception* (assessing one's ability to recognize and identify someone else's emotions and mental state derived from facial expressions and/ or body language). The *perspective taking* and *mental state perception* subscales are focused on cognition, whereas the *empathic concern*, as well as the *emotion mirroring* subscales, are focused on emotion. The questionnaire consists of a total of 30 self-report items. Each item is answered on a five-point Likert scale, ranging from *not true at all* to *very true* (Köhne (2016)). Since both studies of the current research (see sections 2.2.2 and 2.2.3) intended to evaluate the usability and efficacy of the E.V.A. application, only the *mental state perception* subscale of the CEEQ was used (8 items). For each of the items, an additional question, exclusively for the the present research, was added, elaborating on how the behavior/ ability assessed in the item changed within the last two weeks (in general, this was independent of the 2 weeks home training on a six-point Likert scale from *not at all* to *completely*). For the mental state perception subscale, Köhne reports (Köhne (2016), study 1) a Cronbach's alpha value of 0.84 (for n=98). The translated and used version of the questionnaire can be found in Appendix B, section B.1, p. 57.

2.3.1.2 Self Report Emotional Intelligence Test (SREIT)

The self-report emotional intelligence questionnaire (SREIT) (originally: The Assessing Emotions Scale) is a 33 items questionnaire measuring trait emotional intelligence. "Trait Emotional Intelligence [...] is defined as a constellation of emotional perceptions assessed through [self-report] questionnaires" (Chirumbolo, Picconi, Morelli, and Petrides (2019) p.1). The scale is based on the model of emotional intelligence of Salovey and Mayer (Salovey and Mayer (1990)). Most widely used are four subscales: *perception of emotions, managing own emotions, managing others' emotions*, and *utilization of emotions*. However, for both studies of this project (study A and B, see sections 2.2.2 and 2.2.3) only the subscales *perception of emotions*, with the items 5, 9, 15, 18, 19, 22, 25, 29, 32, 33, and *utilization of emotions*,

with the items 6, 7, 9, 17, 20, 27, were employed since the other two subscales (managing own and others' emotions) are rather irrelevant to assess the efficacy of the E.V.A. app. The original questionnaire is in English, however, the scale was translated to German for the current research (please see Appendix B, section B.2, p. 59 for the used questionnaire, selection of the two subscales perception of emotions and utilization of emotions and translation into German) Schutte, Malouff, and Bhullar (2009). Schutte and his colleagues report a Cronbach's alpha value of 0.9 for n=346 (Schutte et al. (1998)).

2.3.1.3 German version of the Interpersonal Reactivity Index (IRI): Saarbrückener Persönlichkeitsfragebogen (SPF)

The "personality questionnaire from Saarbrücken, Germany" (Saarbrückener Persönlichkeitsfragebogen (SPF)) is based on the Interpersonal Reactivity Index (IRI), first developed by Davis in 1980 (Davis et al. (1980), Davis (1983)). Davis was already taking different dimensions into account, *perspective taking, empathic concern, personal distress*, and *fantasy*. The *perspective taking* scale assesses the ability to see something from someone else's point of view. The *empathic concern* scale is measuring the ability to feel sympathy for someone else or being worried about them, hence it is more "other"-related. Contrary to this is the more "self"-related scale, *personal distress*, to measure feelings of oneself as restlessness and discomfort in certain situations. The *fantasy* scale is focusing on the ability to put yourself into the emotional state of a fictional character e. g. from a book or movie. Previous translations of the IRI have been found not sufficient and lacking reliability (Beven, O'Brien-Malone, and Hall (2004), Ireland (1999)). That is the reason Paulus (Paulus (2009)) created and (re-)evaluated a self translated German version of the IRI in an iterative process of reformulation, reevaluation, and extrication of the negatively presented items, to reduce the mental load for participants. The result of this process is a 35 item questionnaire with the five subscales *empathy (e), perspective taking (p), competence (k), distress (d)*, and *fantasy (f)*, where *competence* is one's ability to "navigate and manage the social and emotional experiences in their lives" (Collie (2020), p. 663). Most items are designed to lead to a strong agreement, whereas some items (4p, 5e, 9f, 15f, 16d, 17e, 19p, 23e, 24d) are designed to lead to strong disagreement by the participants, that is why those items need to be inverted to get a valid score. In the current studies, the version from 2004 was used to collect the data, which can be found in the Appendix B, section B.3, p. 60. Despite the *competence* subscale was eliminated in the more recent versions of the SPF scale, for analysis the more recent version SPF-IRI_V7.0 as of 2019[2] is used. Since both studies (section 2.2.2 and 2.2.3) focused on the training of correctly perceiving, understanding, and interpreting emotions, only the subscales *empathy, perspective taking*, and *competence* were used within this scale. Paulus reports a Cronbach's alpha value of 0.78 (n=339), which is good reliability, especially when taken into account the questionnaire only consists of 10 items (Paulus (2009)).

[2] `http://bildungswissenschaften.uni-saarland.de/personal/paulus/homepage/files/SPF-IRI-_V7.0.pdf`, last seen 04/01/2021

2.3.2 Usability Assessment

2.3.2.1 Theory based questionnaire for modular evaluation of technologies: Theoriebasierter Fragebogen zur modularen Evaluation von Technik (meCue)

Nowadays user experience goes way beyond simple usefulness. A more subjective valuation from a user's perspective including aesthetics, user emotions, and user-system relationship plays an important role in the acceptance of new technologies. A more holistic questionnaire, to allow the standardized assessment of that subjective valuation, was developed by Minge and colleagues (Minge, Riedel, and Thüring (2014), Minge and Riedel (2013), Thüring and Minge (2014)), based on the well-known CUE-Model by Thüring and Mahlke (2007). According to Hassenzahl (2001), the CUE-Model differentiates between task-related and non-task-related qualities. In addition, the model takes user emotions into account. After construction, validation, revision, and re-validation the result was the "theory-based questionnaire for modular evaluation of technologies" (theoriebasierter Fragebogen zur modularen Evaluation von Technik (meCUE)) with a modular structure of four modules. The first module is for assessing *product perception* (task-related and non-task-related) with the subscales *usefulness, usability, visual aesthetics, status*, and *relationship*. The second module, *user emotions*, comprises the subscales *positive* and *negative emotions*. *Consequences* represents the third module with the subscales *intention to use* and *product loyalty*. The last module represents the *total judgment* as a one-item scale. Each subscale is assessed on a seven-point Lickert scale, but for the last module *total judgment*, a score between minus five and plus five is possible. All the above-mentioned subscales, except *status* within the first module (*product perception*) and the complete third module *consequences*, were used to assess data from the participants. With Cronbach's alpha values for all subscales between 0.69 and 0.83 as well as between 0.83 and 0.94 (Minge et al. (2014)) the questionnaire has good reliability. For analysis of the data, the more recent version of the questionnaire, meCUE 2.0, is used. The difference to the previous version is, that the first module is divided into two separate modules with the subscales *usefulness* and *usability* as part of the first module and the subscales *visual aesthetics, status*, and *relationship* forming a new module. The content of the questionnaire is the same, some items appear just in a different order (for further details see Minge (2018)). Questionnaire data was assessed online, the printed version of the used meCUE questionnaire can be found in Appendix B, section B.4, p. 64.

2.3.2.2 Software (System) Usability Scale (SUS)

The software (system) usability scale (SUS) was developed to assess an overall view of subjective usability, which can be used for a wide range of contexts for software systems. It is a 10 item questionnaire and each item is answered on a five-point Likert scale (*strongly disagree* to *strongly agree*). For this method, he asked 20 people from an engineering office[3] (from secretary throughout system programmer) to rate 50 potential items for two different

[3]Usability Engineering Program in Integrated Office Systems Development at Digital Equipment Co Ltd., Reading, UK

software systems. The items with the strongest tendency (to each side of the spectrum), were chosen for the software (system) usability scale. The resulting questionnaire consists of alternating (positive/negative) items to prevent a response bias. Scores of the SUS are obtained by adding the even and odd items and summing up both scores into a total by multiplying by 2.5. The overall SUS score can range from 0 to 100 (Brooke (1996)). Bangor and colleagues (Bangor, Kortum, and Miller (2008)) report a Cronbach's alpha value of 0.91 (for n= 2,324). The in the current studies employed SUS can be found in Appendix B, section B.5, p. 69.

2.4 Emotion Recognition - Assessment

Two behavioral tasks were employed during the lab session. The Face Puzzle task, developed by Kliemann and colleagues (Kliemann et al. (2013)), and a Face Morphing Task like the one used in Schwenck et al. (2012). Both tasks yield to measure participants' ability to correctly recognize and understand emotions. The Face Puzzle Explicit task was selected to get analyzed.

2.4.1 Face Puzzle Task

The Face puzzle, both the explicit and implicit version, are two video-based behavioral tasks, designed to evaluate the ability to recognize emotions from facial expressions developed by Kliemann and colleagues (Kliemann et al. (2013) and Kliemann (2013)). The complete set of stimuli for construction and validation of this task, consisted of 1910 videos, for each emotional state about 45 videos. Both versions are independent web-based applications, accessible with any browser application. Before task onset, a short introduction was provided. Both tasks require intuitive mouse interactions such as hoovering over items and drag and drop. Completion of each version lasts 15 to 20 minutes respectively. There is no time limit on completion, but participants were always prompted to conclude as rapidly and accurately as possible.
Outcome measures yielding from these versions are (1) accuracy (percentage of correct answers), (2) reaction time for choosing the correct label (explicit) or the correct lower face video (implicit) for the target video or emotion, respectively. Another outcome measure is obtained by combining, accuracy and reaction times, the (3) accuracy adjusted response time, which takes into consideration compensatory strategies (e.g. speed-accuracy trade-off, etc., see Sucksmith, Allison, Baron-Cohen, Chakrabarti, and Hoekstra (2013)). For both versions, implicit and explicit, Kliemann and colleagues report a Cronbach's alpha value of 0.81, with n=48 (Kliemann et al. (2013)). For the current work, the reaction times across all responses as well as the correct responses separately and hits and misses were chosen to get analyzed.

Figure 2.10: Explicit task: selection of the appropriate label (screenshot from Kliemann et al. (2013))

2.4.1.1 The Explicit Task

In this version, a target video is presented in the center of the screen exhibiting an actor representing a complex social emotion during a few seconds. The user is required to click on the video for visualization. Below the target video, four emotion labels are presented, requiring the user to select the suitable answer and drag it with the mouse beneath the target video. When a label is chosen and placed under the video the task stops and is completed. The four available label options are not presented randomly. Of the three distractor emotions, two encompass the same valence degree as the target emotion, one of them is similar in arousal while the other shows an inverse arousal level. The last distractor is an emotion of opposite valence (see figure 2.10).

2.4.1.2 The Implicit Task

This version displays a video of the target emotion from the upper part of the actor (region of the eyes) on the center of the screen and four lower-face videos (region of the mouth) on the inferior part of the screen (see figure 2.11). The target video starts playing right away whereas the four options of mouth videos enlarge and commence playing as the participant hoovers with the mouse onto the item. The task is completed when the participant selects one inferior video option and places it via drag and drop right beneath the target video to complete the face.

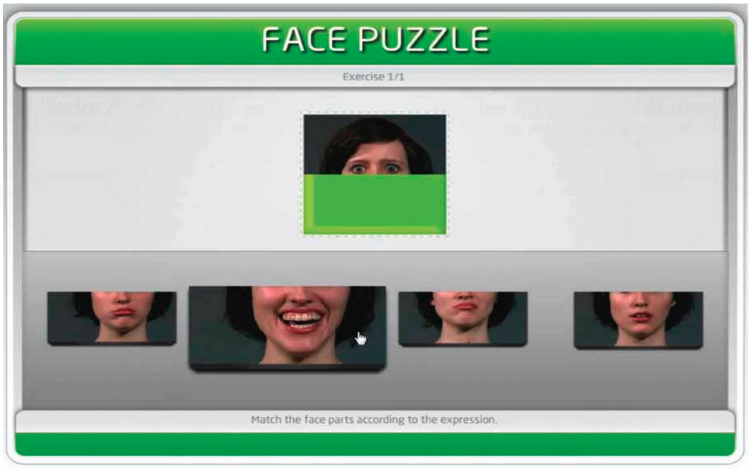


Figure 2.11: Implicit task: selection of the matching lower-face video (screenshot from Kliemann et al. (2013))

3 Analysis

The data sets of all 44 participants for the prior mentioned five questionnaires (see section 2.3) and the Face Puzzle task (see section 2.4) were statistically analyzed. Both samples taken together consists of 24 females and 20 males, with an average age of 34.43 years (SD = 6.76). The sample in study A comprises 31 participants (NT = 21, ASC = 10, mean age M = 34.03, SD = 7.15) whereas the sample in study B comprises a total of 13 participants (NT = 10, ASC = 3, mean age M = 35.38, SD = 6.14).

Study A was aiming to compare the social-emotional skills and emotion recognition abilities between NT and ASC participants. Study B, on the other hand, was aiming to compare pre- (T1) and post-training (T2) differences for NT participants after the 2 week home training session with the E.V.A. app in their social-emotional skills and emotion recognition abilities.

3.1 Social-Emotional Assessment - Analysis

3.1.1 Cognitive and Emotional Empathy Questionnaire (CEEQ)

Data for the *mental state perception* subscale from the CEEQ was collected and analyzed.

Study A

For the CEEQ subscale *mental state perception (MSP)*, a U Mann-Whitney test showed significant between group differences (U = 192.00, p = $<$ .001, figure 3.1) with a large effect size of r_{rb} = 0.83, for the 21 NT participants (M = 29.00, SD = 5.45, min = 18.00, max = 39.00) compared to the 10 ASC participants (M = 17.30, SD = 6.30, min = 10.00, max = 31.00, for boxplots representing mean, minimum and maximum individually per group see figure C.1, p. 71).

Study B

The scores from the *mental state perception (MSP)* subscale of the 10 NT participants, did not reach statistical significance between pre- (T1: M = 26.90, SD = 3.14, min = 22.00, max = 31.00) to post-training (T2: M = 29.30, SD = 5.45, min = 19.00, max = 40.00), as shown by a non-parametrical within-group Wilcoxon test W = 12.50, p = .48, with a medium effect size of r_{rb} = 0.31 (see figure 3.2, for boxplot representation of mean, minimum and maximum for T1 and T2 separately please see figure C.5, p. 73).

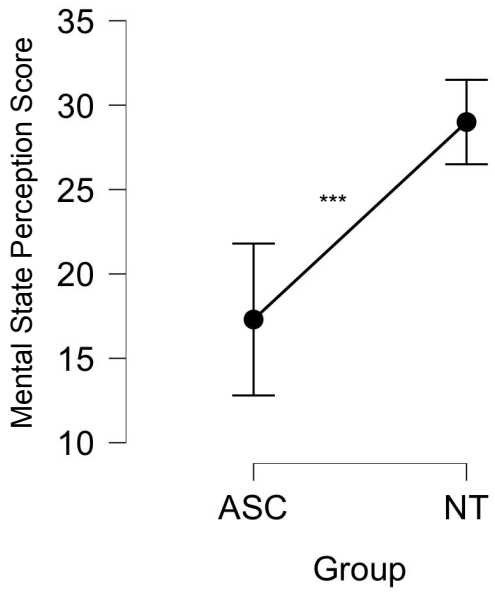


Figure 3.1: Differences in mean *mental state perception* scores (CEEQ), compared ASC and NT

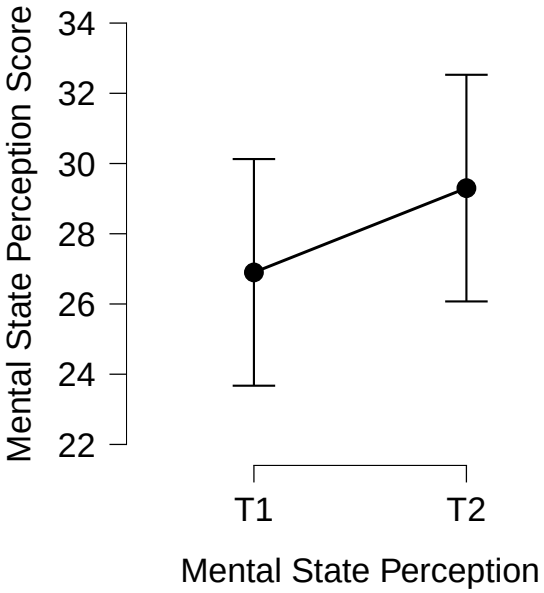


Figure 3.2: Differences in *mental state perception* scores (CEEQ) for NT participants, compared pre- to post-training

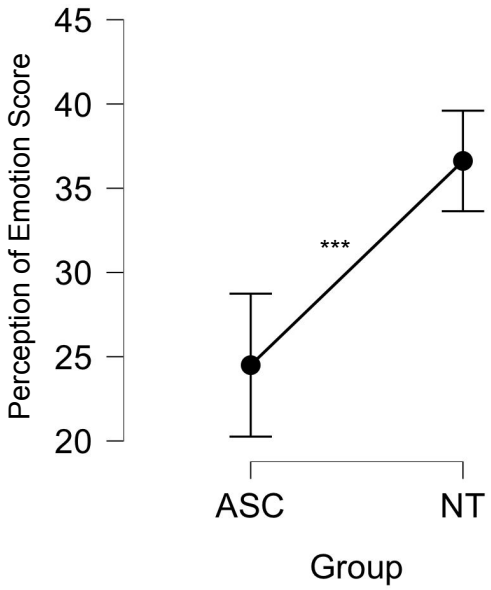


Figure 3.3: Differences in mean *perception of emotion* scores (SREIT) between ASC and NT

3.1.2 Self Report Emotional Intelligence Test (SREIT)

The two subscales *perception of emotions* and *utilization of emotions* from the SREIT questionnaire were collected and analyzed.

Study A

Perception of emotion
The 21 NT participants ($M = 36.62$, $SD = 6.55$, min $= 22.00$, max $= 48.00$) and the 10 ASC participants ($M = 24.50$, $SD = 5.93$, min $= 14.00$, max $= 35.00$) show significant differences (figure 3.3) in the *perception of emotions* subscale, as indicated by a U Mann-Whitney test, $U = 15.00$, $p < .001$, with a large effect size $r_{rb} = 0.86$ (for boxplot representation of mean, minimum and maximum for both groups please see figure C.2a, p. 71).

Utilization of emotion
For the subscale *utilization of emotions* the NT participants ($M = 23.33$, $SD = 3.55$, min $= 18.00$, max $= 28.00$) showed significantly higher scores compared to the ASC participants ($M = 20.00$, $SD = 2.94$, min $= 16.00$, max $= 25.00$) as indicated by a U Mann-Whitney test, $U = 53.50$, $p = .03$, with a medium effect size $r_{rb} = 0.49$ (see figure 3.4 and for boxplot representation of mean, minimum and maximum for both groups please see figure C.2b, p. 71).

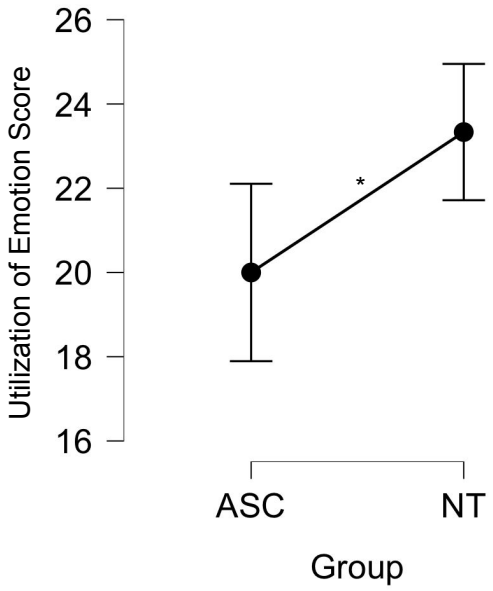


Figure 3.4: Differences in mean *utilization of emotion* scores (SREIT) between ASC and NT

Study B

Perception of emotion
For the 10 NT participants no significant difference (figure 3.5) between pre-training (T1: $M = 34.40$, $SD = 4.48$, min = 28.00, max = 42.00) and post-training (T2: $M = 32.40$, $SD = 4.25$, min = 26.00, max = 37.00) was detected through a Wilcoxon test $W = 39.00$, $p = .26$ with a medium effect size of $r_{rb} = 0.42$ for the *perception of emotions* subscales (for boxplot representation of mean, minimum and maximum for T1 and T2 separately please see figure C.6, p. 73).

Utilization of emotion
A Wilcoxon test for the *utilization of emotions* subscale did not reach statistical significance (figure 3.6) compared pre- (T1: $M = 22.90$, $SD = 3.00$, min = 20.00, max = 29.00) to post-training (T2: $M = 20.70$, $SD = 3.27$, min = 15.00, max = 26.00), with $W = 30.00$, $p = .10$ and a large effect size of $r_{rb} = 0.67$ (for boxplot representation of mean, minimum and maximum for T1 and T2 separately please see figure C.7), p. 74.

3.1.3 German version of the Interpersonal Reactivity Index (SPF)

From the SPF questionnaire the subscales *empathy, perspective taking* and *competence* were collected and analyzed.

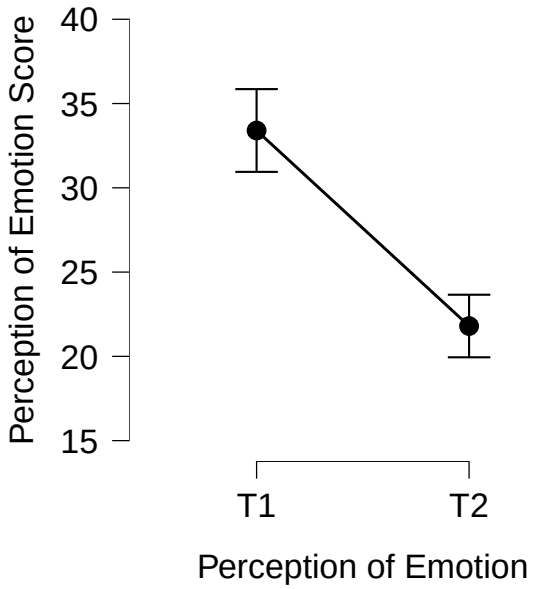


Figure 3.5: Differences between pre- and post-training for the *perception of emotion* subscale (SREIT)

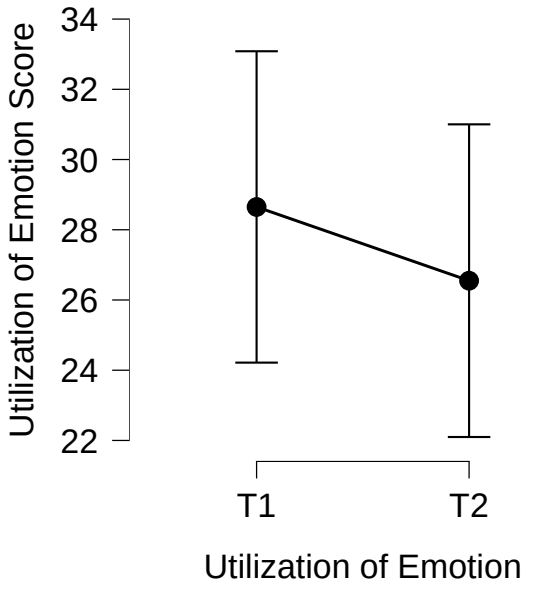


Figure 3.6: Differences between pre- and post-training for the *utilization of emotion* subscale (SREIT)

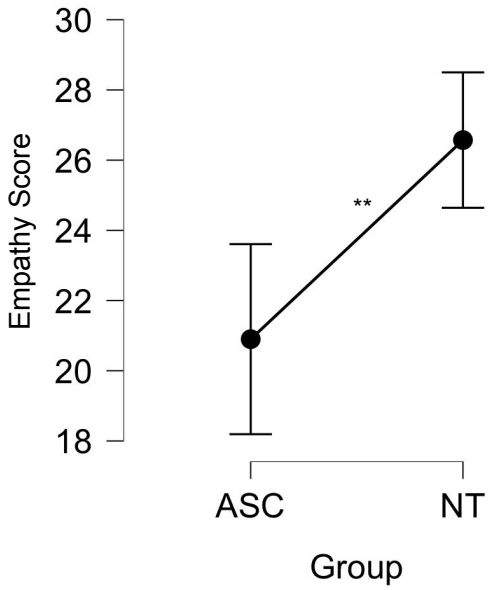


Figure 3.7: Differences in mean *empathy* scores (SPF) between ASC and NT

Study A

Empathy

Significant differences were detected with a U Mann-Whitney test with $U = 31.50$, $p = .002$ and a large effect size of $r_{rb} = 0.70$ (see figure 3.7), comparing the 21 NT participants ($M = 26.57$, $SD = 4.24$, min = 19.00, max = 33.00) to the 10 ASC participants ($M = 20.90$, $SD = 3.78$, min = 16.00, max = 29.00; for boxplot representation of mean, minimum and maximum for both groups please see figure C.3a, p. 72).

Perspective taking

The subscale *perspective taking* showed a significant difference with a Mann-Whitney test, $U = 43.50$, $p = .01$, and a large effect size of $r_{rb} = 0.59$ (see figure 3.8) between the 21 NT participants ($M = 25.48$, $SD = 4.18$, min = 18.00, max = 33.00) and the 10 ASC participants ($M = 20.50$, $SD = 4.06$, min = 12.00, max = 26.00; for boxplot representation of mean, minimum and maximum for both groups please see figure C.3b, p. 72).

Competence

For the *competence* subscale, no significant differences (figure 3.9) between the 21 NT participants ($M = 23.28$, $SD = 2.84$, min = 19.00, max = 29.00) and the 10 ASC participants ($M = 22.80$, $SD = 3.99$, min = 16.00, max = 27.00), analyzed through a U Mann-Whitney test resulting in $U = 102.50$, $p = .93$ and a small effect size of $r_{rb} = 0.02$, were found (for boxplot representation of mean, minimum and maximum for both groups please see figure C.3c, p. 72).

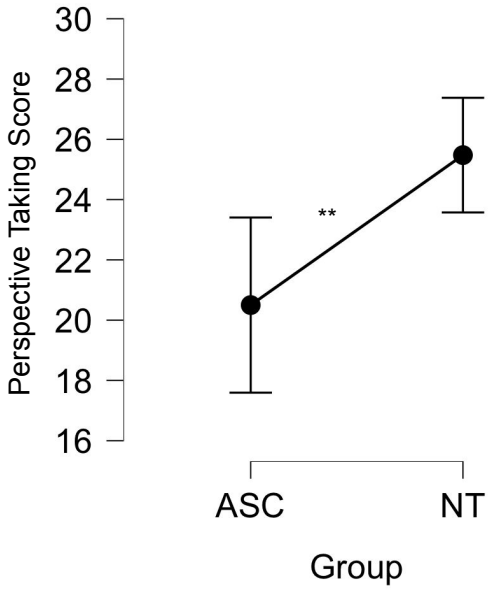


Figure 3.8: Differences in mean *perspective taking* scores (SPF) between ASC andNT

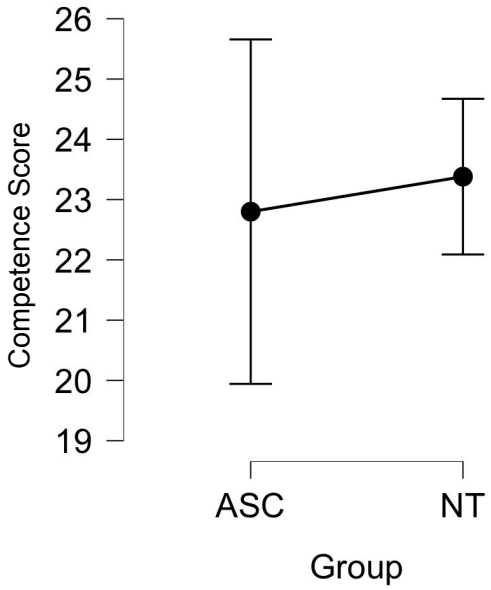


Figure 3.9: Differences in mean *competence* scores (SPF) between ASC and NT

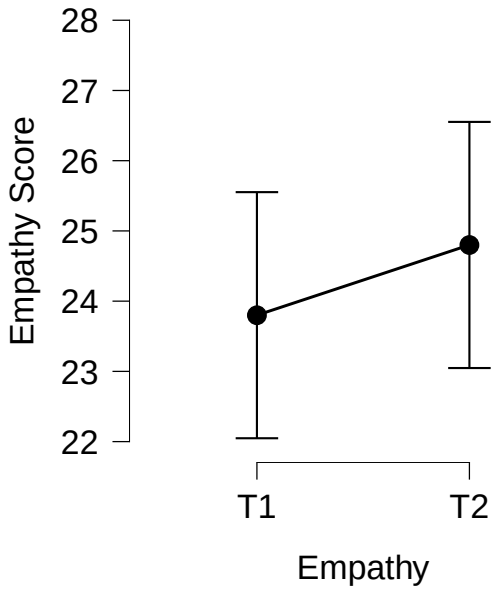


Figure 3.10: Differences between pre- and post-training for *empathy* (SPF)

Study B

Empathy
The 10 NT participants showed no significant differences comparing pre- (T1: $M = 23.80$, $SD = 3.77$, min = 19.00, max = 31.00) to post-training (T2: $M = 24.80$, $SD = 4.66$, min = 17.00, max = 30.00), with $W = 18.00$, $p = .35$ and a medium effect size of $r_{rb} = 0.35$ (see figure 3.10, for boxplot representation of mean, minimum and maximum for T1 and T2 separately please see figure C.9, p. 75).

Perspective taking
Comparing pre- (T1: $M = 23.30$, $SD = 3.47$, min = 16.00, max = 28.00) to post-training (T2: $M = 23.80$, $SD = 4.08$, min = 17.00, max = 30.00) did not reveal any difference for the 10 NT participants, with $W = 22.00$, $p = 1.0$ and a small effect size $r_{rb} = 0.02$ (see figure 3.11, for boxplot representation of mean, minimum and maximum for T1 and T2 separately please see figure C.10, p. 75).

Competence
For *competence* was no difference detected for the 10 NT participants comparing pre- (T1: $M = 22.40$, $SD = 2.37$, min = 17.00, max = 26.00) to post-training (T2: $M = 22.70$, $SD = 2.94$, min = 17.00, max = 26.00), with $W = 16.00$, $p = .83$ and a small effect size of $r_{rb} = 0.11$ (see figure 3.12, for boxplot representation of mean, minimum and maximum for T1 and T2 separately please see figure C.8, p. 74).

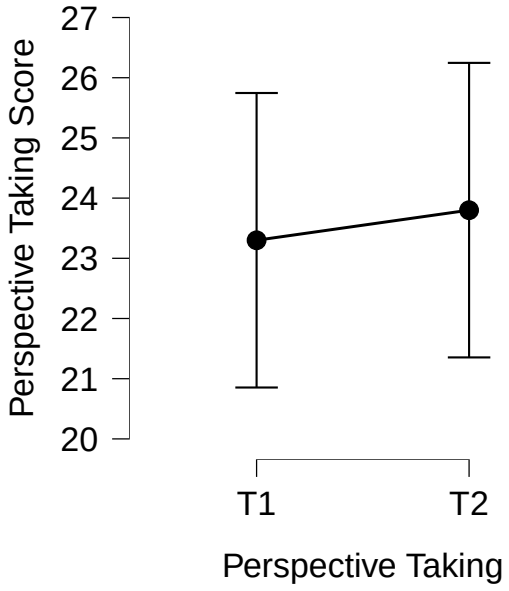


Figure 3.11: Differences between pre- and post-training for *perspective taking* (SPF)

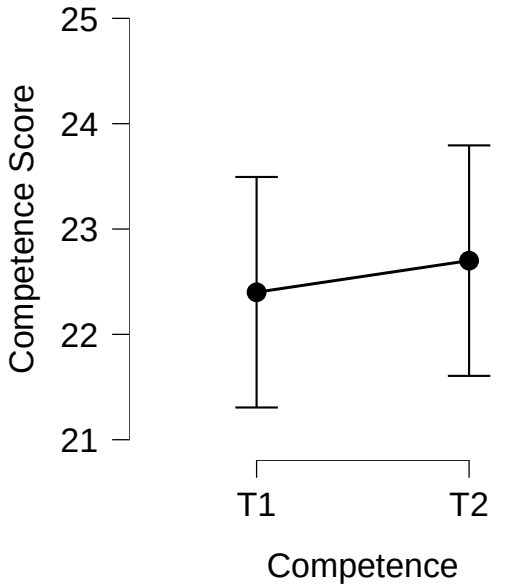


Figure 3.12: Differences between pre- and post-training for *competence* (SPF)

3.2 Emotion Recognition - Analysis

3.2.1 Face Puzzle Task

The Face Puzzle task consists of 25 trials, where a maximum of 25 correct responses can be achieved. For all 31 participants in study A and all 13 participants in study B, the mean reaction time for all trials taken together and the reaction times for the trials with correct responses, each in milliseconds (ms), as well as the total number of correct responses (hits) and the number of misses were recorded.

Study A

Number of correct responses (hits)
Between the two groups, ASC participants ($M = 18.20$, $SD = 4.36$, min = 7.00, max = 22.00) and the NT participants ($M = 18.67$, $SD = 2.67$, min = 13.00, max = 22.00) were no significant differences detected, as indicated by a U Mann-Whitney test ($U = 108.00$, $p = .92$, with a small effect size of $r_{rb} = 0.03$), for number of correct responses.

Mean reaction time for correct responses
For the mean reaction time for correct responses statistical significance was detected by a U Mann-Whitney test, $U = 174.00$, $p = .003$ with a large effect size of $r_{rb} = 0.66$ (figure 3.13), for the 21 NT participants ($M = 9660.36$, $SD = 2646.33$, min = 5981.69, max = 17571.15) compared to the 10 ASC participants ($M = 16459.65$, $SD = 7725.25$, min = 6677.86, max = 30046.95, for boxplot representation of mean, minimum and maximum for both groups please see figure C.4a, p. 72).

Reaction time mean
The 21 NT participants ($M = 10678.86$, $SD = 3064.33$, min = 6506.96, max = 17827.24) showed a significant difference in the mean reaction time across all responses (figure 3.14), compared to the 10 ASC participants ($M = 17070.58$, $SD = 7586.09$, min = 8005.00, max = 31473.88), as shown by a U Mann-Whitney test, $U = 168.00$, $p = .007$ with a large effect size of $r_{rb} = 0.60$ (for boxplot representation of mean, minimum and maximum for both groups please see figure C.4b, p. 72).

Number of misses
For the misses no significant differences were detected between the ASC participants ($M = 6.80$, $SD = 4.37$, min = 3.00, max = 18.00) and the NT participants ($M = 6.33$, $SD = 2.67$, min = 3.00, max = 12.00) with a U Mann-Whitney test ($U = 102.00$, $p = .92$, with a small effect size $r_{rb} = 0.03$).

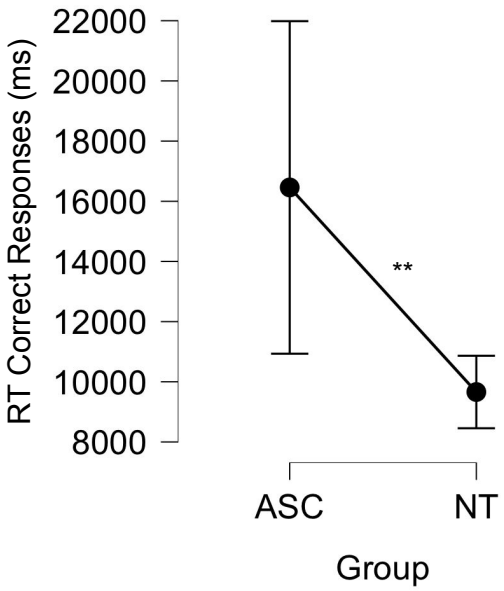


Figure 3.13: Differences in mean reaction times (RT) for correct responses (in ms) for ASC and NT participants

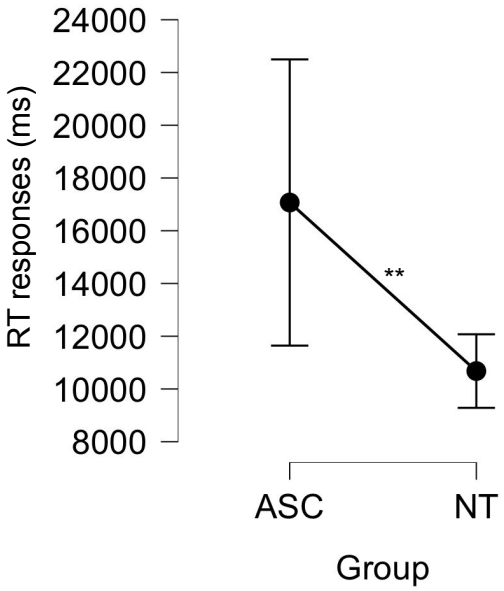


Figure 3.14: Differences in mean reaction times (RT in ms) for ASC and NT participants (across all responses)

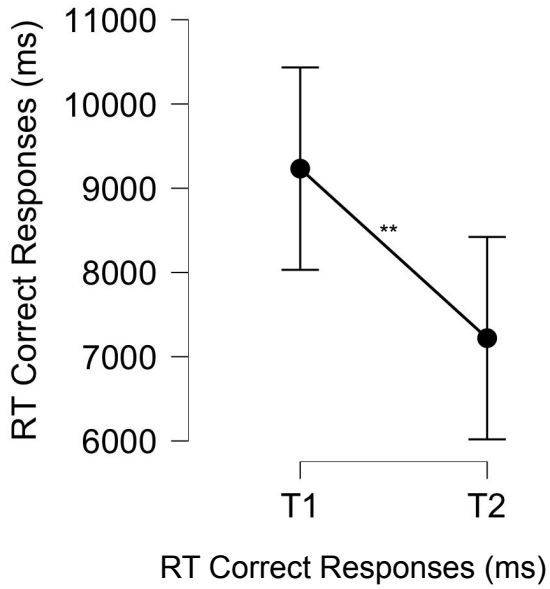


Figure 3.15: Differences in mean reaction time (RT) for correct responses (in ms) compared pre- to post-training

Study B

Number of correct responses (hits)

For the number of correct responses, the mean of the 10 NT participants, was almost the same for pre-training (T1: $M = 18.10$, $SD = 3.48$, min = 14.00, max = 24.00) and post-training (T2: $M = 18.20$, $SD = 3.12$, min = 13.00, max = 24.00) and therefore did not reach statistical significance, as indicated by a Wilcoxon test ($W = 21.50$, $p = .95$, with a small effect size $r_{rb} = 0.04$).

Reaction time for correct responses

In mean reaction time for correct responses significant differences (figure 3.15) were detected with $W = 55.00$ and $p = .002$ and a large effect size of $r_{rb} = 1.0$ compared pre-training (T1: $M = 9232.59$, $SD = 2589.62$, min = 6457.43, max = 15398.94) to post-training (T2: $M = 7220.25$, $SD = 1205.86$, min = 5787.77, max = 9632.71, for boxplot representation of mean, minimum and maximum for T1 and T2 separately please see figure C.11, p. 76).

Reaction time mean

In the mean reaction time of the 10 NT participants for pre-training (T1: $M = 10054.12$, $SD = 2843.27$, min = 7118.96, max = 15667.08) compared to post-training (T2: $M = 7596.84$, $SD = 1341.77$, min = 5897.28, max = 10309.64) revealed a significant difference with $W = 53.00$, $p = .006$ with a large effect size of $r_{rb} = 0.93$ (figure 3.16, for boxplot representation of mean, minimum and maximum for T1 and T2 separately please see figure C.12, p. 76).

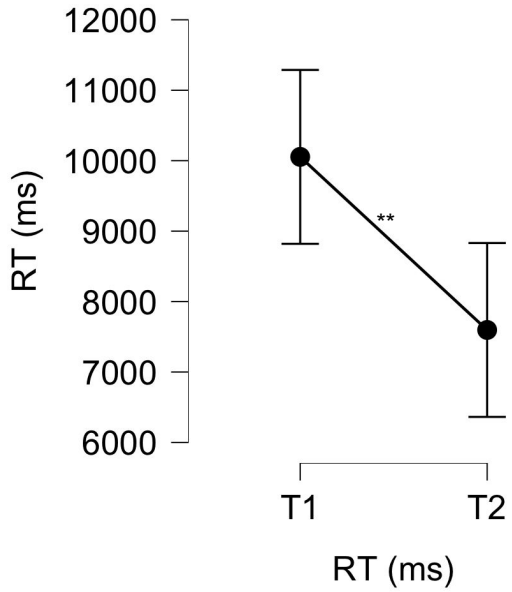


Figure 3.16: Differences in mean reaction times (RT in ms) between pre- and post-training

Number of misses
The mean number of misses across the 10 NT participants was almost the same for pre-training (T1: $M = 6.90$, $SD = 3.48$, min $= 1.00$, max $= 11.00$) and post-training (T2: $M = 6.80$, $SD = 3.12$, min $= 1.00$, max $= 12.00$) and therefore a Wilcoxon test did not reveal any significant differences ($W = 23.50$, $p = .95$, with a small effect size $r_{rb} = 0.04$).

3.3 Usability Assessment - Analysis

A special area of interest in both studies was the overall perceived usability of the E.V.A. app, thus several usability questionnaires were included in the original studies. This master thesis focused on two selected questionnaires, in order to obtain an overview on the overall perceived usability of the application E.V.A. The usability measures were assessed only once, in study A after the use of E.V.A., and in study B after the two week home-training session with E.V.A. For this reason, participants from study A and from study B were taken together to display the perceived usability by each group (ASC and NT) and to check for differences between the two groups. If no differences were detected, all participants of both groups were taken together to get an overview on the perceived usability by all participants.

	Usability		Usefulness		Aesthetics		Relationship	
	ASC	NT	ASC	NT	ASC	NT	ASC	NT
Valid	13	31	13	31	13	31	13	31
Missing	0	0	0	0	0	0	0	0
Mean	5.077	5.161	3.564	3.322	4.232	3.849	0.409	0.473
Std. Deviation	1.571	1.004	1.273	1.234	1.059	1.021	0.513	0.648
Minimum	1.330	2.000	1.330	1.330	2.000	1.330	0.000	0.000
Maximum	6.000	6.000	5.670	6.000	5.670	5.330	1.670	2.000

Table 3.1: Mean, *SD*, minimum and maximum for ASC and NT for the subscales *usability, usefulness, aesthetics* and *relationship*

	Positive Emotions		Negative Emotions		Total Judgment	
	ASC	NT	ASC	NT	ASC	NT
Valid	13	31	13	31	13	31
Missing	0	0	0	0	0	0
Mean	1.757	2.167	1.897	1.565	2.654	2.274
Std. Deviation	1.575	1.200	1.361	1.277	2.154	1.999
Minimum	0.000	0.000	0.330	0.000	-3.500	-4.000
Maximum	4.670	4.170	4.830	4.170	5.000	5.000

Table 3.2: Mean, *SD*, minimum and maximum for ASC and NT for the subscales *positive emotions, negative emotions* and *total judgment*

3.3.1 Theory based questionnaire for modular evaluation of technologies (meCUE)

3.3.1.1 Differences between ASC and NT

Between the 31 NT participants and the 13 ASC participants no significant differences were detected for none of the seven subscales (for *Mean, SD*, Min and Max please see tables 3.1 and 3.2 as well as figure 3.17, p. 38). Each subscale was assessed on a seven-point Lickert scale. A U Mann-Whitney test for the subscale *usability* resulted in $U = 230.50$, $p = .45$ and a small effect size of $r_{rb} = 0.14$. For the subscale *usefulness* a U Mann-Whitney showed no significant differences between the two groups with $U = 222.00$, $p = .61$, and small effect size of $r_{rb} = 0.10$. Also for the *aesthetics* subscale no significant differences were detected with $U = 245.50$, $p = .26$ and a small effect size of $r_{rb} = 0.22$. A U Mann-Whitney test for the *relationship* subscale resulted in $U = 209.00$, $p = .85$ and a mall effect size $r_{rb} = 0.04$. For the *positive emotions* subscale a $U = 162.00$ with $p = .32$ and a small effect size of $r_{rb} = 0.20$ was calculated. The U Mann-Whitney test for the *negative emotions* resulted in $U = 225.00$ with $p = .55$ and a small effect size of $r_{rb} = 0.12$. Last, the *total judgment*, assessed on a scale from -5 to +5, did also not reveal any differences as indicated by a U Mann-Whitney test with $U = 222.00$, $p = .60$ and a small effect size of $r_{rb} = 0.10$.

3.3.1.2 Overall

Since there were no evident difference detected in the scores for the subscales, between ASC and NT participants the overall scores across all 44 participants were calculated (figure 3.17, p. 38 and for frequencies of each response for each subscale please see figures in Appendix C, section C.3, 77). The score of the *usability* scale resulted in a mean of $M = 5.14$ points ($SD = 1.18$, min = 1.33, max = 6.00). *Usefulness* is scored with a mean of $M = 3.39$ points ($SD = 1.24$, min = 1.33, max = 6.00). The score of the *aesthetics* scale encompasses a mean of $M = 3.96$ points ($SD = 1.03$, min = 1.33, max = 5.67). The lowest scores were assessed in the *relationship* subscale with a mean of $M = 0.45$ points ($SD = 0.61$, min = 0.00, max = 2.00). Positive emotions related to E.V.A. reached a mean score of 2.05 points ($SD = 1.32$, min = 0.00, max = 4.67), whereas negative emotions reached a mean score of 1.66 points ($SD = 1.30$, min = 0.00, max = 4.83). The *total judgment*, assessed on a scale from -5 to +5 where 0 displayed neutrality, reached a mean score of 2.39 points ($SD = 2.03$, min = -4.00, max = 5.00).

3.3.2 Software (System) Usability Scale (SUS)

No difference in the scores of the SUS were detected comparing all 31 NT participants ($M = 80.65$, $SD = 12.80$, min = 55.00, max = 100.00) to all 13 ASC participants ($M = 77.31$, $SD = 19.64$, min = 20.00, max = 97.50), with a U Mann-Whitney test resulting in $U = 187.00$, $p = .72$ and a small effect size of $r_{rb} = 0.10$, although the ASC participants minimum is clearly lower than the NT participants minimum (figure 3.18).

The overall usability was perceived rather high across all participants with a mean score of 79.66 points as well as a *SD* of 14.98 points, minimum was 20.00 points and maximum was 100.00 points, which is also the highest possible score for the SUS (figure 3.18).

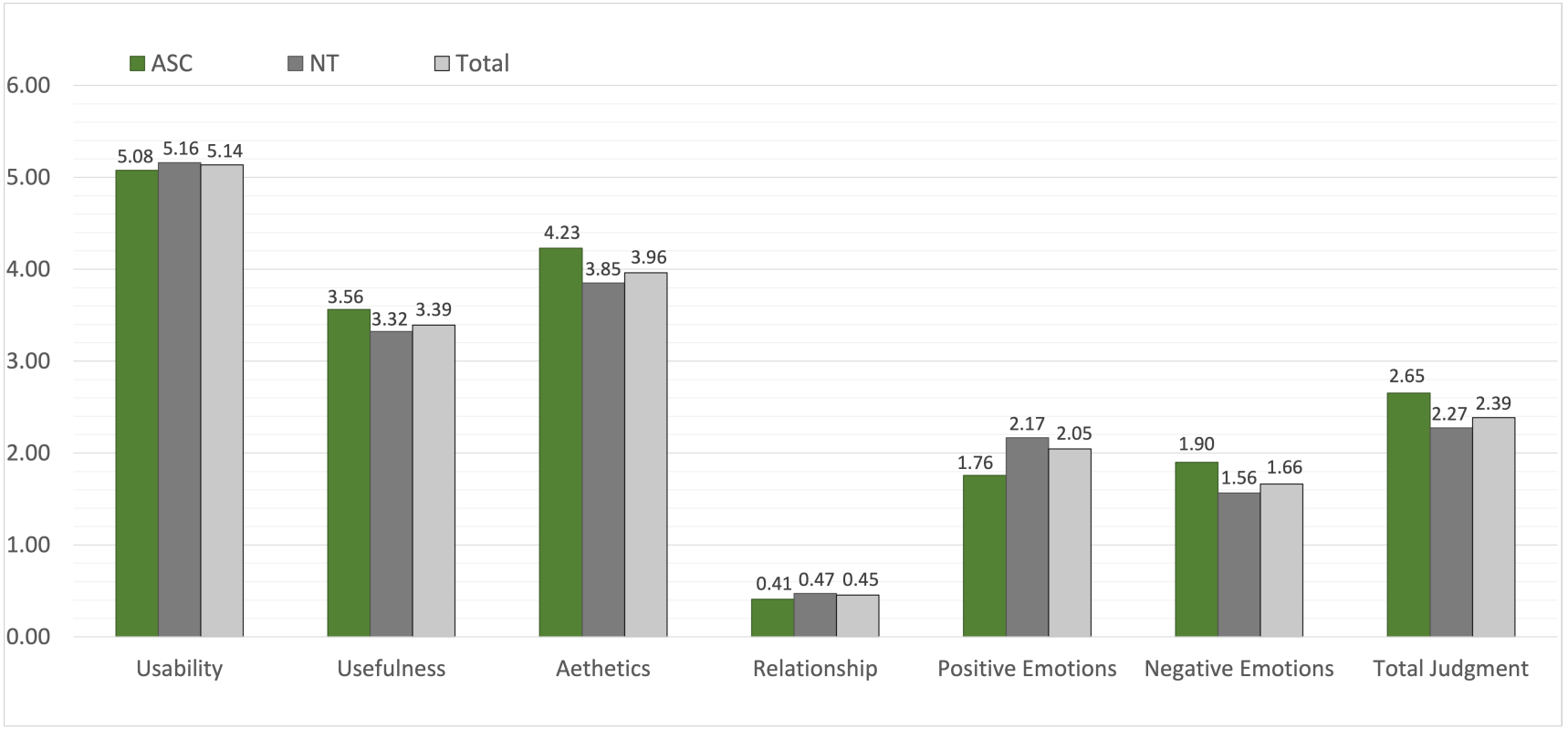


Figure 3.17: Mean differences between ASC, NT and in Total (independent of group) for the seven meCUE subscales

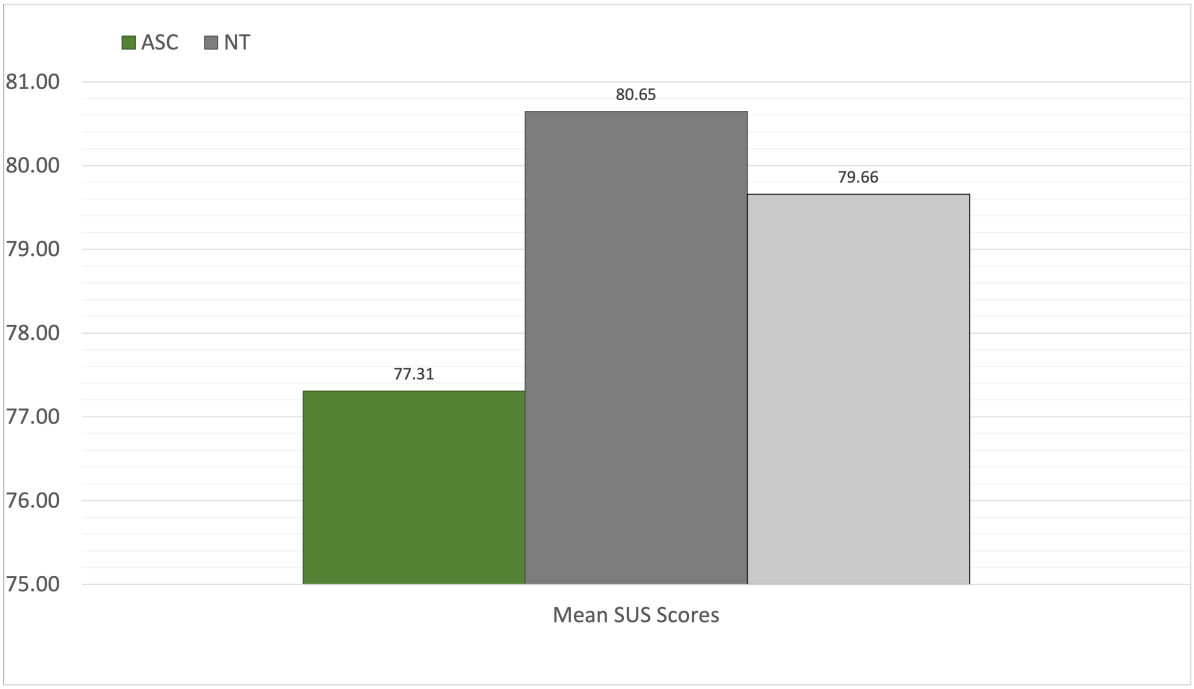


Figure 3.18: Mean differences in SUS score between ASC, NT and in total (independent of group)

4 Discussion

A new social cognition training tool, E.V.A. is described and data from two studies is analyzed to assess usability and effectiveness of the app. E.V.A. was conceived as a new social cognition training tool for adults with ASC and the current studies are the first of their kind. One cross-sectional study with a one-day lab session and a longitudinal study with two lab sessions and a two-week home-training in between, each with NT participants as well as with an autism spectrum condition (ASC), were conducted. Social-emotional measures for characterization of the samples, reaction times for an emotion recognition task as well as usability measures were collected.

4.1 Summary of Results

The overall results from study A indicate that there are indeed some highly significant differences between ASC and NT participants in social-emotional abilities. Results indicate lower cognitive empathic abilities, e.g. the ability to infer mental states of others, in ASC participants, in line with the results of other research (Bos and Stokes (2019), Mazza et al. (2014), Dziobek, Wolf, Bahnemann, Kirchner, and Heekeren (2008)). A lower ability in the comprehension of someone else's point of view from an emotional perspective was also inferred from behavioral measures. In addition, the ability to appropriately recognize emotions expressed by others is found to be lower in people with ASC, also in line with existing research (Kennedy and Adolphs (2012), Walsh, Creighton, and Rutherford (2016)). Lastly, people with ASC achieved lower scores in the ability to use emotional arousal effectively, which is a crucial process for effective emotion regulation (Izard et al. (2011), Cai, Richdale, Uljarević, Dissanayake, and Samson (2018)). All subscales from the self-report questionnaires for social-emotional assessment (except the competence subscale from the SPF-German version of IRI), showed in-between differences amid the two tested samples, highlighting lower scores for the ASC participants. Research indicating why and how exactly the two samples are different is warranted to address interventions designed to improve social cognition skills and therefore lead to greater life satisfaction. The results from the experimental behavioral task underpin the above-mentioned findings. There are differences in the ability to recognize facial emotions rapidly since participants with ASC did not make more mistakes in the emotion recognition task but took significantly longer to respond. These longer reaction times could account for the fact that participants with ASC do usually pay less attention to the facial gestures of others (Klin et al. (2002), Dawson et al. (2005)) especially the eyes (Pelphrey et al. (2002), Baron-Cohen et al. (1997)), and therefore need to control themselves for intentionally contemplating and deciphering the emotional expressing face. These findings suggest that there

could be a need for a training tool such as E.V.A. in ASCs, as the training tool can help to train and improve daily social interaction and social-emotional skills. No differences between the two samples regarding their social and emotional competence were detected (for further details on social and emotional competence please refer to Collie (2020)), indicating that people on the autism spectrum have the same abilities in managing their social-emotional experiences as neurotypically developed people. This finding is indeed conflicting with existing research which reports deficits in social and emotional competence in people with ASC (e.g. Reyes (2013), Reyes, Factor, and Scarpa (2020)). In addition, these findings extend with the results from study B (comparing pre- to post-training in NT participants, see section 3) which suggest that there might not always be an immediate effect on social-emotional skills. Although the results from study B show no pre-post differences in the social-emotional behavioral assessment, the results from the Face Puzzle task indeed indicate improvements in the participants' reaction times for recognizing emotions correctly. Another study supports these findings. Haut, Dodell-Feder, Guty, Nahum, and Hooker (2019) employed a computerized social cognitive training program and also found pre-post improvements on an Empathic Accuracy task, although there were no significant changes in the Interpersonal Reactivity Index (IRI - original English version of the SPF) subscales detected. Other studies also show promising results in the area of social cognition training (please refer to the review by Kouo and Egel (2016)). In addition, there might have been slight improvements in everyday social interactions, but the participants might not have noticed any changes themselves. Reviewing research that indicates that social-emotional abilities and the ability of emotion recognition develop over time (Rump et al. (2009)) suggests that the effects might become visible first in highly sensitive measures such as reaction times. Training effects are already visible for neurotypically developed participants, suggesting that one can improve their social skills with a training tool like E.V.A. while greater improvements in a clinical sample could be expected. Other studies focused mainly on basic emotions or showed only static displays of emotions to their participants while they had unlimited time to respond (and sometimes could view the stimuli for unlimited time). The current research addressed all these characteristics of previous studies as follows. In the Face Puzzle (see section 2.4) basic and more sophisticated emotions were included in a dynamic display of the stimuli by a heterogeneous group of actors. In addition, although the participants could view the stimuli as many times as they needed to, the reaction time for answers was measured (for all responses and the correct responses, separately). Taking this into account with the results from the Face Puzzle Task (significantly higher reaction times for ASCs compared to NTs and no differences in hits and misses for ASCs compared to NTs), they suggest that persons with ASC seem not to have an overall deficit in emotion recognition (no differences in hits and misses for the Face Puzzle task), they rather only need more time to process the presented cues (no differences in hits and misses, but significantly longer reaction times for persons with ASC).

4.2 Overall Usability

No differences were found in the usability measures between the NT and ASC participants. This indicates that ASC and NT participants like E.V.A. equally. Since no differences were found and the usability measures were assessed only once (at the end of the lab session in study A and lab session 2 in study B), the scores across all participants (N=44) were calculated to get an overview of overall perceived usability. The results of the meCUE questionnaire brought insights into the intuitiveness and easiness to use (usability) and the usefulness as task-related subscales. Whereas, relationship and the emotion subscales, on the other hand, shed light on non-task-related usability. More than three quarters of the participants rated E.V.A. as usable, indicating that implementation of the functions was done very well. According to Minge (2018), the results on the usability subscale can be considered as relatively high, whereas those on the usefulness subscale are moderate, for these task-related subscales. These findings suggest that there is still potential for further development of the application with a more user-centered approach, taking detailed feedback from potential users into account. Almost half of the participants rated E.V.A. as visually aesthetic, which can be considered as high in the light of a non-task-related subscale (see Minge (2018)).
The low scores for the relationship subscale are not surprising, since with minimum use of only six hours in two weeks the application could not have become an everyday companion for the participants. In addition, the E.V.A. app is designed to be used with a tablet computer that can hardly become an everyday companion such as a mobile phone, simply because of its size. The findings suggest that these scores could increase with usage time, but cannot rise steadily. Although no significant differences were found for the emotion subscales (meCUE see section 3.3.1), the direction of the differences for both emotion subscales (positive emotion and negative emotion) is interesting. In the positive emotion subscale, the mean score for the NT participants is slightly higher and for the negative emotions, the mean score is slightly higher for the participants with ASC. Here the participants with ASC could feel forced to pay more attention to the face and especially to the eyes of the actor and therefore experiencing higher negative and less positive emotions related to the use of E.V.A. The results from the SUS questionnaire underpin the findings from the meCUE subscale usability since the results from the SUS indicate that the participants rate E.V.A. as "good" to "excellent" (each group separately as well as taken together, independent of the group), according to Bangor and colleagues' suggested categorization (Bangor et al. (2008), Bangor, Kortum, and Miller (2009)). These results indicate that participants from the autism spectrum perceive E.V.A. as useful and user-friendly in accordance with neurotypically developed participants, which accounts for a good underlying concept for the E.V.A. development. Taking these findings together E.V.A. is promising to become a social cognition training tool beyond studies in autism therapy, accompanying therapy, or maybe even for self-training. The adaptive manner of the current application also contributes to the positive acceptance among the participants, which is also influencing their motivation positively. Before making it available for everyone different ethical issues need to be discussed. The overall findings regarding usability, suggest that the newly developed E.V.A. application is accepted very well throughout the two samples of the current two studies, suggesting a promising future for this application and its potential

use in a clinical setting.

4.3 Limitations

The two current studies face some limitations, which can also be hypothetical reasons why no differences were detected for some measures. In general, the sample size is not big enough. Twenty-one NT participants and ten with an autism spectrum condition (for study A) is not optimal, but reasonable as a basis to get an idea of the characteristics of the two populations and how they differ. Although training effects were found in study B for the ten NT participants, such a small sample size might undermine the ability to detect an effect, suggesting that a larger effect size could be found in a bigger sample. The three ASC participants from Study B were mainly left out of the analysis because with such a small sample size no reliable effect can be expected. Only for usability entries, they were included since participants from both studies were taken together.
Results from questionnaires can always be subject to response bias, the effect of unconsciously being influenced by multiple different information. Especially when the same questionnaire is used two times with a relatively short span in between, the second measure can lack sensitivity, due to a repetition effect and therefore miss any training effects. Although in study B effects in the Emotion Recognition task were detected, with such a small sample size replication of the study with a larger sample is desirable. In addition, a matched sample of participants with an autism spectrum condition of greater size should be included to investigate if the effects found in this work will manifest in an ASC sample as well since the findings from study A suggest differences between the two groups for a setting as in study B. The slight differences in the emotions subscales (positive, negative) from the meCUE might just have not reached significance, because of the small sample size. Also, the questionnaires could be subject to ceiling effects, meaning that the participants already achieved the highest scores at the first lab session (T1). Another limitation, for study B, is the relatively short time the participants were exposed to the training material. Two weeks of training with a total of six hours is not enough time, a longer exposure to the training material is needed as done by Golan and Baron-Cohen (2006) with the MindReading Tool, where the intervention group indeed improved. Although their participants engaged with the intervention for ten to fifteen weeks they categorize it as "a relatively short period of time" (Golan and Baron-Cohen (2006), p. 591). Especially social and emotional competence (competence subscale from the SPF questionnaire) is something we humans learn over years, so the two-week home training session could have been just too short to significantly improve this skill. The relatively short time the participants used E.V.A. should be considered when viewing these results, improvements could manifest later and also become visible in a self-report measure after longer exposure. For study B, the time participants engaged in training with the app was estimated by subjective participant ratings. The objective usage time recording within the app needs to be improved. When the length of the study will be increased the possible dropout rate will increase too, which needs to be taken into account when recruiting participants, which might increase the difficulty to recruit participants on the autism spectrum.

4.4 Future Work

Future work could go in two directions, the empirical part as well as the developmental part. There is potential for further development of the application. From a technological perspective, the Voice Module (see section 2.1.1.2) could get extended. An implicit version was already planned to be implemented, where the user needs to match two voice recordings expressing the same emotion, whereas the expressed emotion can be congruent or incongruent to the content. In addition to this task, an improved engagement time recorder could be implemented that the user can retrace the exact time played with the application, which can also serve as a motivational booster. The implementation of the total time used E.V.A., since registration, as well as monthly, weekly, and daily usage, could help the user to motivate themselves and set their own goals on how much time they want to spend training.
There is also potential for further development regarding the content of E.V.A. Not all forty emotions are included yet. That 15 emotions are not yet included becomes visible in the 'net of emotions' (see figure A.1, p. 53 in appendix Appendix A). The circles for already included emotions are solid filled with color and the others only have a solid circle around. Including these 15 additional emotions entails many subtasks which need to be addressed. New training material for all modules (Face, Voice and Film Module, and Library of Emotions) within E.V.A. would need to be recorded and implemented. Another important and vast task will be the detailed inspection and evaluation of the materials before including them. More research-oriented development could be to create a behavioral task, similar to the Face Puzzle task (section 2.4) from Kliemann et al. (2013) and Kliemann (2013)) for voice recordings, where reaction times, as well as hits and misses, can get measured. Reaction times are desirable measures since they are sensitive and objective. An additional approach could be the development of a similar task for the Film Module. To clarify the findings in this work, at least study B should be replicated with a larger sample size and with matched autistic participants as well. Positive and negative emotions related to the use of E.V.A. (subscales from the meCUE questionnaire) should be further investigated as well. A longer training period would be helpful, to make sure participants get accustomed to the tool and its usage. Of course, much larger sample size is needed to compensate for possible dropouts. Another study will be conducted at Humboldt University of Berlin (expected to start February/ March 2021, as of December 2020) to compare internet-based social cognition training in the form of E.V.A. to a regular group therapy, which is specifically designed for the autism spectrum condition.

4.5 Conclusion

The new social cognition training tool is positively accepted by both groups of participants, NT as well as ASC. Current findings suggest that persons on the autism spectrum do not have a general deficit in emotion recognition, since they did not make more mistakes in matching and labeling emotions. They rather suggest that autistic participants need more time to recognize emotions correctly, underpinned by longer reaction times for autistic participants. The results from the behavioral task show promising training effects, for the

emotion recognition task even for NT persons, suggesting that improving emotion recognition abilities with an intervention as E.V.A. is possible. Further studies are needed to verify if these effects manifest in an ASC sample as well. From the gathered findings from both studies, the overall conclusion can be drawn that E.V.A. is a promising tool for improving social cognition in ASC with high usability. Further studies are needed, especially with greater samples and longer training times.

Bibliography

Alpert, J. (2020). Autism: A spectrum disease. *The American journal of medicine*.

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction*, *24*(6), 574–594.

Bangor, A., Kortum, P. T., & Miller, J. T. (2009). Determining what individual sus scores mean: adding an adjective rating scale. *Journal of Usability Studies archive*, *4*, 114-123.

Baron-Cohen, S., Golan, O., Wheelwright, S., & Hill, J. (2004). *Mind reading: The interactive guide to emotions. london: Jessica kingsley.*

Baron-Cohen, S., Wheelwright, S., & Jolliffe, T. (1997). Is there a "language of the eyes"? evidence from normal adults, and adults with autism or asperger syndrome. *Visual Cognition*, *4*, 311-331.

Berggren, S., Fletcher-Watson, S., Milenković, N., Marschik, P., Bölte, S., & Jonsson, U. (2018). Emotion recognition training in autism spectrum disorder: A systematic review of challenges related to generalizability. *Developmental Neurorehabilitation*, *21*, 141 - 154.

Beven, J. P., O'Brien-Malone, A., & Hall, G. (2004). Using the interpersonal reactivity index to assess empathy in violent offenders. *International Journal of Forensic Psychology*, *1*(2), 33–41.

Bölte, S., Golan, O., Goodwin, M. S., & Zwaigenbaum, L. (2010). *What can innovative technologies do for autism spectrum disorders?* Sage Publications.

Bos, J., & Stokes, M. (2019). Cognitive empathy moderates the relationship between affective empathy and wellbeing in adolescents with autism spectrum disorder. *European Journal of Developmental Psychology*, *16*, 433 - 446.

Brooke, J. (1996). Sus: a "quick and dirty'usability. *Usability evaluation in industry*, 189.

Brull, S., & Finlayson, S. (2016). Importance of gamification in increasing learning. *The Journal of Continuing Education in Nursing*, *47*(8), 372–375.

Cai, R. Y., Richdale, A. L., Uljarević, M., Dissanayake, C., & Samson, A. C. (2018). Emotion regulation in autism spectrum disorder: Where we are and where we need to go. *Autism Research*, *11*(7), 962–978.

Castelli, F. (2005). Understanding emotions from standardized facial expressions in autism and normal development. *Autism*, *9*(4), 428–449.

Chiarotti, F., & Venerosi, A. (2020). Epidemiology of autism spectrum disorders: A review of worldwide prevalence estimates since 2014. *Brain Sciences*, *10*(5), 274.

Chirumbolo, A., Picconi, L., Morelli, M., & Petrides, K. V. (2019). The assessment of trait emotional intelligence: Psychometric characteristics of the teique-full form in a large italian adult sample. *Frontiers in Psychology*, *9*, 2786. Retrieved from

https://www.frontiersin.org/article/10.3389/fpsyg.2018.02786 doi: 10.3389/fpsyg.2018.02786

Collie, R. (2020). Social and emotional competence: advancing understanding of what, for whom, and when. *Educational Psychology*, *40*, 663 - 665.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of personality and social psychology*, *44*(1), 113.

Davis, M. H., et al. (1980). A multidimensional approach to individual differences in empathy.

Dawson, G., Webb, S. J., & McPartland, J. (2005). Understanding the nature of face processing impairment in autism: insights from behavioral and electrophysiological studies. *Developmental neuropsychology*, *27*(3), 403–424.

Deruelle, C., Rondan, C., Gepner, B., & Tardif, C. (2004). Spatial frequency and face processing in children with autism and asperger syndrome. *Journal of autism and developmental disorders*, *34*(2), 199–210.

Downs, A., & Strand, P. (2008). Effectiveness of emotion recognition training for young children with developmental delays. *Journal of Early and Intensive Behavior Intervention*, *5*(1), 75.

Dziobek, I., Wolf, I., Bahnemann, M., Kirchner, J., & Heekeren, H. R. (2008). Cognitive and emotional empathy and their neurofunctional correlates in autism spectrum conditions.

Ekman, P., Kuhlmann-Krieg, S., & Reiss, M. (2010). *Gefühle lesen: Wie sie emotionen erkennen und richtig interpretieren* (2nd Edition ed.). Springer.

Fridenson-Hayo, S., Berggren, S., Lassalle, A., Tal, S., Pigat, D., Bölte, S., ... Golan, O. (2016). Basic and complex emotion recognition in children with autism: cross-cultural findings. *Molecular Autism*, *7*.

Frith, C. D. (2008). Social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1499), 2033–2039.

Golan, O., & Baron-Cohen, S. (2006). Systemizing empathy: Teaching adults with asperger syndrome or high-functioning autism to recognize complex emotions using interactive multimedia. *Development and psychopathology*, *18*(2), 591–617.

Grossman, J. B., Klin, A., Carter, A. S., & Volkmar, F. R. (2000). Verbal bias in recognition of facial emotions in children with asperger syndrome. *Journal of Child Psychology and Psychiatry*, *41*(3), 369–379.

Grossman, R. B., & Tager-Flusberg, H. (2008). Reading faces for information about words and emotions in adolescents with autism. *Research in autism spectrum disorders*, *2*(4), 681–695.

Happé, F., & Frith, U. (2020). Annual research review: Looking back to look forward - changes in the concept of autism and implications for future research. *Journal of child psychology and psychiatry, and allied disciplines*.

Harms, M. B., Martin, A., & Wallace, G. (2010). Facial emotion recognition in autism spectrum disorders: A review of behavioral and neuroimaging studies. *Neuropsychology Review*, *20*, 290-322.

Hassenzahl, M. (2001). The effect of perceived hedonic quality on product appealingness. *International Journal of Human-Computer Interaction*, *13*(4), 481-499. Retrieved from https://doi.org/10.1207/S15327590IJHC1304_07 doi: 10.1207/S15327590IJHC1304_07

Haut, K. M., Dodell-Feder, D., Guty, E., Nahum, M., & Hooker, C. (2019). Change in objective measure of empathic accuracy following social cognitive training. *Frontiers in Psychiatry*, *10*.

Haviland, J. M., & Lelwica, M. (1987). The induced affect response: 10-week-old infants' responses to three emotion expressions. *Developmental psychology*, *23*(1), 97.

Hepach, R., Kliemann, D., Grüneisen, S., Heekeren, H. R., & Dziobek, I. (2011). Conceptualizing emotions along the dimensions of valence, arousal, and communicative frequency–implications for social-cognitive tests and training tools. *Frontiers in psychology*, *2*, 266.

Ireland, J. L. (1999). Provictim attitudes and empathy in relation to bullying behaviour among prisoners. *Legal and Criminological Psychology*, *4*(1), 51–66.

Isaksson, J., Van't Westeinde, A., Cauvet, É., Kuja-Halkola, R., Lundin, K., Neufeld, J., ... Bölte, S. (2019). Social cognition in autism and other neurodevelopmental disorders: a co-twin control study. *Journal of autism and developmental disorders*, *49*(7), 2838–2848.

Izard, C. E., Woodburn, E. M., Finlon, K. J., Krauthamer-Ewing, E. S., Grossman, S. R., & Seidenfeld, A. (2011). Emotion knowledge, emotion utilization, and emotion regulation. *Emotion Review*, *3*(1), 44-52. Retrieved from `https://doi.org/10.1177/1754073910380972` doi: 10.1177/1754073910380972

Kennedy, D., & Adolphs, R. (2012). Perception of emotions from facial expressions in high-functioning adults with autism. *Neuropsychologia*, *50*, 3313-3319.

Kliemann, D. (2013). *Implicit and explicit facial emotion recognition in autism spectrum disorder* (Doctoral dissertation). Retrieved from `http://dx.doi.org/10.17169/refubium-16247`

Kliemann, D., Rosenblau, G., Bölte, S., Heekeren, H. R., & Dziobek, I. (2013). Face puzzle—two new video-based tasks for measuring explicit and implicit aspects of facial emotion recognition. *Frontiers in psychology*, *4*, 376.

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of general psychiatry*, *59*(9), 809–816.

Kouo, J. L., & Egel, A. L. (2016). The effectiveness of interventions in teaching emotion recognition to children with autism spectrum disorder. *Review Journal of Autism and Developmental Disorders*, *3*(3), 254–265.

Kuusikko, S., Haapsamo, H., Jansson-Verkasalo, E., Hurtig, T., Mattila, M.-L., Ebeling, H., ... Moilanen, I. (2009). Emotion recognition in children and adolescents with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*, 938-945.

Köhne, S. (2016). *The role of imitation and synchronization in empathic functions: Insights from autism spectrum disorders* (Doctoral dissertation). Retrieved from `http://dx.doi.org/10.17169/refubium-8563`

Lai, M.-C., & Baron-Cohen, S. (2015). Identifying the lost generation of adults with autism spectrum conditions. *The Lancet Psychiatry*, *2*(11), 1013–1027.

Lee, C., Lam, S., Tsang, S. T. K., Yuen, C. M. C., & Ng, C. K. M. (2018). The effectiveness of technology-based intervention in improving emotion recognition through facial expression in people with autism spectrum disorder: a systematic review. *Review Journal*

of Autism and Developmental Disorders, *5*, 91-104.

Leppänen, J., Moulson, M. C., Vogel-Farley, V. K., & Nelson, C. A. (2007). An erp study of emotional face processing in the adult and infant brain. *Child development*, *78 1*, 232-45.

Lindner, J. L., & Rosén, L. A. (2006). Decoding of emotion through facial expression, prosody and verbal content in children and adolescents with asperger's syndrome. *Journal of autism and developmental disorders*, *36*(6), 769–777.

Macleod, J. (2016). Generalization of learning in adolescents and adults with autism spectrum disorder..

Mazza, M., Pino, M. C., Mariano, M., Tempesta, D., Ferrara, M., Berardis, D. D., ... Valenti, M. (2014). Affective and cognitive empathy in adolescents with autism spectrum disorder. *Frontiers in Human Neuroscience*, *8*.

McCann, J., & Peppé, S. (2003). Prosody in autism spectrum disorders: a critical review. *International Journal of Language & Communication Disorders*, *38*(4), 325–350.

Meiselwitz, G., & Sadera, W. (2008). Investigating the connection between usability and learning outcomes in online learning environments. *Journal of Online Learning and Teaching*, *4*(2), 234–42.

Minge, M. (2018). Nutzererleben messen mit dem mecue 2.0 – ein tool für alle fälle? In R. Dachselt & G. Weber (Eds.), *Mensch und computer 2018 - workshopband*. Bonn: Gesellschaft für Informatik e.V. doi: 10.18420/muc2018-ws16-0485

Minge, M., & Riedel, L. (2013). mecue-ein modularer fragebogen zur erfassung des nutzungserlebens. *Mensch & Computer 2013: Interaktive Vielfalt*.

Minge, M., Riedel, L., & Thüring, M. (2014). Modulare evaluation interaktiver technik. entwicklung und validierung des mecue fragebogens zur messung der user experience. *Grundlagen und Anwendungen der Mensch-Technik-Interaktion*, *10*, 28–36.

Moebert, T., Schneider, J., Zoerner, D., Tscherejkina, A., & Lucke, U. (2019). 4. how to use socio-emotional signals for adaptive training. *Personalized human-computer interaction*, *103*, 132.

Paulus, C. (2009). Der Saarbrücker Persönlichkeitsfragebogen SPF (IRI) zur Messung von Empathie. *Psychometrische Evaluation der deutschen Version des Interpersonal Reactivity Index. URL: http://psydok. sulb. uni-saarland. de/volltexte/2009/2363*.

Pelphrey, K. A., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., & Piven, J. (2002). Visual scanning of faces in autism. *Journal of autism and developmental disorders*, *32*(4), 249–261.

Philip, R., Whalley, H., Stanfield, A., Sprengelmeyer, R. H., Santos, I., Young, A., ... others (2010). Deficits in facial, body movement and vocal emotional processing in autism spectrum disorders. *Psychological medicine*.

Pierce, K., Glad, K. S., & Schreibman, L. (1997). Social perception in children with autism: an attentional deficit? *Journal of Autism and Developmental Disorders*, *27*(3), 265–282.

Prinz, J. (2004). Which emotions are basic. *Emotion, evolution, and rationality*, *69*, 88.

Rea, H., LaMotte, K., & Burrell, T. L. (2018). What is autism spectrum disorder? In *Handbook of parent-child interaction therapy for children on the autism spectrum* (pp. 3–26). Springer.

Reyes, N. M. (2013). *Emotion regulation and emotionality: an examination of correlates of social skills in young children with autism spectrum disorders and their typically developing peers* (Unpublished doctoral dissertation). Virginia Polytechnic Institute and State University.

Reyes, N. M., Factor, R., & Scarpa, A. (2020). Emotion regulation, emotionality, and expression of emotions: A link between social skills, behavior, and emotion problems in children with asd and their peers. *Research in Developmental Disabilities*, *106*, 103770.

Rimland, B. (1964). Infantile autism: The syndrome and its implications for a neural theory of behavior..

Rosenblau, G., O'Connell, G., Heekeren, H., & Dziobek, I. (2019). Neurobiological mechanisms of social cognition treatment in high-functioning adults with autism spectrum disorder. *Psychological medicine*, 1-11.

Rosset, D. B., Rondan, C., Da Fonseca, D., Santos, A., Assouline, B., & Deruelle, C. (2008). Typical emotion processing for cartoon but not for real faces in children with autistic spectrum disorders. *Journal of autism and developmental disorders*, *38*(5), 919–925.

Rump, K. M., Giovannelli, J. L., Minshew, N., & Strauss, M. (2009). The development of emotion recognition in individuals with autism. *Child development*, *80 5*, 1434-47.

Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, cognition and personality*, *9*(3), 185–211.

Schneider, J., Dziobek, I., & Weigand, A. (2019). *Vom Emotionsmodell zum adaptiven Lernen: Emotionssensitive Systeme zum Training sozialer Kognition* $EMOTISK$*: Schlussbericht EMOTISK* $Teilprojekt Berlin$ *zum BMBF-Verbundprojekt: Teilprojekt Berlin: Automatische Klassifikation komplexer Emotionsausdrücke und adaptives Training sozialer Kognition: Laufzeit des Vorhabens: 01.04.2015-31.03.2018, kostenneutral verlängert bis 30.09.2018, Automatische Klassifikation komplexer Emotionsausdrücke und adaptives Training sozialer Kognition, Automatic classification of complex emotion expressions and adaptive training of social cognition* (Tech. Rep.). Berlin: Humboldt-Universität zu Berlin, Berlin School of Mind and Brain. doi: 10.2314/GBV:1663635056

Schutte, N. S., Malouff, J. M., & Bhullar, N. (2009). The assessing emotions scale. In *Assessing emotional intelligence* (pp. 119–134). Springer.

Schutte, N. S., Malouff, J. M., Hall, L. E., Haggerty, D. J., Cooper, J. T., Golden, C. J., & Dornheim, L. (1998). Development and validation of a measure of emotional intelligence. *Personality and individual differences*, *25*(2), 167–177.

Schwenck, C., Mergenthaler, J., Keller, K., Zech, J., Salehi, S., Taurines, R., ... others (2012). Empathy in children with autism and conduct disorder: Group-specific profiles and developmental aspects. *Journal of Child Psychology and Psychiatry*, *53*(6), 651–659.

Sharma, S. R., Gonda, X., & Tarazi, F. I. (2018). Autism spectrum disorder: classification, diagnosis and therapy. *Pharmacology & therapeutics*, *190*, 91–104.

Sucksmith, E., Allison, C., Baron-Cohen, S., Chakrabarti, B., & Hoekstra, R. A. (2013). Empathy and emotion recognition in people with autism, first-degree relatives, and controls. *Neuropsychologia*, *51*(1), 98–105.

Thüring, M., & Minge, M. (2014). Nutzererleben messen–geht das überhaupt? *WIsseNsCHAFT TRIFFT PRAXIs*, 45.

Thüring, M., & Mahlke, S. (2007). Usability, aesthetics and emotions in human–technology interaction. *International Journal of Psychology*, *42*(4), 253-264. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1080/00207590701396674 doi: 10.1080/00207590701396674

Tracy, J. L., Robins, R. W., Schriber, R. A., & Solomon, M. (2011). Is emotion recognition impaired in individuals with autism spectrum disorders? *Journal of autism and developmental disorders*, *41*(1), 102–109.

Uljarević, M., & Hamilton, A. (2013). Recognition of emotions in autism: A formal meta-analysis. *Journal of Autism and Developmental Disorders*, *43*, 1517-1526.

Walsh, J. A., Creighton, S. E., & Rutherford, M. (2016). Emotion perception or social cognitive complexity: what drives face processing deficits in autism spectrum disorder? *Journal of autism and developmental disorders*, *46*(2), 615–623.

World Health Organization. (2020). International classification of diseases for mortality and morbidity statistics - ICD-11. Retrieved from https://icd.who.int/browse11/l-m/en

Zoerner, D., Moebert, T., & Lucke, U. (2017). IT-gestütztes Training sozio-emotionaler Kognition für Menschen mit Autismus. *Informatik-Spektrum*, *40*(6), 546–555.

Appendix A: Additional Screenshots

A.1 Net of Emotions

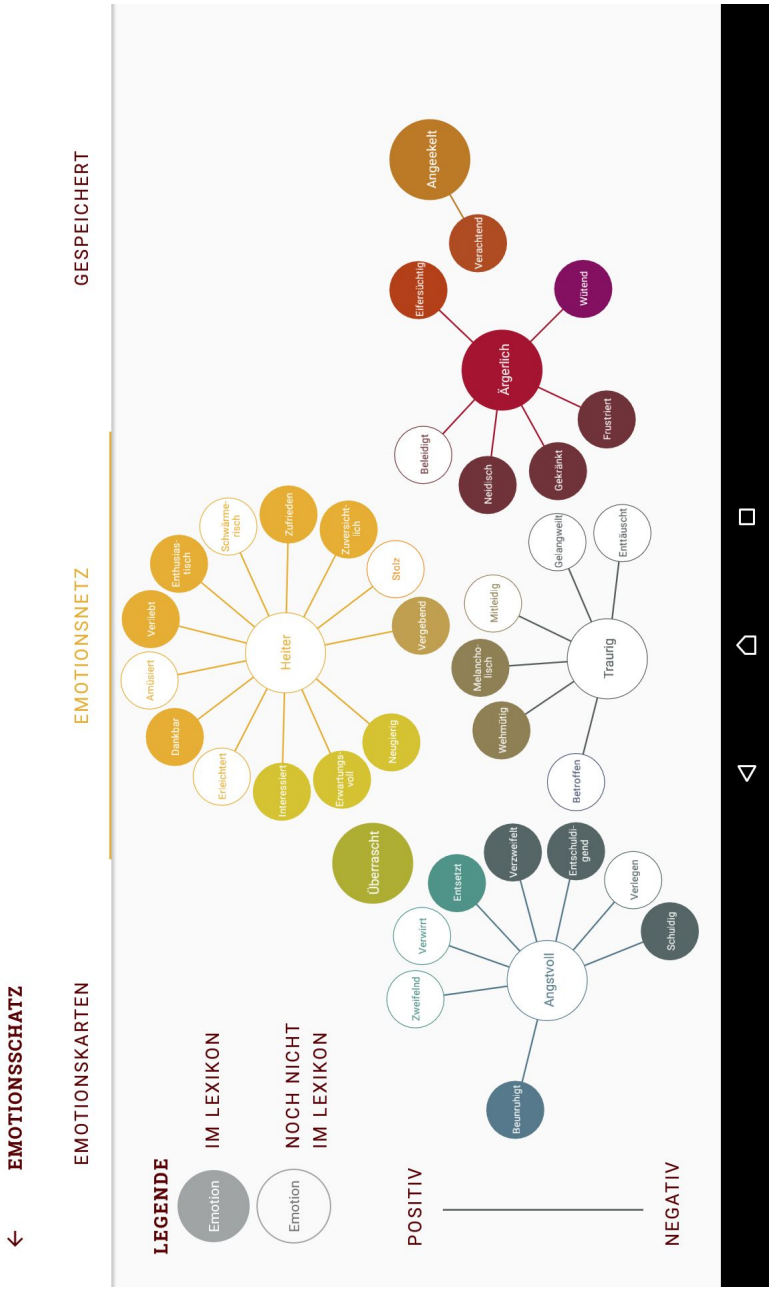


Figure A.1: The "net of emotions": shows the 40 emotions included in E.V.A. The net groups the emotions according to whether they are more positive or negative and to which group of basic emotions this particular emotion belongs to. The solid coloured circles indicate that this particular emotion has it's own entry in the library of emotions.

A.2 Emotion Library Overview - Example: Reliefed

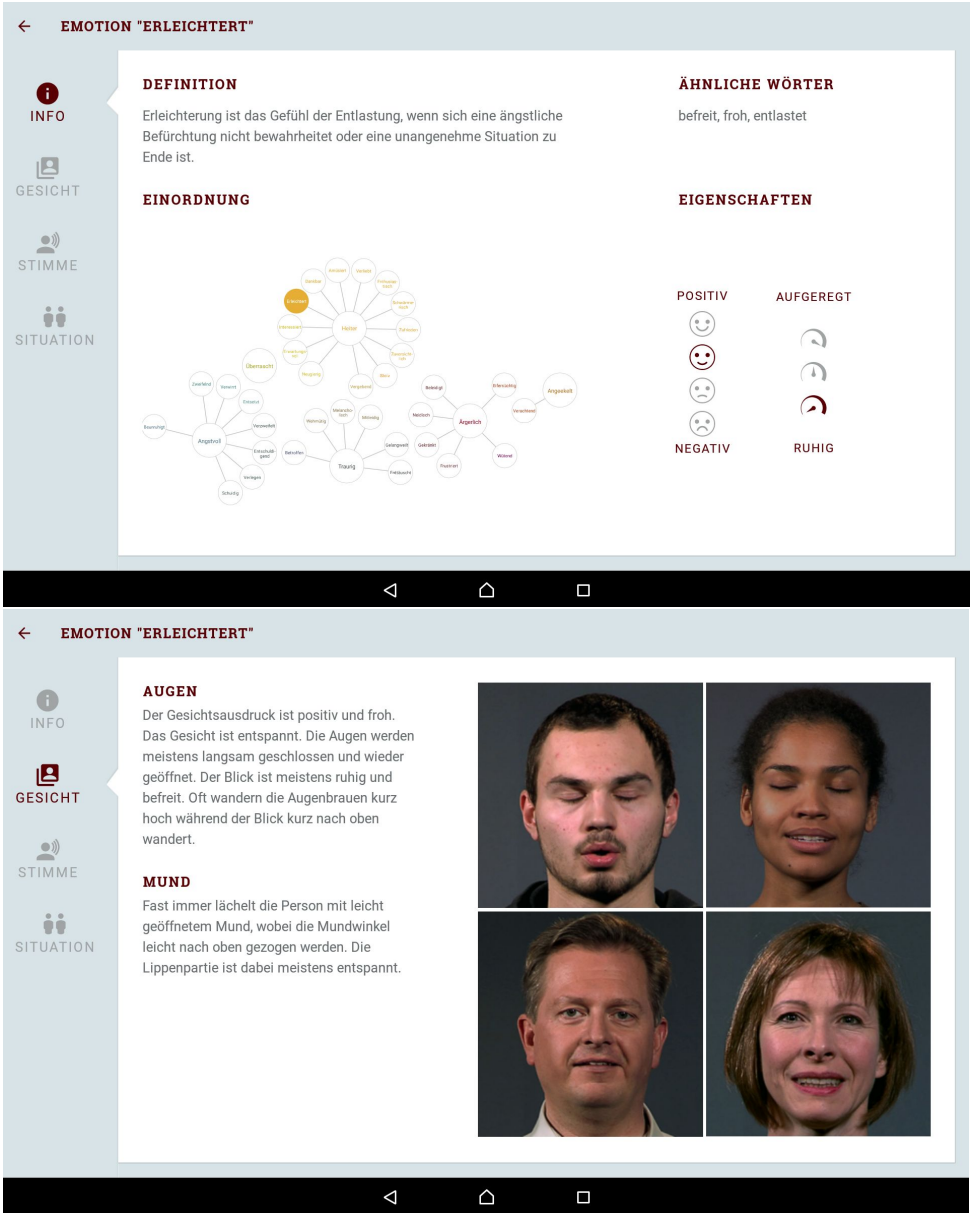


Figure A.2: Top: Overview with detailed description of particular emotion
Bottom: Details on facial changes and example pictures

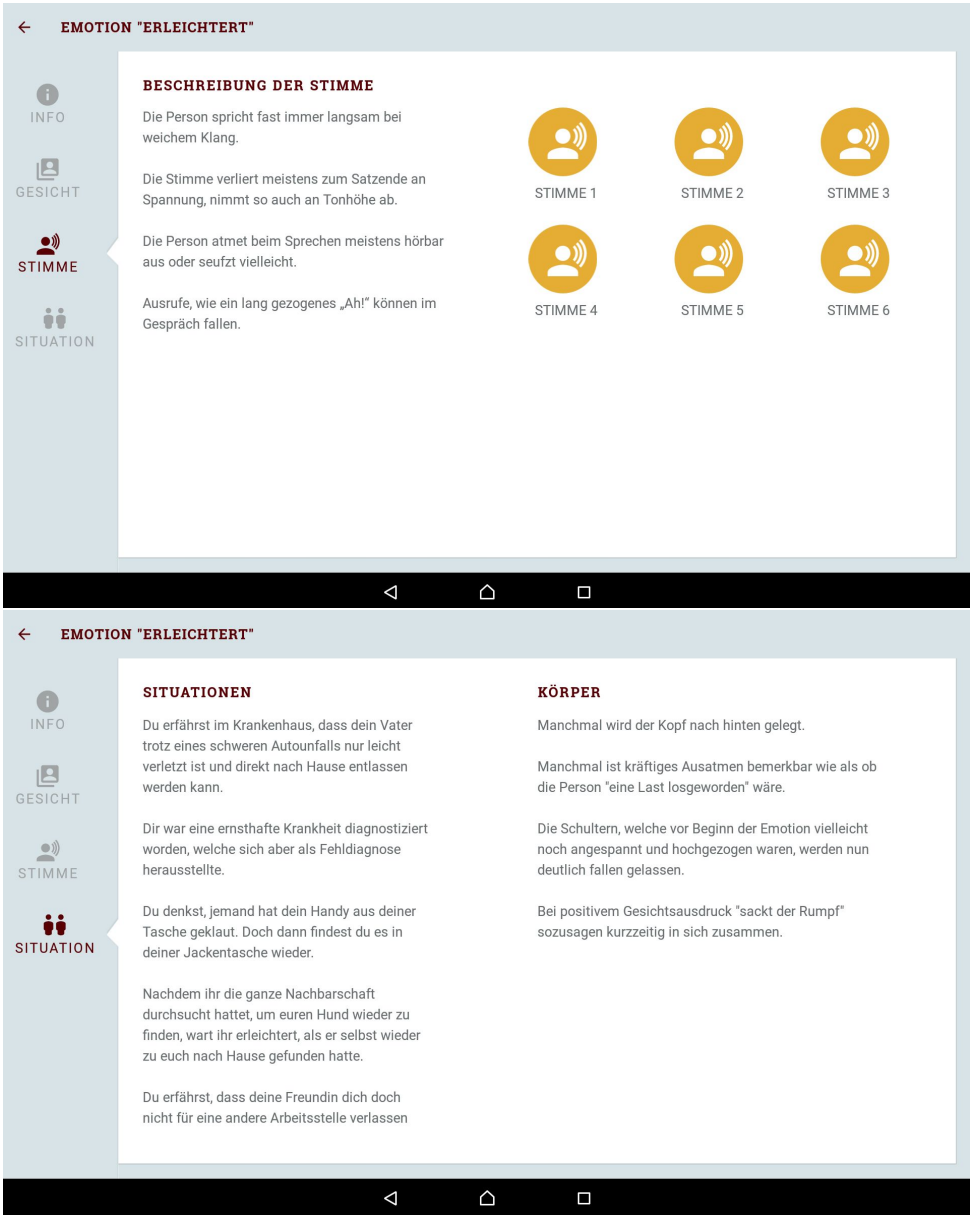
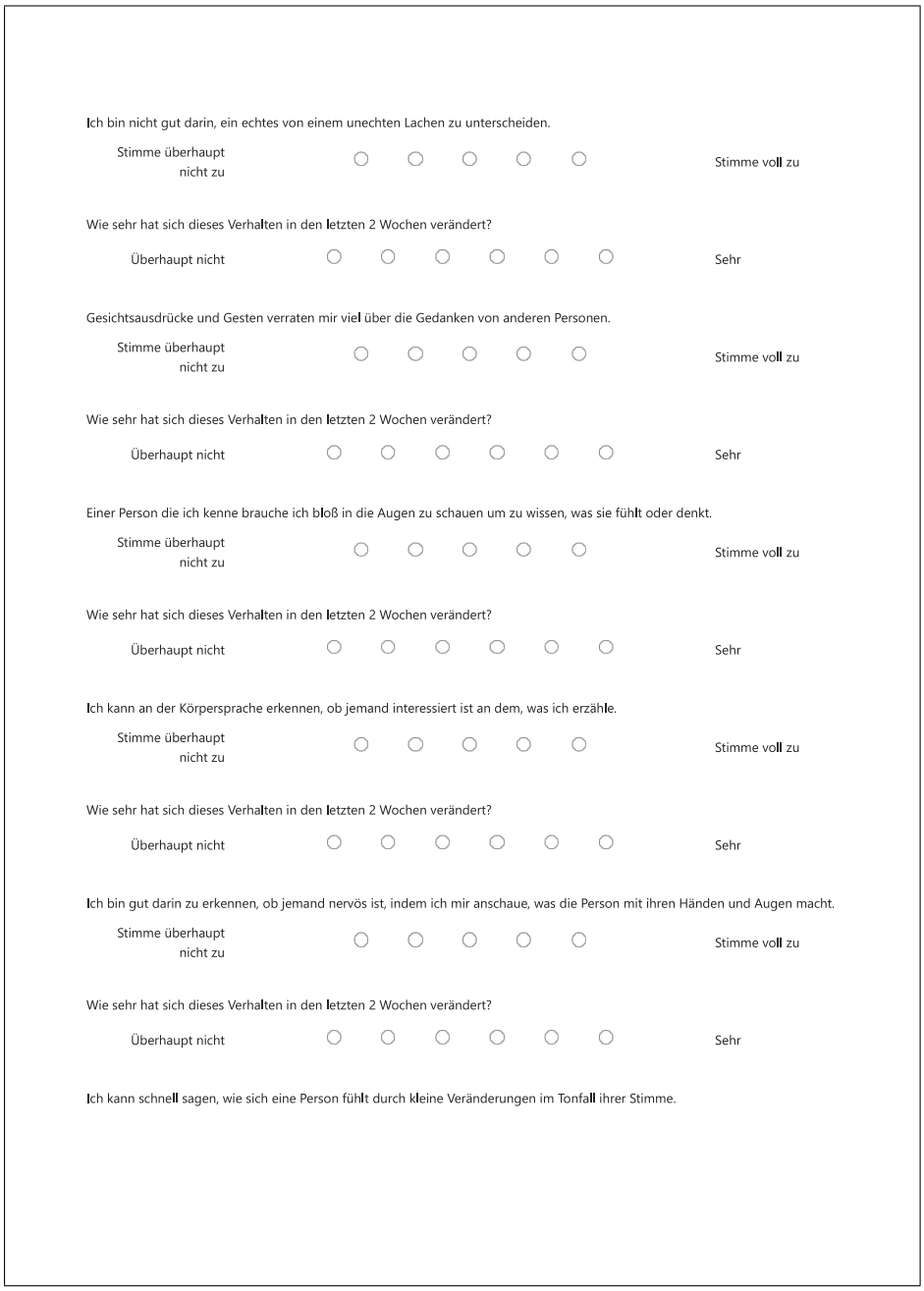


Figure A.3: Top: Details on changes in the voice and example recordings where this particular emotion is expressed
Bottom: Example descriptions of situations one could experience this emotion and changes in body posture

Appendix B: Questionnaires

B.1 CEEQ Questionnaire



Ich bin nicht gut darin, ein echtes von einem unechten Lachen zu unterscheiden.

Stimme überhaupt nicht zu ○ ○ ○ ○ ○ Stimme voll zu

Wie sehr hat sich dieses Verhalten in den letzten 2 Wochen verändert?

Überhaupt nicht ○ ○ ○ ○ ○ ○ Sehr

Gesichtsausdrücke und Gesten verraten mir viel über die Gedanken von anderen Personen.

Stimme überhaupt nicht zu ○ ○ ○ ○ ○ Stimme voll zu

Wie sehr hat sich dieses Verhalten in den letzten 2 Wochen verändert?

Überhaupt nicht ○ ○ ○ ○ ○ ○ Sehr

Einer Person die ich kenne brauche ich bloß in die Augen zu schauen um zu wissen, was sie fühlt oder denkt.

Stimme überhaupt nicht zu ○ ○ ○ ○ ○ Stimme voll zu

Wie sehr hat sich dieses Verhalten in den letzten 2 Wochen verändert?

Überhaupt nicht ○ ○ ○ ○ ○ ○ Sehr

Ich kann an der Körpersprache erkennen, ob jemand interessiert ist an dem, was ich erzähle.

Stimme überhaupt nicht zu ○ ○ ○ ○ ○ Stimme voll zu

Wie sehr hat sich dieses Verhalten in den letzten 2 Wochen verändert?

Überhaupt nicht ○ ○ ○ ○ ○ ○ Sehr

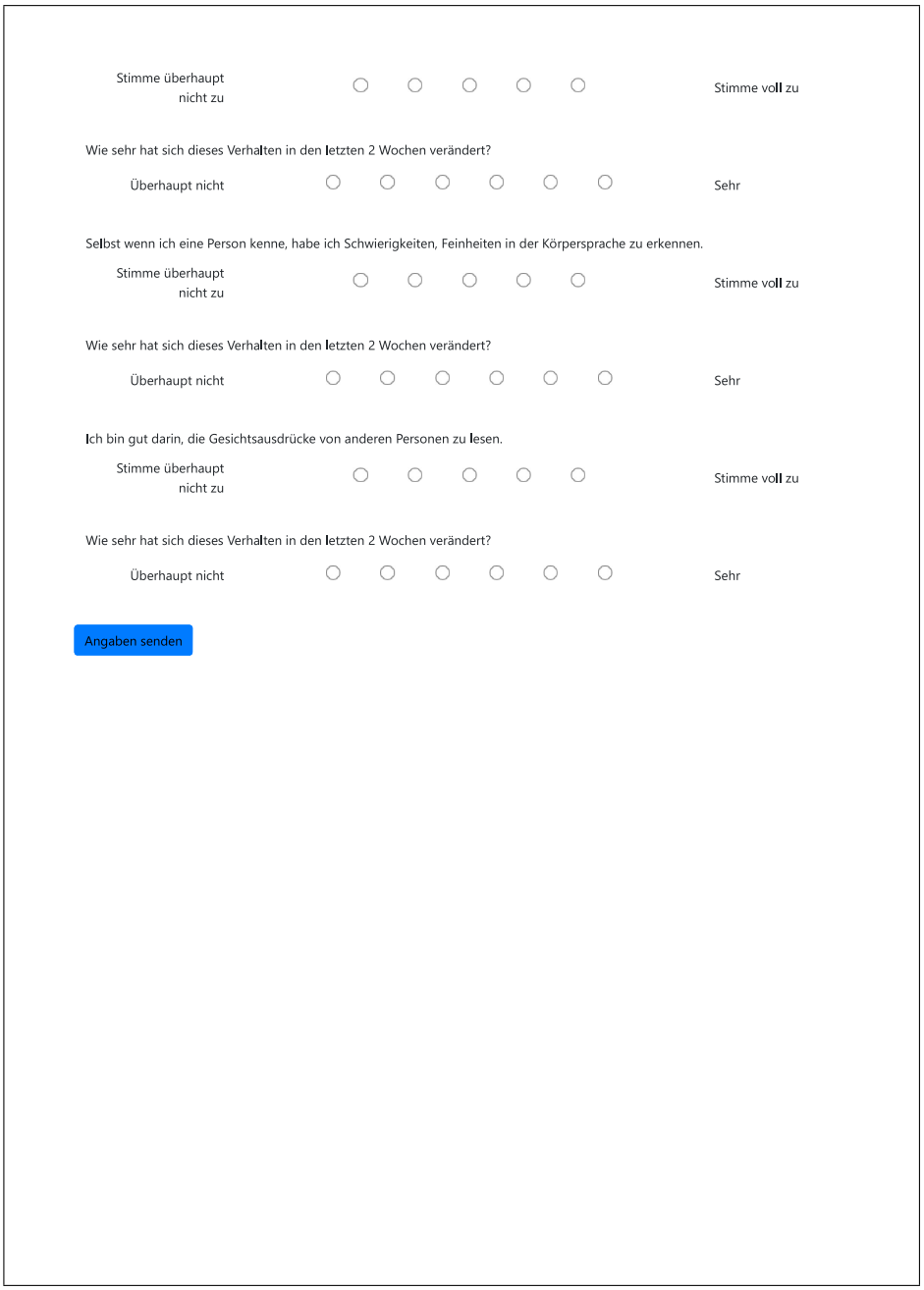
Ich bin gut darin zu erkennen, ob jemand nervös ist, indem ich mir anschaue, was die Person mit ihren Händen und Augen macht.

Stimme überhaupt nicht zu ○ ○ ○ ○ ○ Stimme voll zu

Wie sehr hat sich dieses Verhalten in den letzten 2 Wochen verändert?

Überhaupt nicht ○ ○ ○ ○ ○ ○ Sehr

Ich kann schnell sagen, wie sich eine Person fühlt durch kleine Veränderungen im Tonfall ihrer Stimme.


Stimme überhaupt nicht zu
Stimme voll zu
Wie sehr hat sich dieses Verhalten in den letzten 2 Wochen verändert?
Überhaupt nicht
Sehr
Selbst wenn ich eine Person kenne, habe ich Schwierigkeiten, Feinheiten in der Körpersprache zu erkennen.
Stimme überhaupt nicht zu
Stimme voll zu
Wie sehr hat sich dieses Verhalten in den letzten 2 Wochen verändert?
Überhaupt nicht
Sehr
Ich bin gut darin, die Gesichtsausdrücke von anderen Personen zu lesen.
Stimme überhaupt nicht zu
Stimme voll zu
Wie sehr hat sich dieses Verhalten in den letzten 2 Wochen verändert?
Überhaupt nicht
Sehr
Angaben senden

B.2 SREIT Questionnaire

SREIT (Schutte, 1998) - Skala zur Beurteilung von Emotionen

Anweisungen:
Jedes der folgenden Elemente fragt Sie nach Ihren Gefühlen oder Reaktionen, die mit Emotionen in Verbindung gebracht werden. Bitte lesen Sie sich jede Aussage sorgfältig durch, bevor Sie antworten. Kreuzen Sie, je nachdem wie zutreffend Sie jede Aussage für sich beurteilen, die entsprechende Zahl aus der oben stehenden Antwortskala an.
Es gibt keine „richtigen" oder „falschen" Antworten. Bitte kreuzen Sie die Antwortoption an, die Sie am besten beschreibt.

TRIFFT GAR NICHT ZU **0 1 2 3 4** **TRIFFT VÖLLIG ZU**

PE 5. Es fällt mir schwer, nonverbale Botschaften anderer Menschen zu verstehen.

UE 6. Einige der wichtigsten Ereignisse in meinem Leben haben mich dazu veranlasst, neu zu bewerten, was wichtig ist und was nicht.

UE 7. Wenn sich meine Stimmung verändert, sehe ich neue Möglichkeiten aufkommen.

UE 8. Emotionen gehören zu den Dingen, die mein Leben lebenswert machen.

PE 9. Ich bin mir meiner Emotionen bewusst während ich sie erlebe.

PE 15. Ich bin mir der nonverbalen Botschaften bewusst, die ich an andere Menschen aussende.

UE 17. Wenn ich in einer positiven Stimmung bin, fällt es mir leicht, Probleme zu lösen.

PE 18. Durch das Betrachten ihrer Gesichtsausdrücke erkenne ich die Emotionen, die Menschen erleben.

PE 19. Ich weiß, warum sich meine Gefühle verändern.

UE 20. Wenn ich gut gelaunt bin, bin ich in der Lage, neue Ideen zu entwickeln.

PE 22. Mir fällt es leicht, meine Gefühle zu erkennen während ich sie erlebe.

PE 25. Ich bin mir der nonverbalen Botschaften bewusst, die andere Menschen aussenden.

UE 27. Wenn ich wahrnehme, dass sich meine Emotionen verändern, neige ich dazu, neue Ideen zu entwickeln.

PE 29. Ich muss andere Menschen nur ansehen, um zu wissen, was sie fühlen.

PE 32. Ich kann anhand des Tonfalls in der Stimme anderer Menschen beurteilen, wie sie sich fühlen.

PE 33. Es fällt mir schwer zu verstehen, warum Menschen so fühlen, wie sie fühlen.

PE = perception of emotions
UE = utilization of emotion

B.3 SPF Questionnaire

UNIVERSITÄT DES SAARLANDES
© 2004
Saarbrücker Persönlichkeits-Fragebogen (SPF)
Dr. Christoph Paulus
FR Erziehungswissenschaft
66041 Saarbrücken

Saarbrücker Persönlichkeits-Fragebogen (SPF)

Based on the Interpersonal Reactivity Index (IRI) including new items

Sie werden jetzt eine Reihe von Aussagen lesen, die jeweils bestimmte (verallgemeinerte) menschliche Eigenschaften oder Reaktionen beschreiben, die alle etwas mit Lernen zu tun haben. Bitte kennzeichnen Sie dann auf der 5-Punkte-Skala, inwieweit diese Aussage auf Sie zutrifft; je höher die Zahl, desto höher die Zustimmung. Vielleicht fällt Ihnen auch zu der einen oder anderen allgemeinen Beschreibung ein konkretes Erlebnis ein.

Es gibt dabei keine richtigen oder falschen Antworten.

Falls eine Antwortskala vorgegeben ist**, markieren Sie Ihre Antwort durch einen Kreis um die Zahl** (je größer die Zahl 1-5, desto höher ist Ihre Zustimmung)**,** ansonsten kreuzen Sie einfach die gewünschte Alternative an. Falls Sie eine Änderung vornehmen wollen, so kreuzen Sie die nicht gewünschte Antwort durch und markieren die gewünschte Antwort erneut mit einem Kreis

Ein Beispiel:

Inwieweit trifft die Aussage auf Sie zu?	trifft gar nicht zu ..... 1 (- -)	2 (-)	3 (o)	4 (+)	.....trifft sehr gut zu 5 (+ +)
Ich esse gerne Spaghetti	1 (- -)	2 (-)	3 (o)	4 (+)	5 (+ +)
Ich gehe gerne ins Kino	1 (- -)	2 (-)	3 (o)	4 (+)	5 (+ +)
Ich trinke gerne Bier	1 (- -)	2 (-)	3 (o)	4 (+)	5 (+ +)

Irrtum → 2 (-)

4 (+) ← richtig

Wir danken Ihnen bereits jetzt für Ihre Mitarbeit und Ihre ehrliche Beantwortung der Fragen.
Bitte beginnen Sie nun .....

UNIVERSITÄT DES SAARLANDES
© 2004

Saarbrücker Persönlichkeits-Fragebogen (SPF)

Dr. Christoph Paulus
FR Erziehungswissenschaft
66041 Saarbrücken

		trifft gar nicht zu ..... trifft sehr gut zu				
1f	Manchmal träume ich so vor mich hin von Dingen, die mir passieren könnten.	1 (- -)	2 (-)	3 (o)	4 (+)	5 (+ +)
2e	Ich empfinde oft warmherzige Gefühle für Leute, denen es weniger gut geht als mir.	1 (- -)	2 (-)	3 (o)	4 (+)	5 (+ +)
3k	Wenn ich bei Dingen helfen soll, die ich nicht gut kann, fühle ich mich unsicher.	1 (- -)	2 (-)	3 (o)	4 (+)	5 (+ +)
4p	Ich finde, manchmal ist es sehr schwierig, Dinge aus der Sicht Anderer zu sehen.	1 (- -)	2 (-)	3 (o)	4 (+)	5 (+ +)
5e	Manchmal tun mir Leute, die Probleme haben, gar nicht leid.	1 (- -)	2 (-)	3 (o)	4 (+)	5 (+ +)
6f	Die Gefühle einer Person in einem Roman kann ich mir oft sehr gut vorstellen.	1 (- -)	2 (-)	3 (o)	4 (+)	5 (+ +)
7d	In Notfallsituationen fühle ich mich ängstlich und unbehaglich.	1 (- -)	2 (-)	3 (o)	4 (+)	5 (+ +)
8k	Wenn mich jemand um Hilfe bittet, möchte ich schon selber gerne entscheiden, wann ich mit der Hilfe anfange und wie lange ich helfe.	1 (- -)	2 (-)	3 (o)	4 (+)	5 (+ +)
9f	Gewöhnlich bin ich objektiv, wenn ich einen Film oder ein Theaterstück sehe und lasse mich nicht davon fesseln.	1 (- -)	2 (-)	3 (o)	4 (+)	5 (+ +)
10p	Ich versuche, bei einem Streit zuerst beide Seiten zu verstehen, bevor ich eine Entscheidung treffe.	1 (- -)	2 (-)	3 (o)	4 (+)	5 (+ +)
11e	Wenn ich sehe, wie jemand ausgenutzt wird, glaube ich, ihn schützen zu müssen.	1 (- -)	2 (-)	3 (o)	4 (+)	5 (+ +)
12d	Manchmal fühle ich mich hilflos, wenn ich inmitten einer sehr emotionsgeladenen Situation bin.	1 (- -)	2 (-)	3 (o)	4 (+)	5 (+ +)
13k	Ich suche mir oft Probleme, bei denen ich vorher nicht weiß, ob ich sie meistern kann.	1 (- -)	2 (-)	3 (o)	4 (+)	5 (+ +)
14p	Ich versuche manchmal, meine Freunde besser zu verstehen, indem ich mir vorstelle, wie die Dinge aus ihrer Sicht aussehen könnten.	1 (- -)	2 (-)	3 (o)	4 (+)	5 (+ +)
15f	Von einem guten Buch oder einem spannenden Film so richtig in den Bann geschlagen zu werden, passiert mir so gut wie nie.	1 (- -)	2 (-)	3 (o)	4 (+)	5 (+ +)
16d	Wenn ich sehe, wie jemand verletzt wird, versuche ich, Ruhe zu bewahren.	1 (- -)	2 (-)	3 (o)	4 (+)	5 (+ +)
17e	Meistens stört mich das Unglück anderer Leute nicht besonders.	1 (- -)	2 (-)	3 (o)	4 (+)	5 (+ +)
18k	Oft biete ich von mir aus Hilfe an, auch wenn ich nicht gefragt wurde.	1 (- -)	2 (-)	3 (o)	4 (+)	5 (+ +)
19p	Wenn ich mir sicher bin, dass ich Recht habe, verschwende ich keine Zeit damit, mir die Argumente anderer Leute anzuhören.	1 (- -)	2 (-)	3 (o)	4 (+)	5 (+ +)
20f	Nachdem ich einen Film gesehen habe, fühle ich mich manchmal so, als ob ich eine der Personen aus diesem Film sei.	1 (- -)	2 (-)	3 (o)	4 (+)	5 (+ +)

UNIVERSITÄT DES SAARLANDES
© 2004
Saarbrücker Persönlichkeits-Fragebogen (SPF)
Dr. Christoph Paulus
FR Erziehungswissenschaft
66041 Saarbrücken

21k	Bei Dingen, die ich besser kann als andere, bin ich gerne behilflich.	**1 (- -)**	**2 (-)**	**3 (o)**	**4 (+)**	**5 (+ +)**
22d	In einer gespannten emotionalen Situation zu sein, beängstigt mich.	**1 (- -)**	**2 (-)**	**3 (o)**	**4 (+)**	**5 (+ +)**
23e	Wenn ich sehe, wie jemand unfair behandelt wird, habe ich manchmal überhaupt kein Mitleid mit ihm.	**1 (- -)**	**2 (-)**	**3 (o)**	**4 (+)**	**5 (+ +)**
24d	Ich kann in Notfällen gewöhnlich recht effektiv handeln.	**1 (- -)**	**2 (-)**	**3 (o)**	**4 (+)**	**5 (+ +)**
25e	Oft berühren mich Dinge sehr, die ich nur beobachte.	**1 (- -)**	**2 (-)**	**3 (o)**	**4 (+)**	**5 (+ +)**
26k	Ich kann ganz gut abschätzen, ob jemand Hilfe benötigt oder nicht.	**1 (- -)**	**2 (-)**	**3 (o)**	**4 (+)**	**5 (+ +)**
27p	Ich glaube, jedes Problem hat zwei Seiten und versuche deshalb beide zu berücksichtigen.	**1 (- -)**	**2 (-)**	**3 (o)**	**4 (+)**	**5 (+ +)**
28e	Ich würde mich selbst als eine ziemlich weichherzige Person bezeichnen.	**1 (- -)**	**2 (-)**	**3 (o)**	**4 (+)**	**5 (+ +)**
29f	Wenn ich einen guten Film sehe, kann ich mich sehr leicht in die Hauptperson hineinversetzen.	**1 (- -)**	**2 (-)**	**3 (o)**	**4 (+)**	**5 (+ +)**
30d	Ich neige dazu, in Notfällen die Kontrolle über mich zu verlieren.	**1 (- -)**	**2 (-)**	**3 (o)**	**4 (+)**	**5 (+ +)**
31p	Wenn mir das Verhalten eines anderen komisch vorkommt, versuche ich mich für eine Weile in seine Lage zu versetzen.	**1 (- -)**	**2 (-)**	**3 (o)**	**4 (+)**	**5 (+ +)**
32f	Wenn ich eine interessante Geschichte oder ein gutes Buch lese, versuche ich mir vorzustellen, wie ich mich fühlen würde, wenn mir die Ereignisse des Buches passieren würden.	**1 (- -)**	**2 (-)**	**3 (o)**	**4 (+)**	**5 (+ +)**
33d	Wenn ich jemanden sehen müsste, der dringend Hilfe in einem Notfall bräuchte, würde ich bestimmt zusammenbrechen.	**1 (- -)**	**2 (-)**	**3 (o)**	**4 (+)**	**5 (+ +)**
34p	Bevor ich jemanden kritisiere, versuche ich mir vorzustellen, wie ich mich an seiner Stelle fühlen würde.	**1 (- -)**	**2 (-)**	**3 (o)**	**4 (+)**	**5 (+ +)**
35k	Nicht die Lösung, sondern das Problemlösen an sich ist oft das eigentlich Interessante.	**1 (- -)**	**2 (-)**	**3 (o)**	**4 (+)**	**5 (+ +)**

Saarbrücker Persönlichkeits-Fragebogen (SPF)
Dr. Christoph Paulus
FR Erziehungswissenschaft
66041 Saarbrücken

Auswertungsschlüssel:

Die Items sind mit mit einem Buchstaben gekennzeichnet:

F = fantasy
E = empathy
K = competence (ist für die Empathietestung irrelevant)
P = perspective taking
D = distress

Aus den einzelnen Gruppen können Sie Subscores bilden, z.B.
Empathiefähigkeit = E + P

Die farblich abgesetzten Items müssen bei der Auswertung umkodiert werden.

Quelle und weitere Hinweise:

http://www.uni-saarland.de/fak5/ezw/abteil/motiv/paper/empathie.htm

e-mail des Autors: cpaulus@mx.uni-saarland.de
Homepage: http://www.uni-saarland.de/fak5/ezw/personal/paulus/welcome.htm

B.4 meCUE Questionnaire

VP-Code	Produkt	0

Fragebogen zur Bewertung interaktiver Produkte

Auf den folgenden Seiten finden Sie verschiedene Aussagen, mit deren Hilfe Sie das vorliegende Produkt bewerten können.

Kreuzen Sie bitte für jede Aussage an, wie sehr Sie persönlich finden, dass sie auf das Produkt zutrifft. Es kann sein, dass einige Aussagen nicht so gut zum Produkt passen, kreuzen Sie bitte trotzdem immer eine Antwort an.

Denken Sie nicht zu lange über einzelne Aussagen nach, sondern geben Sie bitte die Einschätzung ab, die Ihnen spontan in den Sinn kommt.

Es gibt keine "richtigen" oder "falschen" Antworten - nur Ihre persönliche Meinung zählt!

VP-Code	Produkt	I.1

	lehne völlig ab	lehne ab	lehne eher ab	weder noch	stimme eher zu	stimme zu	stimme völlig zu
Das Produkt lässt sich einfach benutzen.	○	○	○	○	○	○	○
Das Produkt ist kreativ gestaltet.	○	○	○	○	○	○	○
Die Funktionen des Produkts sind genau richtig für meine Ziele.	○	○	○	○	○	○	○
Das Produkt verleiht mir ein höheres Ansehen.	○	○	○	○	○	○	○
Ohne das Produkt kann ich nicht leben.	○	○	○	○	○	○	○
Das Design wirkt attraktiv.	○	○	○	○	○	○	○
Es wird schnell klar, wie man das Produkt bedienen muss.	○	○	○	○	○	○	○
Durch das Produkt werde ich anders wahrgenommen.	○	○	○	○	○	○	○
Ich halte das Produkt für absolut nützlich.	○	○	○	○	○	○	○
Das Produkt ist wie ein Freund für mich.	○	○	○	○	○	○	○
Das Produkt ist stilvoll.	○	○	○	○	○	○	○
Wenn ich das Produkt verlieren würde, würde für mich eine Welt zusammenbrechen.	○	○	○	○	○	○	○
Die Bedienung des Produkts ist verständlich.	○	○	○	○	○	○	○
Meine Freunde dürfen ruhig neidisch auf das Produkt sein.	○	○	○	○	○	○	○
Mithilfe des Produkts kann ich meine Ziele erreichen.	○	○	○	○	○	○	○

VP-Code	Produkt	II.1

	lehne völlig ab	lehne ab	lehne eher ab	weder noch	stimme eher zu	stimme zu	stimme völlig zu
Das Produkt beschwingt mich.	○	○	○	○	○	○	○
Das Produkt macht mich müde.	○	○	○	○	○	○	○
Das Produkt nervt mich.	○	○	○	○	○	○	○
Das Produkt entspannt mich.	○	○	○	○	○	○	○
Durch das Produkt fühle ich mich erschöpft.	○	○	○	○	○	○	○
Durch das Produkt fühle ich mich ausgeglichen.	○	○	○	○	○	○	○
Das Produkt frustriert mich.	○	○	○	○	○	○	○
Das Produkt stimmt mich euphorisch.	○	○	○	○	○	○	○
Durch das Produkt fühle ich mich passiv.	○	○	○	○	○	○	○
Das Produkt beruhigt mich.	○	○	○	○	○	○	○
Durch das Produkt fühle ich mich fröhlich.	○	○	○	○	○	○	○
Das Produkt verärgert mich.	○	○	○	○	○	○	○

VP-Code	Produkt	III.1

	lehne völlig ab	lehne ab	lehne eher ab	weder noch	stimme eher zu	stimme zu	stimme völlig zu
Wenn ich könnte, würde ich das Produkt täglich nutzen.	○	○	○	○	○	○	○
Ich würde das Produkt gegen kein anderes eintauschen.	○	○	○	○	○	○	○
Ich kann es kaum erwarten, das Produkt erneut zu verwenden.	○	○	○	○	○	○	○
Im Vergleich zu diesem Produkt wirken andere Produkte unvollkommen.	○	○	○	○	○	○	○
Ich würde mir genau dieses Produkt jederzeit (wieder) zulegen.	○	○	○	○	○	○	○
Wenn ich mit dem Produkt zu tun habe, vergesse ich schon mal die Zeit.	○	○	○	○	○	○	○

VP-Code	Produkt	IV.1

Geben Sie bitte abschließend an, wie Sie das Produkt insgesamt bewerten.

schlecht  **gut**

-5 -4 -3 -2 -1 0 1 2 3 4 5

B.5 SUS Questionnaire

	Stimme überhaupt nicht zu	---	---	---	Stimme voll zu
Ich denke, dass ich das Trainingsprogramm nach der Studie gerne häufig benutzen würde.	○	○	○	○	○
Ich fand das Trainingsprogramm unnötig komplex.	○	○	○	○	○
Ich fand das Trainingsprogramm einfach zu benutzen.	○	○	○	○	○
Ich hätte mir die Hilfe einer technisch versierten Person gewünscht, um das Trainingsprogramm benutzen zu können.	○	○	○	○	○
Ich fand, die verschiedenen Funktionen in dem Trainingsprogramm waren gut integriert.	○	○	○	○	○
Ich denke, das Trainingsprogramm enthielt viele Inkonsistenzen.	○	○	○	○	○
Ich kann mir vorstellen, dass die meisten Menschen den Umgang mit dem Trainingsprogramm sehr schnell lernen können.	○	○	○	○	○
Ich fand das Trainingsprogramm sehr umständlich zu nutzen.	○	○	○	○	○
Ich fühlte mich bei der Benutzung des Trainingsprogramms sehr sicher.	○	○	○	○	○
Ich musste eine Menge lernen, bevor ich anfangen konnte, das Trainingsprogramm zu verwenden.	○	○	○	○	○

Angaben senden

Appendix C: Additional Plots

C.1 Additional Boxplots for Analysis of Study A

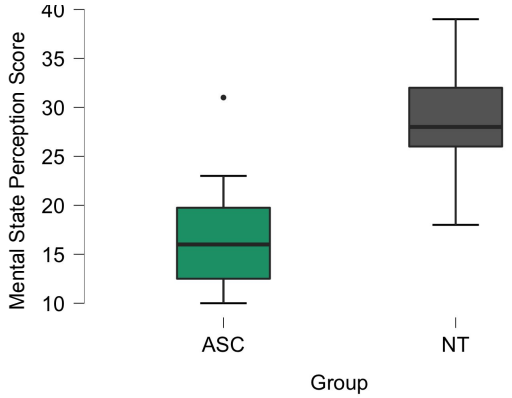
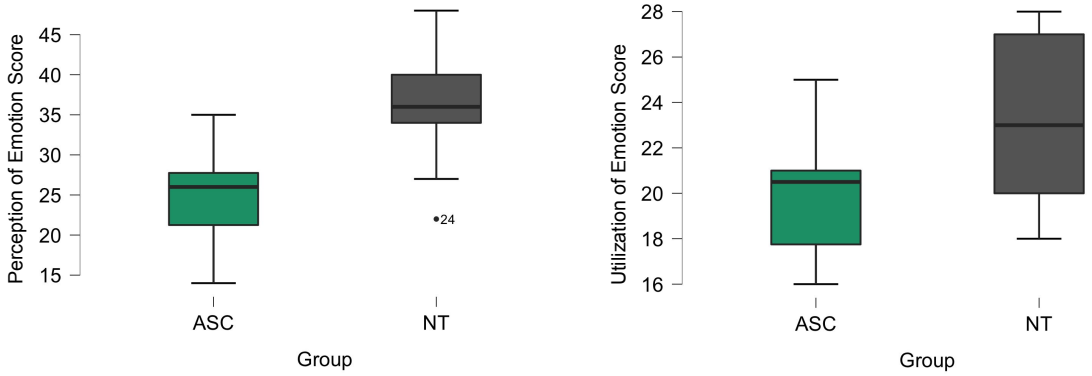


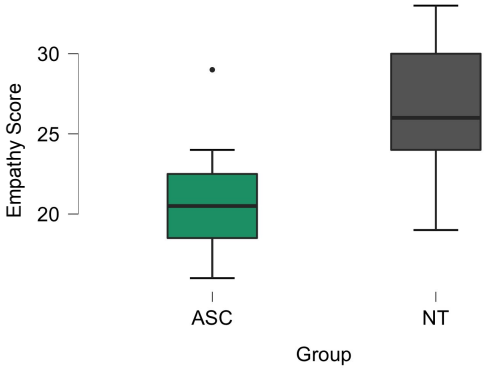
Figure C.1: Mean, minimum and maximum for *mental state perception (MSP)* from CEEQ; ASC and NT



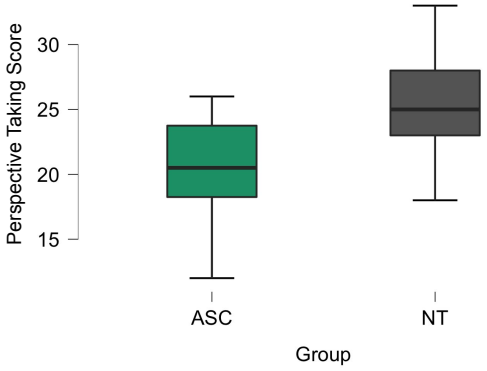
(a) Mean, minimum and maximum for *perception of emotion* from SREIT; ASC and NT

(b) Mean, minimum and maximum for *utilization of emotion* from SREIT; ASC and NT

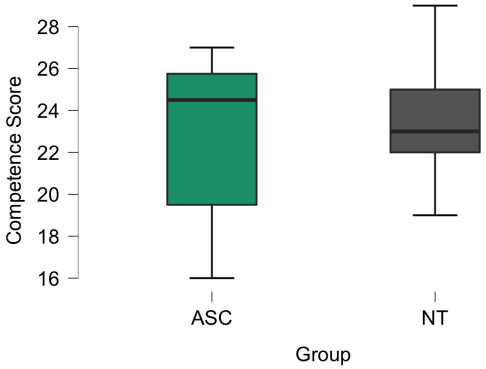
Figure C.2: Additional Boxplots for SREIT



(a) Mean, minimum and maximum for *empathy* from SPF; ASC and NT

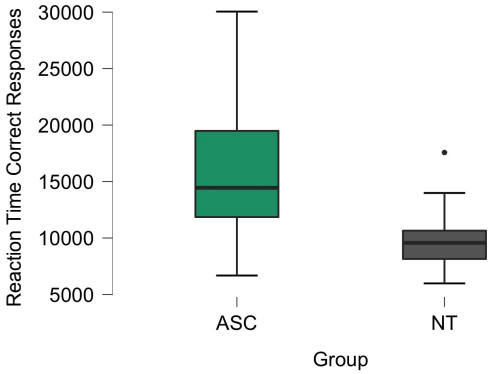


(b) Mean, minimum and maximum for *perspective taking* from SPF; ASC and NT

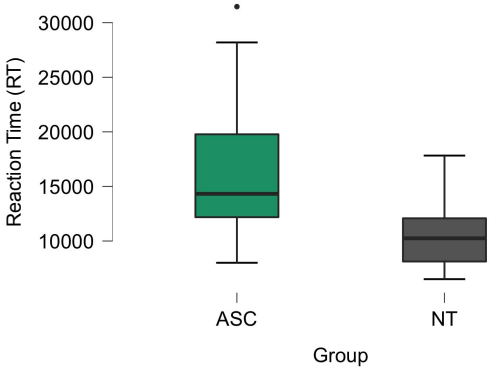


(c) Mean, minimum and maximum for *competence* from SPF; ASC and NT

Figure C.3: Additional Boxplots for SPF



(a) Mean, minimum and maximum reaction times for correct responses in the Face Puzzle task; ASC and NT



(b) Mean, minimum and maximum reaction times in the Face Puzzle task; ASC and NT

Figure C.4: Additional Boxplots for Face Puzzle

C.2 Additional Boxplots for Analysis of Study B

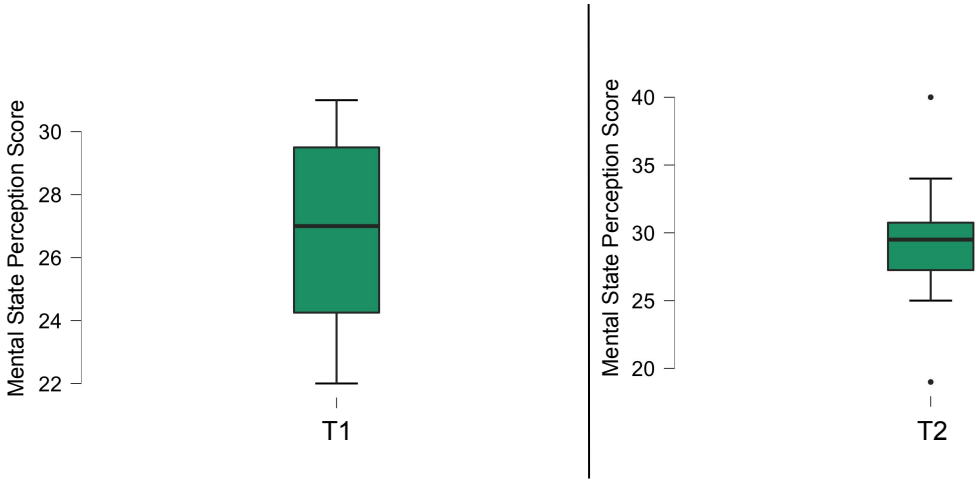


Figure C.5: Mean, minimum and maximum scores for *mental state perception (MSP)*;Pre-post training for NTs

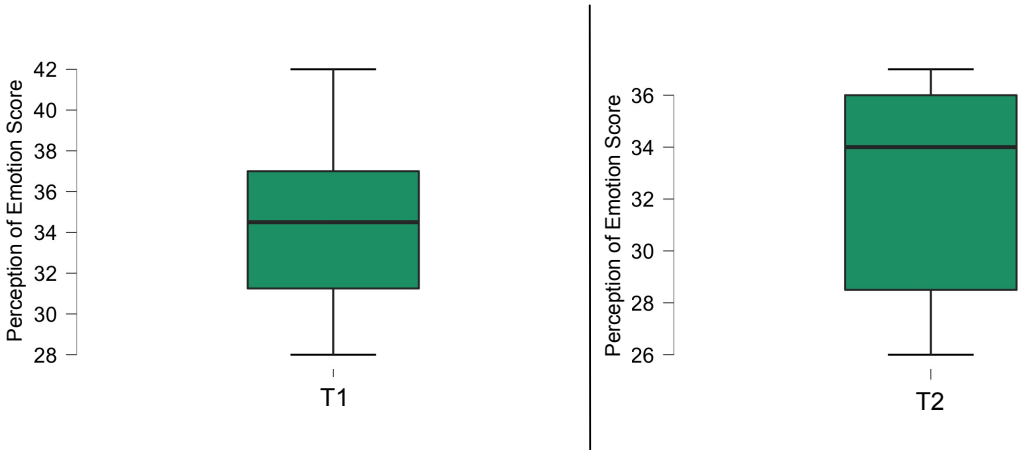


Figure C.6: Mean, minimum, maximum for *perception of emotion* subscale (SREIT); Pre-post training for NTs

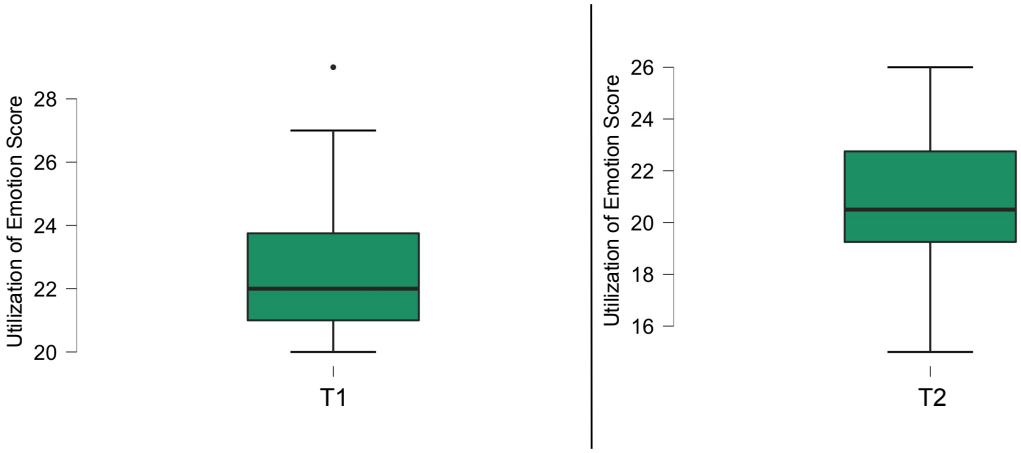


Figure C.7: Mean, minimum, maximum for *utilization of emotion* subscale (SREIT); Pre- post training for NTs

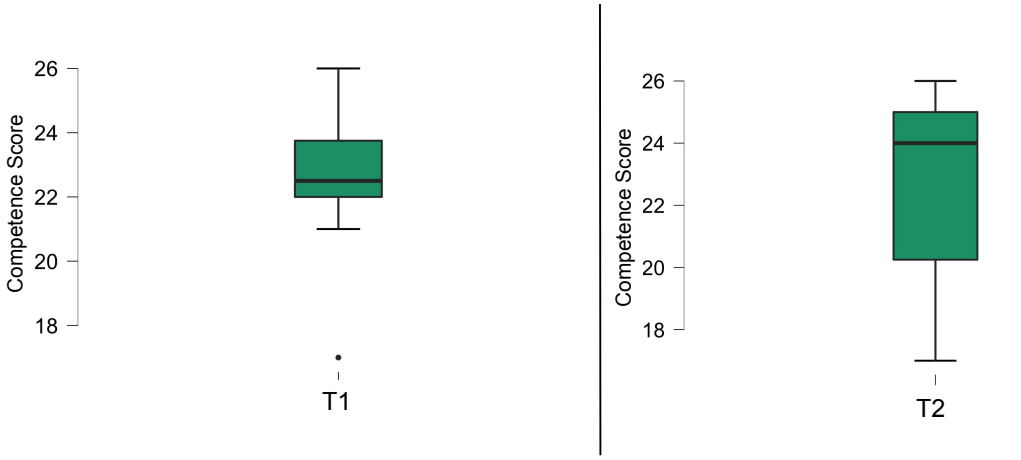


Figure C.8: Mean, minimum, maximum *competence* scores (SPF); pre- post training for NTs

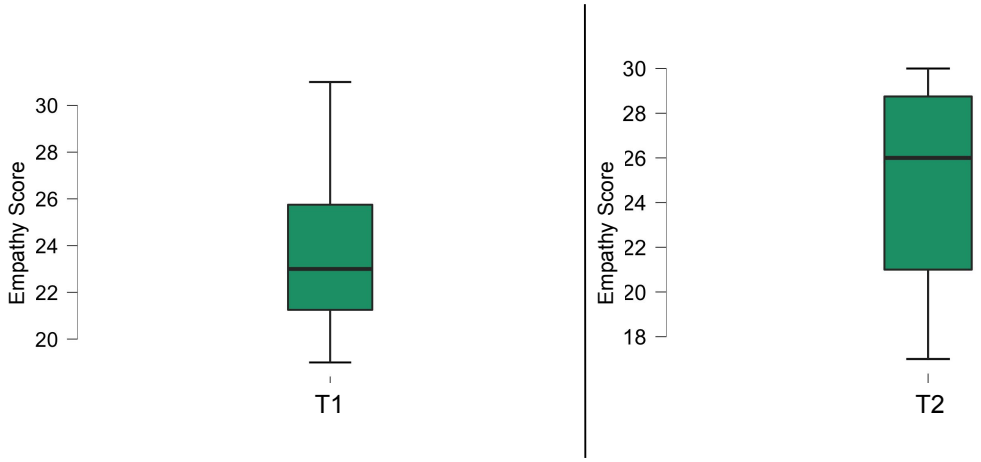


Figure C.9: Mean, minimum, maximum *empathy* scores (SPF); pre- post training for NTs

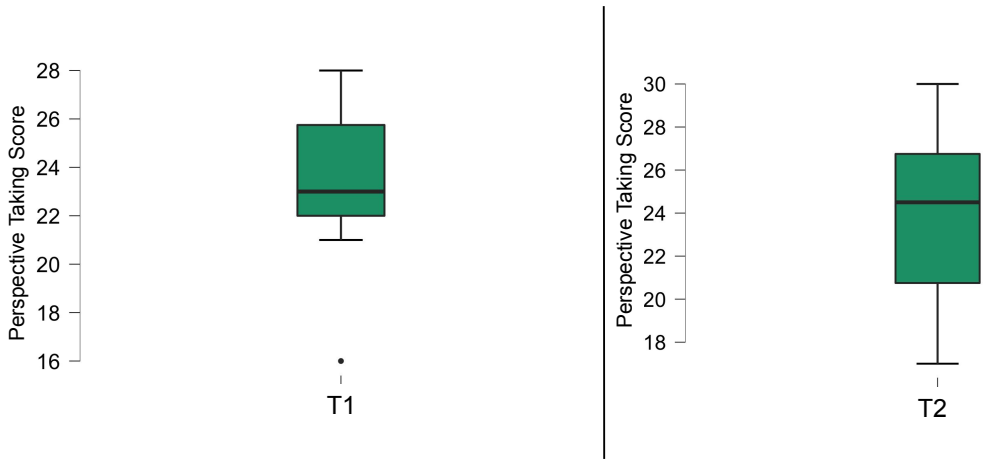


Figure C.10: Mean, minimum, maximum *perspective training* scores (SPF); pre- post training for NTs

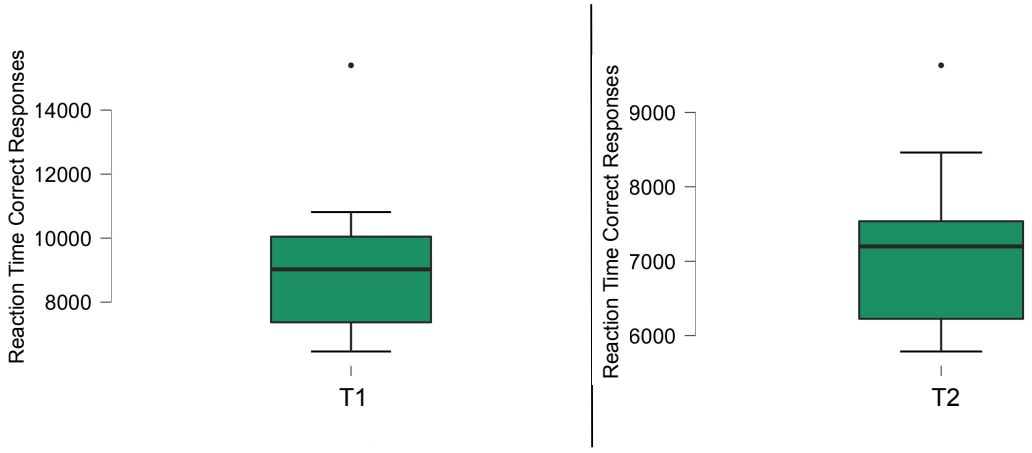


Figure C.11: Boxplots for pre- and post-training; Mean, minimum, maximum for reaction times for correct responses from the Face Puzzle task

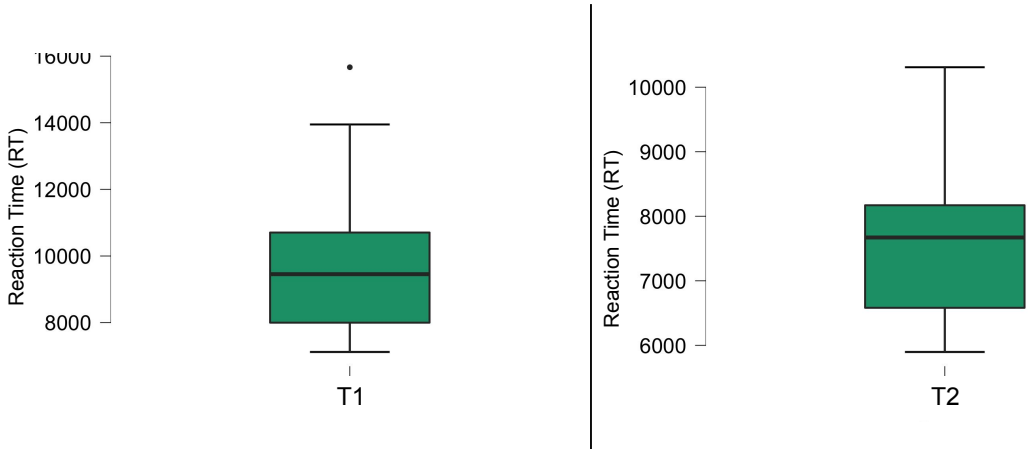


Figure C.12: Mean, minimum, maximum for reaction times (ms) for the Face Puzzle task; pre- post training for NTs

C.3 Response Frequencies for meCUE Subscales

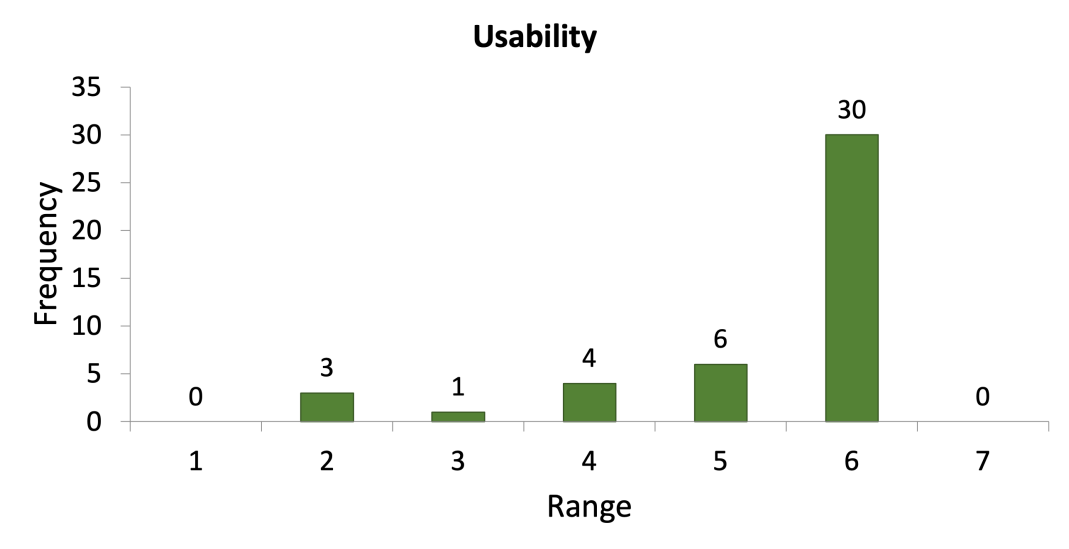


Figure C.13: Response frequencies for *usability*

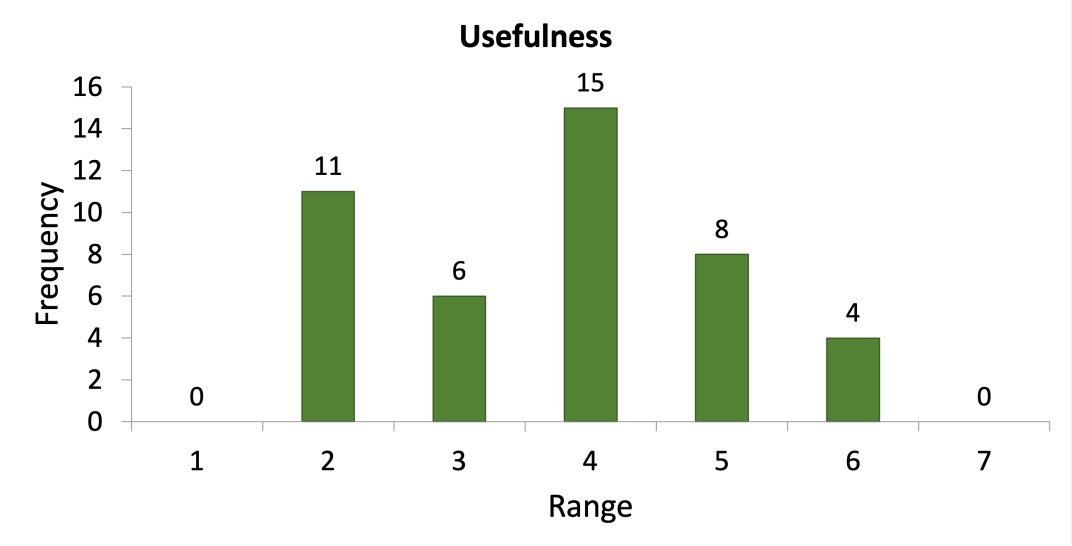


Figure C.14: Response frequencies for *usefulness*

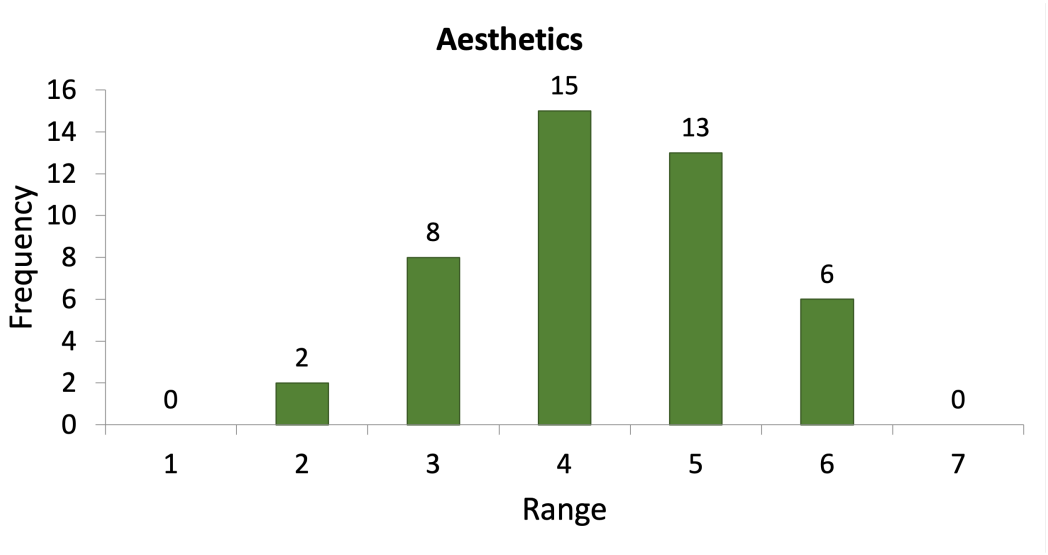


Figure C.15: Response frequencies for *aesthetics*

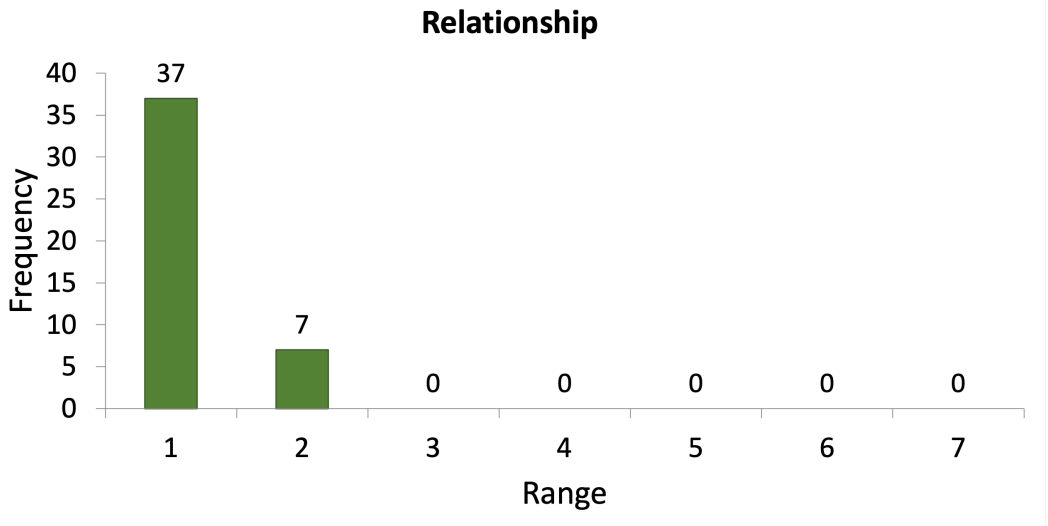


Figure C.16: Response frequencies for *relationship*

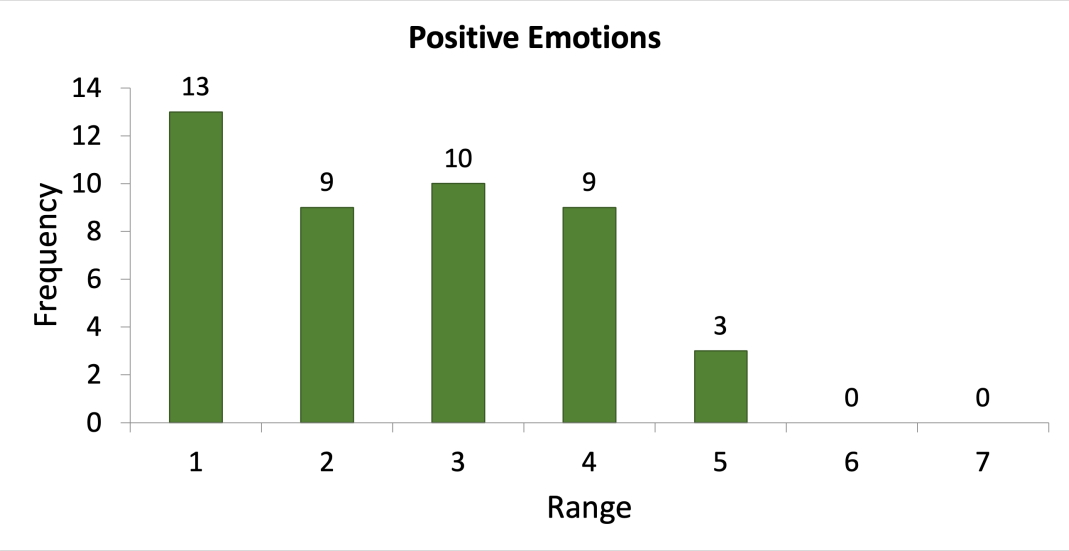


Figure C.17: Response frequencies for *positive emotions*

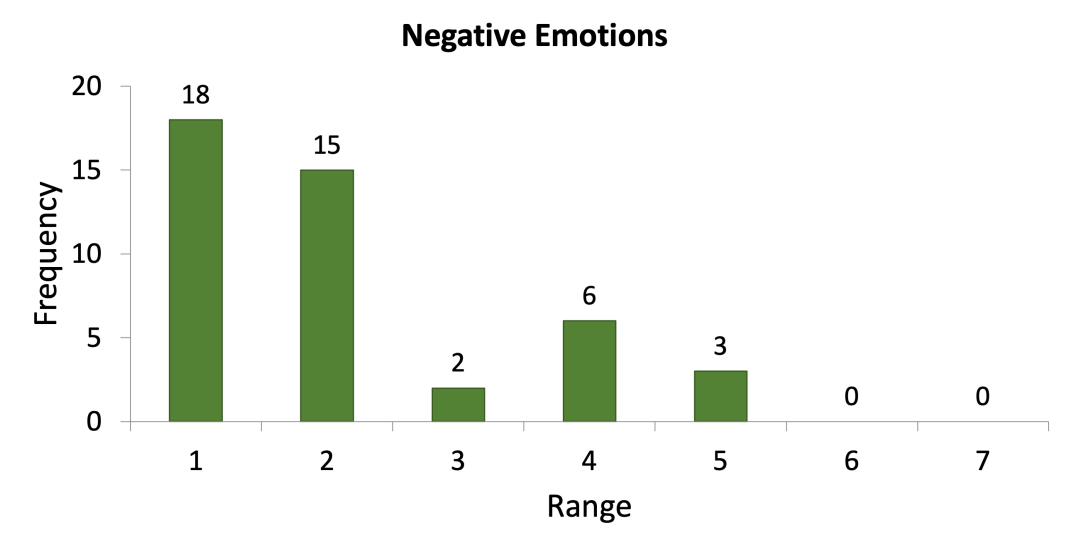


Figure C.18: Response frequencies for *negative emotions*

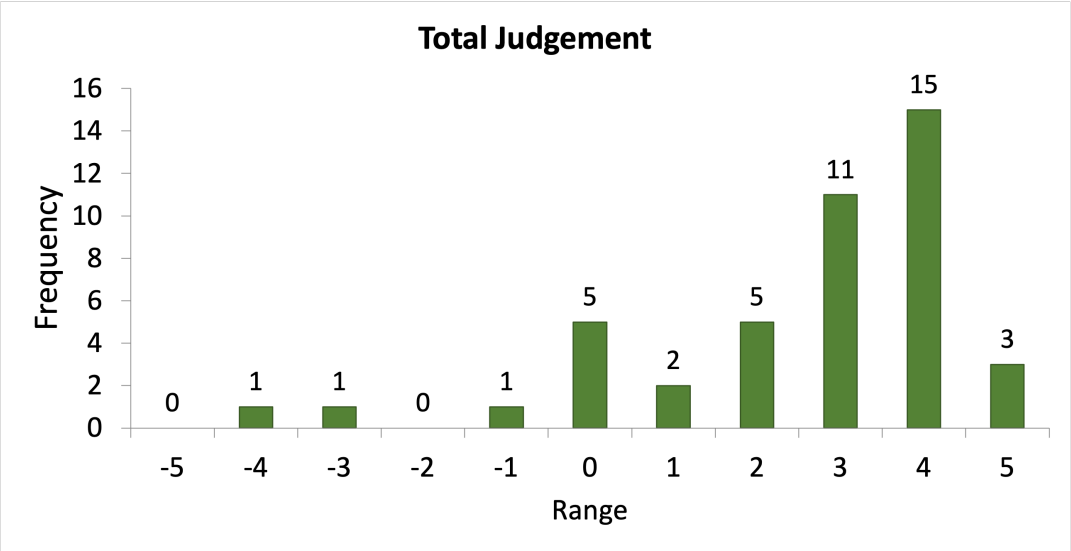


Figure C.19: Response frequencies for *total judgment*