universität wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation /Title of the Doctoral Thesis

„Advancing Assessment and Promotion of Social Cognition and Emotion Understanding in Children“

verfasst von / submitted by

Mag. Jakob Leyrer

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doktor der Naturwissenschaften (Dr.rer.nat.)

Wien, 2019 / Vienna 2019

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 796 605 298
Dissertationsgebiet lt. Studienblatt / field of study as it appears on the student record sheet:	Psychology
Betreut von / Supervisor:	Univ. Prof. PD Dipl.-Psych. Dr. Kristina Hennig-Fast

Abstract

Socio-emotional competence has been identified as a major factor contributing to emotional resilience and mental health. The capability to identify individuals in need of developing these faculties early and to disseminate training tools in a practical and economical fashion could help reduce the global burden of mental health problems. In this light and given that computer games are an integral part of children's lifeworld, computerized interventions can be seen as an extension of the task-shifting approach recently proposed by the World Health Organization. Study one investigates social-cognitive reasoning and mental state talk across the life-span utilizing a new procedure based on cartoon vignettes, the Flexibility and Automaticity of Social Cognition (FASC) to be used in studies two and three. Explaining story character's behaviour flexibly in mental state terms showed an increase from children to adolescents to adults and a decline in older adults. Furthermore the presence of verbal cues and the degree of ambiguity in the vignettes, factors usually neglected in other measures of advanced theory of mind, modulated outcome variables. The FASC is a promising new measure to investigate mentalizing across the life-span. Study two presents the computerized Task of Recognizing and Understanding Emotions (cTRUE), based on the Test of Emotion Comprehension (TEC). The cTRUE total score showed good internal consistency and convergent validity with established measures of emotion understanding. For external emotion understanding and emotion regulation, there was also evidence for cTRUE response time scores to predict academic competence and pro-social role behaviour (respectively) beyond TEC and cTRUE accuracy scores. However, more work needs to be done concerning individual components of emotion understanding which showed great heterogeneity in terms of reliability and validity. Study three describes the development and evaluation of EmoJump, a computer game designed to promote external, mental and reflective emotion understanding in primary school age children. In a randomized controlled trial 12 training sessions of 20-30 minutes resulted in an increased understanding of mixed emotions. Implications for the further development of games to promote emotion understanding are discussed.

Keywords: emotion understanding, social cognition, theory of mind, measurement, promotion, training, computer games, serious games, lifespan development, childhood

To Leonard and Karoline

for teaching me so much about life.

You could imagine sitting on dad’s shoulders to stop being sad

– my son at age 4 to my wive (translated)

Acknowledgments

I want to thank my first supervisor, Priv.-Doz. Mag. Dr. Manuel Sprung, for giving me the opportunity to work and earn experience in a range of areas, such as developing measures and training tools for metacognition, executive functions, theory of mind or emotion understanding. It was a unique learning experience from which I came out a better scholar. I also want to thank my second supervisor, Dipl.-Psych. PD Dr. Kristina Hennig-Fast, for generously taking me in after my first supervisor left university. Assistance by my two lab co-workers, in particular Mag. Gabriela Markova, Ph.D. was greatly appreciated. I would like to thank the - as of this writing yet to be appointed - reviewers of this thesis for their time and the honor to have this work read and evaluated by someone far more experienced than me.

The master students working with me on the projects that were part of this dissertation and who bore the grunt of the “fieldwork” ought to be mentioned individually (in no particular order), as well as important project partners: For the “Flexibility and Automaticity of Social Cognition” project: Kerstin Ganglmayer, Julia Neudorfer, Ann-Kathrin Schmidt. Additional thanks to Elisabeth Hayward for allowing us to use and adapt her new measure and providing assistance. For the “Computerized Test of Emotion Comprehension” project: Hans Haslinger and Stefanie Reichetseder as well as the intern Alexandra Milojevic. I want to warmly thank Francisco Pons, Ph.D. for his encouraging words when I needed them most. For the “EmoJump” project: Sandra Anderl, Tanja Rüscher, Vanessa Zechner, Natascha Schweiger, Katharina Meusburger, Krisztina Halasz, Hans Haslinger, Laura Neumann, Judith Reiss and Theresa Resch. Additional thanks to Univ.-Prof. Dipl.-Ing. Dr. Helmut Hlavacs for supervising the technical realization. At this point I want to extend the acknowledgment to the many other master students that worked with me on several other projects at the games for resilience lab that did not enter this thesis. In computer game terms, you surely did not choose the easy mode to earning your degree.

I would also like to acknowledge the University of Vienna for the opportunities to indulge in my epistemophilia and to take courses from various studies beyond the requirement of a curriculum and in particular Britta Harringer, BA MA from the student service center for being really kind and helpful.

In this pursuit, my (extended) family has been a steady source of support and a safe haven, something invaluable in this world. I am very grateful to my sister Mag. Agnes Leyrer for not flinching once when I asked her to proofread my thesis and for doing so in record time.

Last but not least I want to express my gratitude to my wife for just ‘being’ (and for bearing with me) and for our children which, during this long journey, reminded me of the most important things in life.

Table of contents

List of Tables:

List of Figures:

1. Background

1.1. Children’s use of computers and computer games

Computers and computer games are firmly embedded in children’s and adolescents’ everyday lives in western countries. As early as 1985, there are statistics illustrating children’s interest in computers. In a US-based study about instructional media use in schools, Riccobono (1985) reports that 40 percent of adults did not use their home computer in a typical week compared with only 16-20 percent of children between 6 and 17 years. Today in Austria, almost all adolescents and three in four children have access to a computer (Education Group GmbH., 2015; 2014). Computers and other electronic devices (e.g. smartphones) are not only used for educational or work purposes. A large representative survey in the USA about teen video game use and civic engagement found that 97% of teens aged 12-17 play computer, console, web or portable games (Lenhart et al., 2008). A representative survey of teens in Austria found that three in four teens (11-18 years) play computer and console games (Institut für Jugendkulturforschung, 2008). In a representative survey of adolescents (11-18 years) in Upper Austria, a province in Austria, 61% of youths listed computer games (including console, smartphone, etc.) among their leisure activities. When asked for their computer play time directly, only 20% stated they never play computer or video games while 60% reported to play computer and video games several times a week with an average play time of 72 minute per day (Education Group GmbH., 2015). Seventy-five percent are *rather interested* or *interested* in playing on the computer, console, smartphone or other devices. Computer games are also very commom as a leisure activity with children. In the sister study by the same research firm looking at media use of children (6-10 years) in Upper Austria, 60% listed playing computer games among their leisure activities. Intriguingly this is almost the same percentage as for adolescents (Education Group GmbH., 2014). Fifty-seven percent stated that they play at least once a week. Sixty-nine percent are *rather interested* or *very interested* in computer games. Among all activities exercised on the computer at least once a week, playing computer or console games alone (48%) or with others (44%) was the most frequent, ahead of using it for school (44%), surfing the internet (34%) or listening to music (30%). The same survey asked parents of children between 3 and 10 years about their child’s media usage. Fifty-eight percent listed playing with electronic media (computer, console, smartphone, etc.) among their child’s leisure activities, 55% report that their children play at least once a week and 67% say that their child is *interested* or *very interested* in computer games, showing a high accordance with self-reports of children between 6 and 10 years. These data show that children playing computer games is an everyday phenomenon and it’s not likely going to go away. Therfore, might childrens’ and adolescents’ (including adults’) affinity to playing on an electronic device be harnessed in a therapeutic context? Before we approach existing uses of computers in the field of mental health, I turn to the burden mental health problems pose globally.

1.2. A global perspective: burden of mental health problems

Mental health problems are common, pose a serious challenge to health care and have debilitating effects on the lives of the affected individuals, their families and society as a whole. According to the world mental health surveys of the World Health Organization (WHO) one in three people in the USA suffer from a mental disorder in their lifetime (Kessler et al., 2009). While not leading causes of mortality, mental disorders only reveal their significance for global health when considering the impact on social functioning. A benachmark to estimate this *burden of disease*, the gap between an ideal health situation and the current health status, are the *disability adjusted life years* (DALY). They take into account years of life lost due to premature death as well as years of life lost due to disability (YLD). According to “The Global Burden of Disease: 2004 Update” (WHO, 2008), neuropsychiatric diseases, leading among them depression, are the leading causes of disability and account for approximately one third of YLD among people of 15 years and older. In terms of DALY, mental, neurological and substance-abuse disorders (MNS) account for 14% of life-years. For children and adolescents (aged 5-14) in western Europe, mental and substance use disorders account for 26% of YLD and 23% of DALY with anxiety disorders (AD) conduct disorders (CD) and depressive disorders (DD) taking ranks two, four and six respectively among all diagnoses. More specifically, the proportion of YLD for primary school age children (5-9 years) is 20% (7% anxiety, 6% conduct, 4% autism spectrum, 3% others), for older children (10-14 year) 30% (10% anxiety, 8% conduct, 5% depressive, 7% others) and for adolescents 33% (8% anxiety, 8% depressive, 4% conduct, 3% drug use, 3% bipolar, 7% others) with DALY being only slightly lower due to low morbidity in this age group in high income countries (Institute for Health Metrics and Evaluation, 2017).

However, only a part of the affected people receive treatment ranging from 26% to 60% for mild and severe mental disorders respectively. This problem, called treatment gap, is even more pronounced in lower- to medium-income countries reaching 75% and more (Wang et al., 2007; WHO, 2008b). The causes for this treatment gap are manifold (Kohn et al. 2004). On the part of the individual, the reasons include a lack of problem awareness, the belief that treatment is not effective or that the problem may go away by itself, a lack of knowledge about mental disorders, and fear of stigmatization. On the structural or societal side, there may be limited availability or accessibility. A concern shared between individuals and society is the issue of treatment costs. Yet available treatments could prove cost-effective since mental disorders cause so much impairment in every-day life in the long term.

Mental health promotion and prevention constitutes an important avenue to combat the global burden of disease by reducing both mental and physical illness (WHO, 2002, 2004a, 2004b). In this framework, “a person’s ability to deal with thoughts and feelings […] and emotional resilience” is seen as a central determinant of mental health and mental illness (WHO, 2004a, p.26). In the introduction of study 2 I will return to the topic of emotional competence and will discuss related outcomes in terms of health and social functioning.

In an effort to face the shortage of health services in developing countries, particularly associated with HIV, the World Health Organization (WHO) issued guidelines for an approach called *task-shifting*. The idea is to redistribute tasks among health workforce teams and move specific tasks “where appropriate, from highly-qualified health workers to health workers with less training and fewer qualifications in order to make more efficient use of the available human resources for health” (WHO, 2008a, p.2). This approach has recently been taken up to be applied to the field of mental health (Patel, 2012). In a similar vein, computerized interventions could be seen as lying at the extreme end of this task-shifting approach, where parts of or even whole interventions are delivered by computer programs instead of mental health professionals. Alternatively it could be combined with the original task-shifting idea to support health workers with fewer qualifications to maintain treatment fidelity. Computer and communication technologies could thus play an important role in reducing the treatment gap by making interventions widely available, reducing application costs and overcoming barriers like mental-health worker shortage, geographic isolation or stigmatization. Of course, computers are not ubiquitous (yet), but with the advent of smartphones the number of people with access to devices capable of running computer programs has risen drastically. According to the Q1 2018 Ericsson Mobility Report (http://www.ericsson.com/mobility-report) there were 4.3 billion global smartphone subscriptions at the end of 2017, which accounts for about 60% of mobile phone subscriptions. There are 5.3 billion mobile broadband subscriptions globally and it is forecast that 95% of all subscriptions will be for mobile broadband by the end of 2023. In this context, computerized interventions to improve mental health promotion become a real opportunity to reach the masses.

1.3. The use of computers and telecommunication in (mental) health interventions

The use of computers and telecommunication in mental health and health promotion falls under terms like telemental health, telepsychiatry, e-health or online / internet / computerized or computer-assisted therapy and encompasses a wide range of different technologies and methodologies from the use of synchronous (e.g. chat, video-conference) or asynchronous (e.g. email) telecommunication to therapist-guided computer-assisted therapies to fully autonomous computerized self-help interventions. Telemedicine and telepsychiatry are one of the oldest uses of information technology in the field of health services. It commonly denotes the use of telecommunication technology (e.g. telephone, video-conferencing) to share information in real time between two or more individuals (Myers & Cain, 2008). There is evidence that it is effective for assessment (Hilty, 2013) and psychotherapeutic treatment (Backhaus, et al., 2012) and well perceived by patients in terms of outcome and therapeutic alliance (Jenkins-Guarnieri, Pruitt, Luxton, & Johnson, 2015). Additionally reviews have found it to be cost effective if used beyond the break-even-point of introductory costs (Hyler & Gangure, 2003) and a viable alternative in resource constrained environments (Chipps, Brysiewicz, & Mars, 2012). Although there are few studies directly comparing the effectiveness of telemental health treatment and face-to-face treatment studies, two studies have found it to be only slightly inferior or comparable to face-to-face treatments for depression or eating-disorder (De Las Cuevas, Arredondo, Cabrera, Sulzenbacher, & Meise, 2006; Mitchell, 2008). In respect to the treatment gap, telemental health can help to alleviate the lack of mental health services in rural areas and to address people shying from face-to-face contact. In terms of treating widely prevalent mild mental health problems or even providing mental health promotion for everyone it is not enough to meet the need because the health worker’s time is still taxed one-to-one deducting potential travel times.

The use of computer programs and interactive websites delivering (mental) health treatment, prevention or promotion is a new, emerging field although in cognitive rehabilitation, computers have been used since the mid-eighties (Lynch, 2002). If a certain degree of supportive interaction with a human is maintained, the intervention is termed *guided*. Support can be of technical and/or therapeutic nature and be given remotely (e.g. phone, video-conference) or on site. Interventions can also be standalone computer programs. In respect to psychotherapy, cognitive behavioural therapy (CBT) is spearheading this development and known as computerized cognitive behavioural therapy (cCBT) or internet-based/administered CBT. Elements of CBT have been mainly used in computerized treatments of depression and anxiety. Evidence regarding efficacy is mixed but tends to be favourable when seen in a stepped-care context where computer-assisted therapy is applied in mild to moderate mental health problems or when face-to-face therapy is not available (see chapter 1.2). Other health problems where cCBT has been applied to successfully include pain (Velleman, Stallard, & Richardson, 2010), insomnia (Cheng & Dizon, 2012) and chronic somatic conditions (van Beugen et al., 2014). An early literature-review, conducted for the UK-based NHS R&D Health Technology Assessment (HTA) programme that informs the guidelines of the National Institute for Clinical Excellence (www.nice.org.uk), concluded that there is some evidence for cCBT to be as effective as therapist-led CBT for phobias and panic-disorders and superior to treatment as usual for depression and anxiety (Kaltenthaler, Brazier, de Nigris, Tumur, & Ferriter, 2006). In terms of cost-effectiveness there were substantial uncertainties concerning purchasing costs and likely throughput numbers.

Regarding depression, a recent meta-review (review of reviews) cautiously concluded that the reviewed packages can have positive effects on symptoms of depression but had to exclude many reviews because they did not describe their methodology adequately (Foroushani, Schneider, & Assareh, 2011). A recent meta-analysis of computer-based psychological treatments for adults with depression (roughly pertaining to CBT but including interventions with elements of problem-solving-therapy, schema-therapy, structured writing, mindfulness, etc.) found them to be efficacious and effective for the treatment of depression with an overall moderate treatment effect (Richards & Richardson, 2012). Their analyses also yielded several interesting additional findings. Therapist-supported interventions had better outcomes and retention than unsupported interventions. More surprisingly, studies that used less than eight sessions yielded significantly higher treatment effects than studies with more sessions. Effect sizes did not differ between settings (community vs. primary / secondary) but studies targeting general clinical populations had better outcomes than studies with specific populations. In contrast, two large randomized controlled trials speak against the efficacy and effectiveness of cCBT, at least for depression. The Randomised Evaluation of the Effectiveness and Acceptability of Computerised Therapy (REEACT) trial (Littlewood et al., 2015), a multi-site study within the Health Technology Assessment (HTA) programme compared the effects of cCBT packages *MoodGYM* and *Beating the Blues* to general practitioner care for patients with depression at three time points (4, 12 and 24 months). Across all time points there was no difference between either of the two cCBT packages compared to general practitioner care. Only at 12 months follow-up, *MoodGYM* had a small effect over general practitioner care in depressive symptom reduction. Other measures of mental health and quality of life did not show any benefits of the cCBT packages. Similarly, a randomized controlled trial, comparing a cCBT package (*MoodGYM*) to using a website with general mental health information, did not find evidence for beneficial effects on work- and social-adjustment, depression or anxiety over the website-control (Schneider, 2012). Several factors possibly contributing to these disparate findings come to mind. First, many trials of cCBT packages included in the reviews are conducted by the authors of these treatments, possibly leading to an allegiance bias. Second, despite weekly telephone support calls, uptake of cCBT program use in the REEACT trial was very low (modal number of sessions was only one) which could have limited the treatment effect over all participants. The authors argue that their trial was a pragmatic one that aimed to evaluate cCBT in a primary care context as it is currently offered by the National Health Service in the UK.

For anxiety, a meta-analysis of randomized controlled trials of computer-aided psychotherapy for adults with anxiety disorders, found a large effect size compared to contrast conditions (any non-computerized control) and no difference to face-to-face psychotherapy (Cuijpers et al., 2009). A meta-analysis evaluating cCBT for adults with a diagnosis of major depression or anxiety (panic disorder, social phobia, generalized anxiety disorder) also found a large effect, concluded that it is effective and acceptable in terms of adherence and satisfaction (Andrews, Cuijpers, Craske, McEvoy, & Titov, 2010).

Evidence for young people seems to be more favourable overall. A recent meta-analyses of randomized controlled trials for cCBT for anxiety and depression in children and adolescents found it to be effective (Ebert et al., 2015). Another recent meta-analysis of randomized controlled trials found positive effects for cCBT on symptoms of depression and anxiety for adolescents and young adults but not for children (Pennant et al., 2015). Results for children are only based on two studies with cCBT for anxiety in children though. Evidence for other computerized interventions, e.g. cognitive bias modification, was inconclusive. It should be noted that the two meta-analyses differ in some important aspects. While the second study was more inclusive (general to clinical populations, all types of control conditions) the first only included trials with non-active control conditions and samples with elevated depressive or anxiety scores. A common theme found in the discussions of these reviews is, that quality of the reviewed studies is often low and design-heterogeneity high, making it hard to unequivocally answer the question of effectiveness. The conclusion of these reviews and meta-analyses thus could be that cCBT can be efficacious but evidence for pragmatic effectiveness under realistic conditions has yet to be fully established.

Turning to the field of health promotion, computer-delivered interventions have been shown to lead to behaviour change for tobacco and substance use, sexual behaviour, regular mammography screening, healthy diet, binging/purging episodes and general health maintenance. Evidence for physical activity is mixed (Portnoy, Scott-Sheldon, Johnson, & Carey, 2008; Krebs, 2010). In addition it can impact psychological antecedents of behaviour change like knowledge, attitudes and intentions. It does not seem to be effective for diabetes control, weight loss and weight gain/management (Portnoy, et al. 2008). A systematic review examining online mental health promotion and prevention interventions for youths found evidence for the effectiveness of computerized cognitive behavioural therapy in reducing symptoms of depression and anxiety in samples not fulfilling a diagnosis and tentative evidence for positive effects of module-based online mental health promotion interventions (Clarke, Kuosmanen, & Barry, 2015).

1.4. Therapeutic uses of commercial computer games

Specifically designing computerized tools with the goal in mind to alleviate mental health problems or promote specific skills (chapter 1.3, also see chapter 4.1.4) spiced up with more or less game elements are one thing. Might there also be possible positive effects and uses of commercial computer games (in the sense of “created solely to entertain and sell well”) in a therapeutic context? Before discussing this point I will briefly comment on a topic usually popping up when talking about the influence computer games have on us.

Scientific debate about video games to date has focussed mainly on the presumable effects of violent games on aggression, antisocial behaviour and empathy. I will not discuss the research literature on that topic here because it is not the focus of this thesis and a short excursion cannot adequately address such a delicate matter. I will however add some points for consideration from a meta-scientific viewpoint. The debate about negative effects of computer games is led very emotionally. In contrast to many other scientific debates, it is also carried out on a public and political level. At least since the high-school shooting in Littleton 1999 many politicians and media have been quick to point the finger on violent games as a causal factor for the massacre (Sternheimer, 2007). According to a quick web of science search for the topics video games and violence the first publications emerged in 1991 but the field did not take off until 2004 when the published articles doubled compared to the preceding year. Since 2005, numbers of publication have seen a steady increase each year. The instant inference that violent computer games are to blame, when years of scientific discourse have not come to a definitive conclusion yet, can in part be attributed to a cognitive bias dubbed “what you see is all there is” by Daniel Kahneman (Winerman, 2012) that is explained in terms of a dual-process theory. With incomplete information at hand, a hypothetical automatic but limited information processing path (System 1) tries to construct a coherent explanation which is not necessarily reliable. Other biases and psychological mechanisms like priming or semantic network models (violence in video games is semantically connected to violence in behaviour via the superordinate node violence) likely play a part in this “jumping-to-conclusions”. This phenomenon can also be explained as an effort to fulfill the basic need for control and orientation (Grawe, 2004) where any explanation is better than no explanation. On the other hand, people playing computer games, among them researchers, might feel ostracized and experience reactance leading yet to other cognitive biases and premature discarding of evidence linking video games and agression. Due to ethical reasons, the link between video games and aggression can only be explored correlatively or experimentally via proxies. The former approach cannot give causal answers while for the second approach, questions of validity have been raised. Currently there is a wealth of evidence supporting each stances (e.g. Anderson et al., 2010; Ferguson, 2015), depending on the conceptualization of the research question, outcome measures and interpretation of results, suggesting that the overarching question cannot be answered in a dichotomous manner. With that said I will now turn to the potentially beneficial uses of computer games.

As early as 1991 suggestions were made how commercial computer games could be used as an adjunct in psychotherapy with children for the purpose of building relationship, assessing and improving competences (Gardner, 1991; Griffiths, 2003; Ceranoglu, 2010; Steadman, Boska, Lee, Lim, & Nichols, 2014). One of the more obvious uses is as an aid to build a relationship with the client through a shared experience grounded in the child’s or adolescent’s lifeworld. Another area of use is the opportunity to observe behaviours and proxies for cognitive and affective processes like problem-solving, memory, cooperation, frustration tolerance, emotional reactivity or emotion regulation. More specifically, the strong attention focusing effect can be used as a distractor in pain management. Video-game-time can also be simply used as a reward or token for desirable behaviour. Probably the potential benefit of computer games most focused on is that of cognitive training. According to a recent meta-analysis, playing video games improves information processing (Powers, Brooks, Aldrich, Palladino, & Alfieri, 2013). There is evidence for positive effects on visuospatial processing, attention, cognitive control and flexibility while effects on short-term memory are inconclusive (Bisoglio, Michaels, Mervis, & Ashinoff, 2014). There is a need, however, for more rigorous experimental designs, a formal classification of video games and taking into account individual differences. Another potential benefit of computer games is the cultivation of a persistent, optimistic motivational style in the gamer through the development of an incremental theory of intelligence (see Dweck & Molden, 2005). According to this subjective theory, intelligence is malleable and can be developed through effort and time (Granic, Lobel, & Engels, 2014). Likely, this is more of a long-term effect however and thus can not be easily utilized in a therapy. Nevertheless, persistence and success in a game can be addressed in therapy to strengthen a client's self-esteem and build relationship. With the spread of multiplayer-games the question of social benefits arises. Evidence from correlational, longitudinal and experimental studies suggests that playing prosocial video games is related to and predicts prosocial behaviours (Gentile et al., 2009). Interestingly, also violent video games seem to promote prosocial behaviours if played cooperatively compared to competitively (Ewoldsen et al., 2012).

In psychological treatments or psychotherapy, emotions are often deliberately activated to clarify the significance of an experience or to put new coping strategies to the test. Although research of emotions in video games is scarce, a wealth of internet discussion threads about "most emotional moments in video games" pay tribute to the fact that computer games are not only about feelings like frustration, aggression or reward. Recently, the first awards for emotional computer games were given with the goal to reward the quality of emotions elicited by them (http://www.emotionalgamesawards.com). According to a survey of 535 gamers by a private research firm (https://www.bowenresearch.com/studies.php), half of participants agreed that emotions are somewhat or extremely important in computer games. Two thirds think that games could equal, go beyond, or are already beyond books, movies or music in the ability to inspire emotions. Emotional states elicited playing video games were diverse and ranged from high arousal states like competitiveness, violence/excitement or frustration to more positive feelings of accomplishment, delight, awe and wonder to social emotions like compassion for others and love to sadness. What can make the emotional impact of computer games more powerful and more reflective than that of movies, in my eyes, is that – similar to written stories where you can put aside the book and ponder on that last paragraph – you can often decide the pace to advance with to the next scene. For example after going through an emotional scene where someone, you are emotionally attached to, leaves, you may stand in a now empty room and you decide on your own accord when to exit that room and continue with the story. But maybe even more strongly than it is the case with a book, you are still *present* within that empty room that reverberates with what just happened (e.g. the overthrown chair, the amulet thrown on the floor, etc.).

This emotion activating quality of certain games could be used with children and adolescents who have difficulties opening up to talk about emotions in a less threatening context and to build trust for them to be able to talk about their problems associated with the same feeling. I want to close this chapter with two quotes that demonstrate nicely, how video games can elicit a wide range of feelings and emotional states. The first is the review about a short indie-game by a tech reporter of the news site BuzzFeed: “In order, here are the emotions I experienced during the three minutes it took me to play The Plan, the new free morsel from the Norweigan indie developer Krillbite: confusion, frustration, boredom, fear, amusement, delight, joy, enchantment, and regret.” (Bernstein, 2013).

The second is an excerpt from an online-post on the community forum of Telltale Games, a video game developer which became known for its episodic adventure games that feature hard moral choices. The post is describing the experience of someone playing Game of Thrones, a game based on the A Song of Ice and Fire fantasy novel series by George R. R. Martin and the hugely successful TV adaption.

> And i started to love the story. And I was SO attached to all the characters, i cared about them and was so worried when someone died. Ending of the 5th episode? It was predictable, but after all, i cried like i never did during TWD. And moments, when your characters show their true strange and determination like Rodrik did few times? Man, that was trully amazing... It's like you're in this story together with them. (brbsmoking, 2015).

1.5. Potential roles for computerized interventions and games designed to promote socio-emotional skills

In the following section I want to briefly discuss different settings and goals for computer-assisted trainings of emotion understanding or social-cognitive understanding and the potential role of variables like the use of game-elements and the autonomy with which they are applied. First I want to introduce a formal definition of *games* to start from a common ground for evaluating the degree of game-likeness of computerized interventions:

“A game is a system in which players engage in an artificial conflict, defined by rules, that results in a quantifiable outcome” (Salen & Zimmerman, 2003, p80). This definition contains many elements also found in other definitions of what a game or play is (e.g. Caillois, 1961; Suits, 1967; Avedon & Sutton-Smith, 1971). *Conflict* and a *quantifiable outcome* implies that players strife toward a goal, are impeded by obstacles and know when and how well they have achieved their goal. *Rules* govern the set of actions the player can perform to achieve the goal and are themselves often part of the obstacles to pose a challenge. *System* means that there is a set of separable parts that interact and can be directly or indirectly influenced by the player to achieve his goal. *Artificial* refers to the fact that games take place in a confined time and space separate to “real life”. Many authors also include the voluntary nature of games in their definitions. Although the definition by Salen & Zimmerman (2003) chosen here does not contain this attribute for taxonomic reasons it is obvious that in the context of games for health or training, participation has to be voluntary. But is it really fully voluntarily in every case? Volition has been subject to philosophical, theological and scientific debate for centuries. The degree of voluntariness in using a training game is gradual and depending on context. While rarely anybody is forced to actively participate in an intervention or health promotion program, the authority and expectations of a health care professional or teacher recommending it or the social dynamic in a classroom will make the decision to participate a not entirely voluntary one most of the time. The question of how the degree of "gamification" corresponds to efficacy is not clear and probably depends on many factors (e.g. motivation). Certainly designing an intervention as a computer game puts restraints on how this intervention can look like while the ability targeted on the other hand restricts certain game-design choices. So it seems reasonable to assume that certain intervention/training targets lend themselves more easily to gamification than others. For example, training a rather narrowly defined cognitive function like set-shifting is probably easier to implement in a traditional game design than a complex competence like taking turns in a conversation and aligning with the interlocutor. Another open question not discussed here is whether certain game-types (e.g. action, strategy) are more suitable for specific target abilities. As mentioned in chapter 1.3 the level of autonomy or automation of computer-assisted interventions can vary greatly. They range from unguided, fully standalone software programs to guided online therapy programs supported by a health care professional to the use of computers as an adjunct in a traditional one-on-one treatment setting. A central aspect underlying the variables game-likeness and automation is motivation. As much as any intervention may benefit from a high level of incentive, in a therapeutic setting (ambulant or even stationary) structural factors may in part replace incentives or modulate motivation. One can imagine for example, that in general, the motivational situation for a patient in a stationary clinic with a high degree of suffering, that experiences the computerized intervention as part of the overall professional treatment and no opportunity to play commercial games in the meantime is very different to a person that may have only got an online invitation to a computer game through a health promotion campaign. Similarly, a therapist using a computerized intervention may itself constitute a source of incentive. Not least the hope of reducing suffering from symptoms or the goal of being better able to participate in social interactions can be a strong incentive. Therefore the question of why or to what end does one develop and respectively use a computerized tool for training or intervention is crucial. I will now shortly address different possible scenarios and goals.

First, promotion of emotion understanding and social-cognitive understanding might be deployed as a general health promotion or primary prevention campaign unspecifically. This context and goal requires interventions to be fully autonomous computer programs, freely available over the internet. For maximum dissemination in the general population, these have to be highly motivating to be played for the sake of playing which is probably best achieved through a full-fledged commercial-grade game. In a quasy secondary prevention setting, at-risk populations found in schools or youth centres in socio-economically disadvantaged districts might be targeted. Computerized trainings will likely benefit from game-elements but do not have to be triple-A-quality if embedded in a school project due to extrinsic motivational factors. Alternatively more specifically mediated through social workers and the health care system, children of parents with mental health problems could be targeted. Interventions can also function as an adjunct or adjuvant for populations primarily being treated for mental health problems. Training understanding of others' mental states can be part of a psychological treatment protocol. A special - albeit hypothetical - application could be to promote a child's metacognitive knowledge / knowledge about thinking in preparation to be better able to participate in a cognitive therapy. Aside from training abilities, serious games could encourage involvement with a topic (e.g. how people perceive a situation differently) that can then be discussed with a therapist. Neurological, and more recently also psychiatric, rehabilitation is a field where computerized trainings for cognitive remediation hold a firm place, although programs for emotional or social-cognitive competences are less common according to my own working experience. Most existing programs focus on the important, but narrow aspect of facial affect recognition as I will elaborate in chapter 4.1.4. In these settings, training programs do not necessarily have to be game-like since they are usually implemented in a structured health care environment, although motivational factors do play a role.

In chapter 4, I will return to this topic and review existing computerized training tools for socio-emotional skills before introducing EmoJump, a new computer game designed to promote emotion understanding.

2. Study 1: Flexibility and Automaticity across the Lifespan

2.1. Introduction

2.1.1. Theory of mind and social cognition.

Theory of mind (ToM) was originally introduced as the ability to impute (not directly observable) mental states to oneself and other individuals (Premack & Woodruff, 1978). The term theory was used by these authors because these mental states are “not directly observable” and “can be used to make predictions, specifically about the behaviour of other organisms” (p. 515). A word on wording: many terms have been used to describe understanding of mental states. Among them are theory of mind, mentalizing, mindreading, folk- or naïve psychology. By virtue of prevalence and my academic upbringing I will mostly stick to the term theory of mind but will also use the terms mentalizing and mindreading interchangeably. Bear in mind however that the term theory of mind refers to the scientific subject independently of theoretical positions and should not be seen as an alignment to theory-theory (described later) on my side. Since Premack & Woodruff’s seminal paper and the introduction of Wimmer & Perner’s (1983) false-belief task which has – for good and for worse - been the power-horse of ToM research, focus has been on the preschool years. Yet by then, research dealing with aspects of ToM had already been going on for a while, albeit under different headings like perspective taking (Piaget & Inhelder, 1956), metacognition (Kreutzer, Leonard, & Flavell, 1975), or social attribution (Heider, 1956) and mostly focused on older children, adolescents and adults (see Miller, 2012 for a good review). Research on older age groups has never died out but has again risen to prominence as answers to central questions like when explicit ToM first emerges have found broad consensus (Wellman, Cross, & Watson, 2001). This has shifted the research focus to questions like what happens after initial (explicit) false-belief understanding is in place. As an entry point, I will introduce the classic false belief tasks before turning to theories of theory of mind, precursors and the development of theory of mind more broadly.

In the original unexpected location false belief task (Wimmer & Perner, 1983) subjects are confronted with following scenario which is usually enacted with puppets: Maxi puts a chocolate into a cupboard. While he is away, his mother puts the chocolate into another cupboard. The subject has to say or point where Maxi will look for his chocolate on his return. Most children younger than four years insist that Maxi will look for the chocolate in the new place, failing to acknowledge that Maxi’s knowledge is different from theirs. This task has been modified many times since but the core tenet has stayed the same. The second standard false belief task, dubbed unexpected content, deals with appearance reality distinction (Hogrefe, Wimmer, & Perner, 1986; Gopnik & Astington, 1988). The subject sees a container (e.g. sweets box) and is asked what he/she thinks is in the box. Afterwards the container is opened to reveal something unexpected (e.g. matches or pencils). The subject is asked what she thought was in the box or what she thinks another person will think is in the box. In a large meta-analysis Wellman, Cross, & Watson (2001) concluded that most 5-year-olds succeed while most 3-year-olds fail on the two tasks with 4-year-olds showing mixed performance. They further concluded a range of variations like task type, type of question posed in the content task (about self/other belief), nature of the protagonist or object, have no effect on difficulty. Five variables did have an effect but did not interact with age, thus leaving the developmental trajectory unchanged. Small facilitative effects were found for following variations: the protagonist's motivation is deception, the child actively changes the location of the item, salience of the protagonist's belief (e.g. belief is stated), whether the object is present or not (e.g. the chocolate was eaten) and country (e.g. Australia > US > Japan). If prevalence of use is any measure these two tasks have to be considered *the* proxy measures for theory of mind but their interpretational scope has not remained without criticism (Bloom & German, 2000).

2.1.1.1. Theories.

In this chapter I will shortly outline the three major theoretical positions on the nature of theory of mind: theory-theory, simulation theory and modular theories. This outline rests mainly on the review by Doherty 2009 and the article about theory of mind in the Internet Encyclopaedia of Philosophy (Marraffa, n.d.), complemented by primary literature study. These three positions constitute rather broad orientations under which specific theories (e.g. different propositions for simulation) fall. In the scope of this work it should suffice to say that all three positions have received much support and it is likely that they all describe different aspects of the complex phenomenon coined theory of mind. Astonishingly experimental studies have managed to generate a wealth of interesting insights without (at least explicitly) aligning themselves with either theoretical position. Meanwhile cognitive-neuroscientific work has tried to sketch out theory of mind networks while largely remaining agnostic about psychological or philosophical theories. It has been argued that integrating both avenues might prove fruitful (Mahy, Moses, & Pfeifer, 2014) as might dropping the old psychological theories altogether and formulating a new theory better suited to generate hypotheses for neuroimaging studies (Apperly, 2008). At the end of this chapter I will briefly present Apperly's two-system account.

According to *theory-theory* children acquire an understanding of minds through the gradual development of an informal theory (folk psychology) about mental states and processes (Wellman, 2014). Theory-theory has arguably been the most prevalent view in philosophical and psychological inquiry of peoples capacity to understand their own and others minds. Already in the mid of the last century, Sellar (1956) speculated that our understanding of mental phenomena results from a folk theory of mind. Heider (1958) further lay groundwork for the development of theory-theory, a term which was finally coined by Morton (1980). Some authors, following Piaget, have proposed that this process is not unlike scientific theory-change (Gopnik & Wellman, 1992). It is domain-specific (occupied with representational mental states) and establishes cause and effect relations to predict and explain other people's behaviour. The interplay between theory and data (experience) is critical in this account. Theories about the mind are constructed and revised on the basis of experience. Different levels of theories exist: a framework theory governs general principles about behaviour and underlying mental states. Embedded in and constrained by this framework theory more specific theories are used for specific causal predictions in every-day life. Gopnik and Wellman (2012) proposed how hierarchical Bayesian learning could explain development of these layered theories where direct evidence changes specific theories which in turn can themselves act as evidence and change higher-order framework theories. According to developmental constructivists naïve framework theories also exist for domains such as biology or physics. Different stages regarding representational development can be distinguished. During the first year of life, primary representations about the environment (e.g. objects) are formed. During the second year of life, representations of non-real situations (secondary representations) emerge. They can represent the past, desires or hypothetical situations (e.g. pretence). At the age of about 3 or 4 years children start to form meta-representations and become aware that propositions and their truth can be evaluated independently by different people (Doherty, 2009). Major theorists who have developed accounts of theory-theory are Josef Perner, Alison Gopnik and Henry Wellman (Perner, 1991; Wellman, 1990; Gopnik & Wellman, 1994, 2012; Wellman, 2014). Theory-theory-accounts differ in how they treat first-person, in relation to third-person, mental attributions (Marraffa, n.d.). Most proponents of theory-theory believe that first- and third person mindreading exploit the same external behavioural and situational information that is interpreted (i.e. the "outside access" view / the symmetrical account of self-knowledge). While evidence from social psychology and the study of normal and pathological developmental is mostly interpreted in support of the symmetrical view, Nichols & Stitch (2003) have put forward an asymmetrical account assuming that first-person detection and reasoning of mental states is subserved by different mechanisms (called monitoring mechanism) than third-person mindreading.

Modularity accounts, sometimes subsumed as a variant of the theory-theory account, propose an innate mental module responsible for mentalizing. Fodor (1983) posits that mental modules hold certain properties of which informational encapsulation and domain specificity are most essential. Informational encapsulation (and the related limited central accessibility) refers to the module processing informational input on its own without being able to access other information stored outside and without being accessed by other parts of the system (e.g. higher-order cognitive processes) to influence the output. Other properties that, according to Fodor, *most* modules possess are: fast and efficient processing, mandatory operation (e.g. one cannot *not* attribute mental states), shallow outputs (i.e. restriction to low-level concepts), fixed neural architecture, characteristic ontogenetic sequencing and pace and specific breakdown patterns if the module is damaged. Modular accounts of theory of mind have often used less strict conceptions of mental modules. Two noted accounts are those of Leslie and Baron-Cohen which are closely intertwined with the study of autism. According to Leslie (1987) a *theory of mind mechanism* (ToMM) computes metarepresentations about psychological states on an unconscious, sub-personal level and matures during the second year of life. In response to the objection that such young children do not solve the classic false-belief task he posited that performance is limited by executive demands, namely inhibitory control. In later developments of the theory, he introduced a non-modular executive *selection processor* which may be domain general or specific to theory of mind (Leslie & Thaiss, 1992; Leslie, Friedman, & German, 2004). The selection processor's purpose is to select among belief descriptions and it works in tandem with the ToMM. In the case of conflicting descriptions of beliefs the true belief is set as a default and has to be inhibited by the selection process. This selection process needs time to develop which is supposed to explain why children under 4 years consistently fail false-belief tasks. Baron-Cohen in contrast proposes several modules working together with a final *theory of mind module* to accomplish mindreading (Baron-Cohen, 1995). The strongest support in favour of the modularity approach comes from research on the autism spectrum disorder which suggests a strong biological basis and independence of other cognitive abilities (Miller, 2012, p. 38). One argument brought forth in support of the ToMM was a dissociation of performance on the false-photograph task (Zaitich, 1990) between typically developing children and children on the autism spectrum. This task was devised to parallel the false-belief task without relying on mental concepts. Children are familiarized with a Polaroid camera that ejects a photo instantly which takes some time to develop. A cartoon character takes a photo of another cartoon character lying on a mat in the sun and both go away. Another cartoon character lies down on the mat. Children are then asked who on the photograph was lying on the mat. Three-year but not 4-year-olds had problems solving this task as the false belief task. In contrast, older autistic children are performing rather well on the false-photograph task while failing the false-belief task. Leslie & Theiss (1992) presume that while in typically developing 3-year-olds the ToMM works well, performance is hindered by the later developing selective processor. On the other hand, older autistic children possess an intact selective processor but a deficient ToMM. Subsequent analysis and modification of the false-photograph task found that solving it does not require understanding a misrepresentation and problems arise from referential confusion since modifications clarifying the test questions led to younger children being able to solve it compared to the false-belief task (Slaughter, 1998). Other evidence that has been interpreted in favour of the ToMM are looking time differences in 15-month-olds in a violation of expectation task modelled after the unexpected location false-belief task (Onishi & Baillargeon, 2005) which however can also be explained in a non-mentalistic fashion (Ruffman & Perner, 2005 and Perner & Ruffman, 2005). There is also convincing evidence that speaks against a strong modular account. The first comes from studies on deaf children born to hearing parents who have a markedly delay in theory of mind development of about seven years that is not shown in deaf children born to deaf parents who communicate with their parents in sign language from early on (Wellman, 2014, p.102). The second kind of evidence comes from a large representative behavioural-genetic twin study which found only 15% of variation explained by common genetic influences (Hughes et al., 2005).

In *simulation theory*, the key to understanding and predicting others' mental states is the working model of our own mental states. We simulate what we would see, how we would feel, etc. if we were in the others' situation. While antecedents of this idea can be found in philosophers like David Hume, Adam Smith, Immanuel Kant, Arthur Schopenhauer, Friedrich Nietzsche, W. V. Quine Wilhelm Dilthey, Theodor Lipp and Willard Van Orman Quine (Goldman, 2005; Barlassina & Gordon, 2017), it was transferred to the scientific subject of mindreading in 1986 by philosophers Jane Heal and Robert Gordon and further developed by Alvin Goldman (1989). Theorists differ in the role they assign introspective processes as a way of direct access to mental states or products of simulation however. While Goldman (1993) is a proponent of this view (attenuated in his 2006 account), Gordon (1995, 1996) more radically claimed that simulation amounts to "becoming" the mindreading target utilizing nonconceptual representations, therefore eliminating the need for analogue inferential processes. Low-level aspects of mindreading through simulation (e.g. understanding intentions) has also been linked to the mirror neuron system originally implicated in action recognition (Gallese & Goldman, 1998; Iacoboni et al., 2005).

In developmental psychology the idea of mental simulation was taken up by Harris (1992), who proposed that there are two types of default settings that have to be overridden in theory of mind and related concepts (e.g. pretence). The first is the state of reality as perceived or believed by the agent (e.g. the box contains an apple). The second consists of the agent's mental states (e.g. desires apples). The more of these default settings have to be overridden, the more difficult the task becomes. Pretence for example only requires manipulation of the default settings of reality while acknowledging diverse desires only requires manipulation of the default setting of one's own mental state. Understanding others' false beliefs on the other hand requires manipulation of both, one's own believe and the known state of reality. Developmental trajectory of pretence, understanding of diverse beliefs or desires and understanding of false beliefs is in accord with this theory. On the other hand, Wellman (2014, p. 158) argues, that according to simulation theory attribution of false beliefs should be equally difficult as attribution of false desires but the former proves to be more difficult. Over time the gap between simulation theory and theory-theory has lessened and many theorists now propose that both elements play a role in mentalizing (Doherty, 2009). Hybrid accounts integrating simulation theory and theory-theory to a varying degree have been developed for example by Nichols & Stich (2003), Perner & Kühberger (2006) and Goldman (2006).

Apperly and Butterfill (2009; Apperly, 2011) proposed a two systems account in an effort to reconcile the conflicting findings of early competence in infancy and the comparatively late passing of false-belief tasks at around four years, which according to them cannot be explained sufficiently by existing theories like the modular approach or theory-theory. According to this account, a theory of mind module in the strict sense of modularity is – per definition – not able of solving complex instances of mindreading, encountered in real life which are defined by incomplete information and an infinite number of potential solutions (he uses the example of a jury trial). These instances warrant a flexible, informationally non-encapsulated system which in turn is slower, effortful and highly dependent on language. In his view there also exists an early, relatively automatic and fast system explaining the findings of early competence in infancy. This system can track perceptions, goals, and beliefs without representing propositional attitudes and forms a distinct set of concepts distinct to the later emerging second system. The primitive first system however continues to operate in adulthood. He draws parallels to number cognition where there is good evidence for a fast, automatic system for immediate recognition of small sets and a later developing elaborate system for number processing (Feigenson, Dehaene, & Spelke, 2004). Additionally parts of the domain general flexible system can become modularized with experience. This theory has garnered support (Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010; Schneider, Bayliss, Becker, & Dux, 2012; Surtees, Butterfill, & Apperly, 2012) but has also been opposed (Carruthers, 2016). A recent study of a functional double dissociation between low-level perceptive based and high-level inference based mentalizing in white fibre disconnections also hints at two streams of theory of mind although the authors used other tasks than Apperly and also did not relate their findings to his two systems account (Herbet et al., 2014). Accuracy on the Reading the Mind in the Eyes task (low-level mentalizing) was mainly associated with degree of disconnection in the arcuate fasciculus while accuracy on the comic strip task (high-level mentalizing) was associated with the degree of disconnection in the cingulum.

2.1.1.2. Early forms & antecedents of ToM.

Understanding of minds does not start with the passing of the false-belief task. Some aspects of understanding are present much earlier, in the second year of life or even earlier. I will now briefly review aspects of infant's social cognition often considered precursors (or early instances, depending on the theoretical position) of theory of mind such as gaze following and joint attention, understanding intentionality in actions and pretence. I will also mention behavioural findings that some scholars interpret as evidence for false-belief understanding in infancy. For a review about representational theory of mind in infancy see Caron (2009). For a review of general psychological reasoning in infancy see Baillargeon, Scott and Bian (2016). Without delving too far into the scientific debates and interpretational controversies regarding infant's abilities some critical thoughts shall be mentioned first. In infant research there is evidently more room for interpretational leeway in terms of cognitive theories since inferences for cognitive processes (e.g. concept formation) rely mostly on non-verbal proxy measures (e.g. looking times) which are subject to various, often neglected, methodological issues (Kagan, 2008). Apart from this, there has been more fundamental critique to rich interpretations in infant research (Haith, 1988).

2.1.1.2.1. Intentional experiences.

Although belief understanding is often considered the hallmark of theory of mind, there is a large evidence base that understanding (at least some limited form) of other mental states like perception, emotion and desire emerges earlier. In particular, infants of a few months are already very sensitive to eye contact and gaze (Wellman, 2014, pp. 20-21; Miller, 2009, pp. 26-27). In the second half of the first year of life the dyadic interaction of infant and adult starts to encompass the environment (e.g. an object) on which focus of attention is shared which is termed *joined attention*. The following of an agent's line of sight or head orientation is called *gaze following* (e.g. when a mother unexpectedly turns her head) and emerges at around 8 months. Starting at 12 months, infants also lean or move to follow an adult's gaze behind a barrier. At the same age infants often follow the direction of the head-turn of adults whose eyes are blindfolded while 18 month olds do not, raising the question if early gaze following has more to do with overt tracking of head movement than some implicit understanding of the sense of seeing. Meltzoff & Brooks (2008) provided 12 month olds with experience with blindfolds and 18 month olds with experience with a special opaque blindfold. Consequently the younger infants stopped following a blindfolded adult's head-turn while 18 month olds started to follow it, demonstrating that it is not merely the head directedness but a sense of visual experience that guides their behaviour and that personal experience plays a crucial role to shape that sense. Another early example of infant's social cognition that can be interpreted as mentalizing or theory of mind is *social referencing*, the focusing of attention on and the use of social signals like facial expression or tone of voice (e.g. the infant follows the gaze of its mother toward a stranger entering the room and looks back at the mother's face) (Miller, 2009, pp. 27-28). There is also evidence that toddlers appreciate others' perceptual experiences in an epistemological sense and use the information to guide their behaviour. In a study by O'Neill (1996), a toy was placed out of reach of 2-year-old children while the parent was either present or away. When requesting help in retrieving the toy, children provided more detailed gestures to parents who did not witness the toy being placed high up. Other evidence comes from two studies where infants watched an adult experiencing two objects in joint engagement with the child, individual engagement or by looking at them and subsequently played with a third object while the adult was away (Moll & Tomasello, 2007). Eighteen and even 14-month olds who had previously been in joint engagement inferred that the adult was referring to the new toy when displaying excitement and asking the child to give it to him. Only 18-month olds also made this inference when they had only seen the adult engage with the object alone. Neither age group preferentially selected the new toy when they had only witnessed the adults look at the toy. As the task was basically to infer desire (the adult wants a toy) one can speculate that watching someone play with a toy imparts desire more strongly than just watching him look at it. In fact looking at a toy but not engaging with it might be understood as not desiring the toy in a rich interpretation. In another study where children had to decide which one of three objects a (in my opinion rather spooky looking) clown wants based on his eye gaze, 4-year-olds could reliably, and 2-year-olds after some trials, select the right object although pointing and head turning in conjunction with eye gaze was more facilitative for two-year-olds (Lee, Eskritt, Symons, & Muir, 1998).

2.1.1.2.2. Intentional actions.

Implicit understanding of intentionality in people's behaviour, such as grasping an object, also seems to develop early in infancy (Wellman, 2014, pp. 17-20). Infants even as young as 5 months seem to process actions beyond their perceptual and spatial features and attribute goals to them. In looking time habituation paradigm experiments they look longer at hand movements grasping a new object in the same location an old object used to be than at hand movements grasping an old object in a new location (Woodward, 1998). Similarly when habituated to a grasping of an object over a barrier, making the same hand movement (now impractical) when the hand barrier is removed elicits longer looking times in 8 month olds and older infants than direct grasping (Phillips & Wellman, 2005; Gergely, Nadasdy, Csibra, & Biro, 1993 for the original, slightly different paradigm). If the event was encoded merely in terms of spatial movement the old movement should habituate and elicit shorter looking times than the new movement. Further evidence of early understanding of intentionality and goal orientation comes from studies where infants watched adults perform goal oriented action sequences but fail to complete them and were subsequently given the opportunity to try the actions themselves. Eighteen (Meltzoff, 1995) and even 15 but not 12 months olds (Carpenter, Nagell, & Tomasello, 1998) perform the successful action more often than the failed action thus going beyond simply mirroring the observed action. Similarly when habituated to unsuccessful grasping of an object over a barrier, 10 and 12 but not 8 month olds look longer if the adult does not grasp the object indicating an ascription of intention behind the agent's movements (Brandone & Wellman, 2009). Importantly, longitudinal studies suggest a continuity between these early abilities and preschool theory of mind (Wellman, Lopez-Duran, LaBounty, & Hamilton, 2008; Aschersleben, Hofer, & Jovanovic, 2008).

2.1.1.2.3. Pretence.

Pretend play, as opposed to functional play, involves objects or situations that are not real (Doherty, 2009). Think about a child pretending a small pack of handkerchiefs to be a mobile phone (the classic example of a banana as a phone is not really appropriate anymore given the form of today's phones). This behaviour is first observed in late infancy, around 18 months but becomes more elaborate over the ensuing years. Young 2-year-olds but not old 1-year-olds react appropriately to a pretence transformation of another person (e.g. wiping up a spilled imaginary tea) most of the time. Cooperative social pretend play first emerges between two and three years of age and is followed by complex social pretend play which involves meta-communicating about the pretence play itself (e.g. roles, script) at about age four. However there is considerable variability with children from below-average childcare showing a delay of one year or more. Pretence can be seen as involving the mental state "to pretend" but minimally only requires one to act as if something was true when in fact you know it is false. Thus it might be that children imagine (propose) a situation (e.g. tea-party) and act on props accordingly without having an understanding that other participants have mental representations of the propositions which has been termed the action theory of pretence (Harris, Lillard, & Perner, 1994). According to Perner, Baker and Hutton (1994), children have a shared concept of pretence and belief called "prelief", prior to developing a distinct, complete concept of belief. This prelief enables them to differentiate between false propositions evaluated as false (pretence) as well as true propositions evaluated as true (belief) but not true propositions evaluated as false (false belief). A special case of pretence is divergent (also called discrepant) pretence in which a child has to acknowledge that different people can pretend different things in the same situation (e.g. a protagonist who takes part in pretend play leaves the room and in his absence the pretence, an imaginary milk in a glass, is emptied out and the child is asked what the protagonist thinks is in the glass when he returns). There is equivocal evidence but most studies suggest that such tasks are passed earlier, if not by much, than the standard false belief task (Doherty, 2009, pp. 99-100; Wellman, 2014, p. 51). Pretence is also related to social competence and false belief understanding although causal directions have not been firmly established yet.

2.1.1.2.4. Preference / Desire.

Children as young as 10 months seem to infer psychological states on the basis of statistical violation of physical probabilities (Wellman, Kushnir, & Xu, 2014). Infants who repeatedly see an actor take out five blue balls from a bowl where blue balls are the minority look longer when the actor – after habituation has set in– eventually chooses five red balls from two separate bowls with only red and blue balls. Children who witness the actor behaving the same but taking from a bowl were blue balls are the majority do not look longer when the actor chooses red balls afterwards. Infants apparently are sensitive to background probabilities and evaluate behaviour differently when compatible or incompatible with this background probability. It can be argued that they infer a preference. The same experimental design has been done with 20 month old toddlers earlier (Kushnir, Xu, & Wellman, 2010), demonstrating more direct evidence for the influence of background probabilities on inferences of preference or desire. In two conditions, an experimenter either repeatedly drew toy frogs from a bowl primarily containing toy ducks (minority condition) or toy frogs (majority condition). The experimenter left and the child was given two separate bowls of toy frogs and ducks. She then returned and stretched out her hand in the direction between the two bowls. Children in the minority condition predominantly gave her a toy frog while children in the majority condition were equally likely to give her a duck or a frog.

2.1.1.2.5. False belief understanding in infancy?

If no explicit prediction or explanation of behaviour is warranted, some understanding of false beliefs seems to be present much earlier. Clements & Perner (1994) found that the looking behaviour of children between 2 years 11 months and 4 years 5 months indicated false belief understanding compared to children between 2 years 5 months and 2 years 10 months. Other explanations for the finding have been tested but it indeed seems likely that there is implicit false belief understanding in very young children which acts mostly independently from explicit belief knowledge (Garnham & Perner, 2001; Garnham & Ruffman, 2001; Ruffman, Garnham, Import, & Conolly, 2001). A widely cited study utilizing a violation of expectation paradigm common in infant research pushed the limit even further and suggested implicit unexpected location false belief understanding in 15 month olds (Onishi & Baillargeon, 2005). As mentioned earlier a non-mentalistic explanation of the looking time differences have been laid out (Ruffman & Perner, 2005; Perner & Ruffman, 2005). In a reaction to this objection the task was modified and 13-month-olds found to expect agents with desires to act consistently with their knowledge, obtained through level 1 visual perspective taking (Surian, Caldi, & Sperber, 2007). However, inferring cognitive processes and concept formulation from looking time is subject to fundamental criticism (Haith, 1998; Kagan, 2008), so it was an important step to replicate the findings of early competence with other methods (Southgate, Senju, & Csibra, 2007; Buttelmann, Carpenter, & Tomasello, 2009). False belief understanding in older infants has also been shown with the unexpected content paradigm (Buttelmann, Over, Carpenter, & Tomasello, 2014). The interpretational scope of these findings vary widely depending on the theoretical stance as is the case in the debate about theories of mind in general.

2.1.1.3. Development of theory of mind.

My summary of key points about the development of first-order theory of mind will be short since the focus of this work is advanced theory of mind and emotion understanding.

Although understanding false beliefs has become to be viewed as synonymous with theory of mind (alas sometimes in a reductionist sense) there is a lot more about mental state understanding that develops in the preschool years.

Flavell (1999) reviewed the literature on post-infancy theory of mind developments. While toddlers can infer if a person can see an object based on its position (level 1 visual perspective taking), only in the later preschool years can they appreciate that an object may look different from a different perspective and calculate that perspective (level 2 visual perspective taking). Early sensitivity to a person's attention is visible in infant gaze following and joined attention (see chapter 2.1.1.2.1). The knowledge about the specificities of attention continues to grow during late preschool age and middle childhood however. Children begin to grasp that attention is selective, is constructive, limited and can be present at different levels (e.g. sleep). Desire understanding emerges earlier than belief understanding (Wellman & Woolley, 1990) and children talk earlier about desires than about beliefs (Bartsch & Wellman, 1995). In the second half of the second year of life children start to use desire terms correctly and by age 3 limited understanding of the relations between desires, emotions, actions and outcomes is evident. Comprehending intentionality is among the first things children begin to grasp about the mind as we have seen in chapter 2.1.1.2.2. However acquiring understanding is a gradual process and at least implicit and explicit understanding can be distinguished. At the beginning of early childhood children can distinguish intended actions from non-intended behaviour (e.g. mistakes). Later (around 4-5 years) they learn to differentiate desires, preferences and intentions and that they need not necessarily align. Understanding of mental states that are meant to represent reality (e.g. beliefs, appearance-reality distinction, level-2 visual perspective taking) undergoes dramatic progress in the preschool years. As already mentioned above (see chapter 2.1.1.1) explicit false belief reasoning usually develops between 4 and 5 years (Wellman et al., 2001). The distinction between appearance and reality (e.g. a sponge looking like a rock) develops a bit earlier (between 3 and 4 years) but performance is related between these two competences, as well as level 2 visual perspective taking. Beliefs about other targets than the locations or contents of object namely beliefs about morality or social conventions have shown a similar timeline of emergence (Flavell, Mumme, Green, & Flavell, 1992). Later (at the beginning of middle childhood) children learn that individual differences like expectations influence the interpretation of ambiguous stimuli. Another area of theory of mind concerns understanding of the process of knowledge formation (epistemology). Young elementary school children are far better in knowing how and when they acquired some fact than young preschool age children which often declare they have always known some information when in reality they have only gained it recently. Over the late preschool and middle-childhood years they become also more adept in judging which quality of an object (e.g. colour, weight) one can determine with which sense and how incomplete perception restricts the certainty with which to make a judgement. Children's knowledge about thinking has been another focus of particular scientific interest (Flavell, Green, & Flavell, 1995). Thinking becomes object of thinking early in preschool and generates metacognitive knowledge that only people (or at least animated agents) think, that thinking is different from external, physical events, that it requires a brain and mind and that it has content which may refer to real or imagined things.

Wellman & Liu constructed a theory of mind scale to investigate sequential progressions in theory of mind development (Wellman & Liu, 2004). Tasks in ascending difficulty are: diverse desires, diverse beliefs, knowledge access, false beliefs and hidden emotions. *Diverse desires* is about understanding that another person can have a desire different from the child's own and that this desire predicts his choice. Understanding *diverse beliefs* warrants acknowledging that two people can hold diverging beliefs upon an object and that this belief predicts where a person will search for that object. *Knowledge access* test if the child understands that a person who did not see inside a container does not know its content even if the child itself does. The *false-beliefs* task consists of an unexpected content procedure (understanding that another person judges a container's content after its looks even though the child knows that the content is something else). *Hidden emotions* assesses understanding that experienced and displayed emotions can diverge and lets the child judge them in a specific situation. This sequence of progressions in understanding has been found in the U.S. (Wellman & Liu, 2004; Wellman, Fang, Liu, Zhu, & Liu, 2006; Wellman et al., 2008), Australia (Peterson, Wellman, & Liu, 2005) and Germany (Kristen, Thoermer, Hofer, Aschersleben, & Sodian, 2006). Interestingly, Chinese children show a slightly different sequence than children from the U.S. or Australia. They acquire understanding of knowledge access earlier than understanding of diverse beliefs hinting at the influence of social experiences (Wellman et al., 2006; also see chapter 2.1.1.4 for a discussion). This developmental sequence has also been replicated longitudinally in U.S., Chinese and late-signing Australian deaf children (Wellman, Fang, & Petersen, 2011).

2.1.1.4. Influences on theory of mind development.

Despite a seemingly universal developmental trajectory in false-believe understanding there are large individual differences in its timetable. Numerous variables haven been implicated in accelerating or delaying this development or even constituting a precondition for its emergence. In the following chapters I will illustrate important findings of variables associated with theory of mind development.

2.1.1.4.1. Social experience – siblings, pretend play, mental state talk and parenting.

There are many findings that make clear that social experience plays a vital role in the development of theory of mind (e.g. see Miller, 2012, pp. 33-34 and Wellman, 2014, pp. 25-26). Number of siblings has been linked with earlier theory of mind development (Perner, Ruffman, & Leekam, 1994). Other studies have found evidence of this benefit only for older siblings (Ruffman, Perner, Naito, Parkin, & Clements, 1998; Ruffman, Perner, & Parkin, 1999) or siblings in general (e.g. Peterson, 2000) but not all studies have found this link (e.g. Cutting & Dunn, 1999; Pears & Moses, 2003). Demographic factors which have not been controlled for in most studies might explain the null findings (see Pears & Moses, 2003) yet Cutting and Dunn (1999) didn't find an association even in zero-order correlations. Clearly the existence of siblings alone does not necessarily foster theory of mind. Which factors (e.g. relationship quality) might play a crucial role awaits clarification. Family size and interaction with peers, especially older ones, has also been associated with theory of mind development (Lewis, Freeman, Kriakidou, Maridaki-Kassotaki, & Berridge, 1996). Part of this link might be explained through fostering of pretend play. Pretence has already been introduced as precursor of theory of mind. Pretend and fantasy play (including imaginary playmates) has also been associated with performance on false-belief tasks. More specifically pretend role enactments with siblings (Youngblade & Dunn, 1995) and role assignments and shared negotiations during pretence (Astington & Jenkins, 1995) predict false-belief understanding in children. Additionally, frequency of pretend play in preschool children has also been found to be associated with emotion understanding in particular (Lindsey & Colwell, 2003).

Mind-mindedness refers to a parent's inclination to treat its child as a being with a mind and predicts theory of mind performance (Meins et al., 2003). It is operationalized by evaluating parents' references to mental states when talking with the child or frequency of mental states when talking about the child. Richness in maternal mental state talk in general (e.g. while looking at picture books) is associated with theory of mind development (Bartsch & Wellman, 1995). The influence seems to be genuinely causal since maternal mental state talk predicts the child's theory of mind longitudinally while the reverse is not the case (Ruffman, Slade, & Crowe, 2002). In longitudinal studies Taumoepeau and Ruffman (2006, 2008) investigated the influence of maternal mental state talk on the child's mental state language and emotion understanding. Mother's mental state talk about desires at 15 months predicted toddler's mental state language and understanding of external causes of emotions at 24 months. Mother's mental state talk about her own and others' thoughts and knowledge at 24 months predicted the toddlers' mental state language at 36 months while her talk about the child's desires, thoughts and knowledge predicted the toddlers understanding of external causes of emotions.

Parenting style is another family variable that has been investigated. The influence of parenting on socio-cognitive and emotional understanding seems obvious yet findings are inconsistent. Parents' self-reported use of explanations about the effect of the child's behaviour on the feelings of others' was linked with belief understanding in one study (Ruffman et al., 1999) but unrelated in another (Pears & Moses, 2003). Performance on the theory of mind scale was positively related to authoritative but negatively related to authoritarian parenting style in school age children (O'Reilly & Peterson, 2014). Since the sample was rather old for the scale (mean age of 8 years) and performance on most sub-scales was near optimum the link is probably mostly supported by the last task, understanding hidden emotions. Unfortunately the authors don't report separate data that could answer this hypothesis. Hughes, Deater-Deckard and Cutting (1999) investigated parenting and mental understanding in young pre-schooler twins through a multi-measure assessment. Overall, specific negative control in videotaped dyadic interactions between mother and child was negatively correlated with theory of mind. Furthermore complex gender effects were found. While parental warmth rated by an independent observer was positively correlated with theory of mind competence for girls, disciplinary strategies as reported by the parent in an interview were positively correlated for boys. Another study (Pears & Moses, 2003) investigated demographic variables, parenting styles and mental understanding (perception, desires, beliefs, emotions). Power assertion (e.g. yelling, physical punishment) negatively predicted belief understanding but positively predicted emotion understanding, while consequence responses (e.g. removal of toy) negatively predicted emotion understanding in young pre-schoolers over age, cognitive ability, maternal education, income, marital status and number of siblings. The authors argue that emotional expressions accompanying power assertion might push children to understand emotions but emphasize that there is firm evidence for the overall negative socio-emotional consequences of power assertive discipline. Another explanation offered is that the causality might run the other way. That is, parents of children with poor mental understanding might tend to us power assertion because other parenting strategies might not succeed. Yet since the unexpected findings concerning mental understanding and negative parenting practices of both studies rest on self-report data, they are open to another possible explanation: metacognitive awareness of parents themselves and related validity of self-report data. Parents with more profound metacognitive competencies might be more aware of their power asserting behaviour and thus provide more valid self-report data while parents with lower metacognitive awareness might underreport their behaviour. The influence of metacognitive understanding on self-reported clinical symptoms has been shown with children previously (Sprung 2008; Sprung & Harris, 2010).

Carpendale and Lewis (2004) put forth an account of how understanding of minds is constructed gradually through social communicative interaction. This interaction takes place between a subject, an object (the world) and an interlocutor, coined epistemic triangle. They set this view apart from individualistic accounts of theory construction but also from mere collectivistic enculturation in which children internalize social understanding in an accumulation of cultural norms (e.g. Astington & Olson, 1995). According to Carpendale and Lewis's view, knowledge is grounded in action. Children first acquire sensorimotor action schemes through dyadic interaction with objects and people which constitute knowledge in the form of expectations what can be done with an object and how objects and people interact. Action schemes combine and become more separate from the infant's own actions. The repeated experience that people's behaviour does not always align with the child's expectations and desires under the basic assumption of a stable external world changes their understanding of mind. With the inclusion of objects the interaction becomes triadic and the infant learns early forms of pointing and gaze following which is not yet indicative of a referential understanding of attention. Through further interaction they learn more about the referential nature of attention and how to coordinate it with others. These abilities are the foundational insight for the development of language which greatly opens up the context for further development of social understanding. Following Wittgenstein (Montgomery, 1997) they argue that "Children learn the pattern of interaction for which it is appropriate to use a particular term, for example, mental or emotional, or dealing with pain, and so forth" (Carpendale & Lewis, 2004, p. 88).

Studies that that can be seen as evidence for the influence of social experiences as well as of language are those with deaf children. Deaf children born to hearing parents (late signers) demonstrate a marked delay in theory of mind development that is not visible in deaf children born to deaf parents (native signers) who communicate with their parents in sign language from early on (Peterson & Siegal, 1999; Jackson, 2001). On the theory of mind scale the delay has been shown to be about seven years (Wellman, 2014, p.102). Additional evidence based on this measure comes from trans-cultural comparative studies. Children from the U.S. and Australia first acquire understanding of diverse beliefs prior to understanding of knowledge access while children from China show a reverse developmental pattern (Wellman & Liu, 2004; Wellman et al., 2006). Cultural differences in values and epistemology that influence child rearing practices might explain these findings (Wellman, 2014, pp. 98-100). The United States and Australia are considered individualistic, independent societies where advocating one's own opinion and acknowledging different views is encouraged by parents while in China, a collectivist, interdependent society, Confucian ideals of loyalty and piety towards the elders are held in high esteem and own opinions are expected to be subordinated to shared values and social coherence. Likewise, western epistemology has been described to be centred more on truth while Chinese epistemology favours pragmatic knowledge acquisition. These differences in emphases are also reflected in U.S. and Chinese parents' conversational use of the mental states thinking and knowing with their children. This differing sequence in acquiring belief vs knowledge access understanding has also been found in Iranian vs. Australian children (Shahaeian, Peterson, Slaughter, & Wellman, 2011).

2.1.1.4.2. Incongruence experiences.

Bartsch (2002) emphasizes the role of inconsistency experiences in the development of belief understanding. Inconsistency experiences are defined as "any experience in which mental states and/or actions under the child's consideration are perceived as being inconsistent or incoherent with each other" (Bartsch, 2002, p. 153). One type of inconsistency experience arises from situations in which behaviour leading to frustration of desires can only be understood by reverting to underlying (false) beliefs. Another type of potential inconsistency is between other familiar mental states like for example, an emotion and a belief (e.g. the display of surprise facilitates inference of a false belief). Bartsch' assumptions can also be seen in the light of a general psychological inconsistency theory (Grawe, 2004) where inconsistency is broadly understood as incompatibility of parallel running psychological processes. One such kind of incongruence termed cognitive dissonance (Festinger, 1957) has been the subject of extensive investigation in social psychology. In the belief-disconfirmation paradigm in cognitive dissonance research, people are confronted with information contradicting their belief which may – among several alternatives – alter their belief to reduce dissonance (Grawe, 2004). To utilise an example of Bartsch (2002), a child who demands a cookie and is confronted with her mother looking in the wrong place is pressed to use an auxiliary construct of belief to make sense of the situation and reduce inconsistence/dissonance. This might not explain how the concept of belief originates in the first place (from a theory-theory stance) but how nascent auxiliary hypotheses of false-beliefs corroborate. I will shortly demonstrate some neuroscientific evidence that is consistent with such a learning mechanism.

The ability to comprehend, predict and control critical aspects of the environment arguably constitutes an evolutionary advantage. As social animals, the behaviour of other members of our species can be seen as such critical aspects of the environment. Epstein (1990) defined the need for orientation and control as one of human's basic needs in his cognitive-experiential self-theory. Violation of needs or in other words, incongruence between motivational goals and perceptions indicative of the fulfilment of these goals, leads to negative emotions and stress. Neuro-chemically, this stress is characterized by increased release of adrenalin and noradrenalin and eventually activation of the hypothalamic–pituitary–adrenal axis and the release of corticosteroids like cortisol. Central Noradrenalin (in the Locus Coeruleus) modulates reorienting of attention, interrupting "the activity of existing functional networks and facilitate their reorganization to promote rapid behavioural adaptation" (Sara, 2009, p. 219) and facilitates memory consolidation through cellular long term potentiation (Harley, 2004; Sara, 2009; Sara, 2015). Reduction of aversive conditions (i.e. stress, incongruence, e.g. the child which with the help of an auxiliary hypotheses of false belief regains orientation and control over an otherwise confusing situation) leads to strengthening of neuronal excitability patterns that led to the reduction of the aversive condition (Huether, 1998), a learning mechanism which is supposedly facilitated by dopaminergic modulation (Rada, Mark, & Hoebel, 1998; Harley, 2004; Lisman, Grace, & Duzel, 2011). Let us return to the example of the child which is frustrated (apart from not getting the cookie) by the mother's behaviour of not getting the cookie despite its learned expectation that mother cares for her. This inconsistency experience increases noradrenergic modulatory effects which "resets" currently activated networks to adapt behaviour (including information processing). This allows for the updating of the internal working model of reality to incorporate some representation of beliefs which better explain mother's behaviour and lead to a decrease of the aversive state of confusion and frustration which in turn facilitates learning. These mechanisms might explain from a neurobiological view why explanatory hypotheses like false beliefs consolidate. The medial prefrontal cortex and in particular the anterior cingulate cortex has been implicated in a wide range of functional roles conflict or error monitoring, outcome unexpectedness, error likelihood or volatility (Jahn, Nee, Alexander, & Brown, 2014). The anterior cingulate cortex has recently also been conceptualized as an action-outcome predictor (Alexander & Brown, 2011). In this view it is recruited in the instance of surprise, when actions do not lead to the expected outcome (Egner, 2011). According to another recent account (Karlsson, Tervo, & Karpova, 2012) anterior cingulate cortex also signals the “updating of beliefs and internal models of the environment” in events of unexpected uncertainty to adapt behaviour (Kolling, Behrens, Wittman, & Rushworth, 2016, p.36). Therefore it would not be surprising if the anterior cingulate cortex is also involved in incongruence experiences that facilitate developing internal models of false-belief.

2.1.1.4.3. Language and executive functions.

There are also strong links to be found between theory of mind abilities and language or executive functions (see Doherty, 2009, pp. 129-175 for a review and Miller, 2012 for a consideration of these factors in first-, second order and advanced theory of mind). Concerning language, the influence is probably bidirectional but more strongly for language on theory of mind (first order false-belief tasks) as a large meta-analysis has shown (Milligan, Astington, & Dack, 2007). Another meta-analysis found lower theory of mind performance in children between 4 and 12 years with specific language impairment compared to typically developing children (Nilsson & de López, 2016). How language abilities promote theory of mind is not clear however. Aspects like semantics, syntax and pragmatics have been proposed to play a role. Semantics and lexical abilities are concerned with understanding the meaning of linguistic constructs and have been linked with theory of mind generally (Dunn, 1988) and more specifically (e.g. knowing the meaning of the word think; Olson 1988). Syntax addresses the structure of language and how words combine to sentences. Syntactic competency in general (Astington, 1999) or the ability to build and understand embedded sentences (i.e. object complements) specifically (de Villiers, 2005) has been proposed to play a critical role in the development of representational theory of mind. Finally aspects of language pragmatics (how contextual information like knowledge and intent influences meaning) have been discussed as the main promoting factor (Harris, de Rosnay, & Pons, 2005). Alternatively general language abilities might better explain the influence on theory of mind development than for example syntax or semantics alone (Ruffman, Slade, Rowlandson, Rumsey, & Garnham, 2003). The meta-analysis by Milligan et al. (2007) looked at different types of language ability as moderators (general language, receptive vocabulary, semantics, syntax, memory for complements) and found that all effects were significant and ranged from moderate to strong effects. The only significant difference was that the effect of general language ability was larger than the effect of receptive vocabulary.

Another venue of research about language and theory of mind development are studies on deaf children already mentioned above. Additional evidence for the pivotal role of language on theory of mind development comes in the form of training studies. Lohmann &

Tomasello (2003) utilized several different training groups to investigate experience with deceptive objects that seemed to be one thing first (e.g. a flower) but turned out to be something else (e.g. a pen), syntactic characteristics, and use of mental state or communication verbs. They demonstrated that perspective-shifting discourse with only simple linguistic clauses and without using mental state or communication verbs improved false belief understanding while simply drawing attention to the deceptive nature with expressions and single words did not. Another training condition in which experimenters used sentential complements (e.g. X thinks that it is Y), including mental state or communication verbs, but did not introduce the deceptive nature of the object also improved false belief understanding. Finally, a full training condition including perspective-shifting discourse and sentential complements led to the largest improvement.

In the case of executive functions there is firm evidence that executive functions predict theory of mind abilities (Devine & Hughes, 2014). There is controversy however whether they are necessary for expression of competence in ToM tasks, emergence or both. I will discuss the interrelation of executive functions and theory of mind in more depth in section 2.1.3, because of their importance to the studies described later. Of particular note is the question how these two broad constructs are associated more specifically. A recent study with young adults showed that specific executive functions were related to some higher order theory of mind tasks but not others (Ahmed & Miller, 2011). Verbal fluency and deductive reasoning explained variance on the Strange Stories task, while verbal fluency, problem solving and gender predicted performance on the Faux Pas task. No executive functions predicted performance on the Reading the Mind in the Eyes Test.

2.1.1.5. Consequences of theory of mind.

Diverse competences and behaviours are assumed to be influenced by theory of mind ability (see Astington, 2003 and Wellman, 2014 for reviews). The focus in investigating these links has been on basic theory of mind measures like the false belief task however. One area of consequence of early or well developed theory of mind understanding concerns prosocial behaviours and getting along well with peers. Children with better false-belief understanding seem to be more accepted and popular among peers as ascertained by methods like peer nomination (Slaughter, Dennis, & Pritchard, 2002) and socio-metric peer ratings (Cassidy, Werner, Rource, Zubernis, & Baralaman, 2003). In several longitudinal studies theory of mind understanding predicted prosocial behaviour (Caputi, Lecce, Pagnin, & Banerjee, 2012) or social skills (Razza & Blair, 2009). Theory of mind can also be linked to specific real world social behaviours. In studies with preschool children false-belief understanding was correlated with teacher rated behaviours categorized as intentional but not with behaviours that are defined by social conventions (Lalonde & Chandler, 1995; Astington, 2003). However the link is only moderate and theory of mind understanding is not sufficient for the display of prosocial behaviour (Astington, 2003). Sophisticated theory of mind has also been associated with negative outcomes such as effective bullying (Sutton, Smith, & Swettenham, 1999a), oversensitivity (Dunn, 1995) and sensitivity to criticism (Lecce, Caputi, & Hughes, 2011).

Another area of consequence are abilities that per-se are not good or bad (in a socially normative sense) but indicative of an understanding of other minds: lies and deception, physical hiding and keeping secrets and persuading others. Wellman (2014, p.63) defines lying as "*making a false statement with the intention to deceive (where deception itself involves the intention to produce a false belief)*". He further elaborates that research has demonstrated that lying emerges a bit earlier than classic false-belief understanding although the occurrence of lying increases drastically between three and five years (Polak & Harris, 1999). A possible explanation is that different kinds of lying can be distinguished. While early lie-telling may merely be learned behaviour with the goal to avoid negative consequences there have indeed been found links with theory of mind. In experiments about temptation and peeking that did not include punishment, false-belief understanding predicted lying about peeking and not communicating knowledge gained by peeking in order to not reveal the lie (Talwar & Lee, 2008).

Playing hide-and-seek requires understanding what kind of information (e.g. location) someone should or should not have according to the rules of the game and in which ways this information can be obtained (e.g. through telling or seeing). In experiments by Peskin & Ardino (2003) only few 3-year-olds but most 4year-olds and even more 5-year-olds could successfully play hide-and-seek or keep a secret when asked for. Furthermore, behaviours were strongly correlated with false-belief understanding. Persuading other people beyond begging affords understanding information should be provided to a person to change her beliefs and desires and subsequently her behaviour. The work of Bartsch and colleagues about persuasion through information management revealed a transition from at chance performance at three years to above chance at four years and further improvement at five. The correlation with false-belief understanding was high, interestingly similar to hide-and-seek play and keeping secrets in the Peskin & Ardino studies.

2.1.1.6. Advanced theory of mind and measures.

What is advanced in advanced theory of mind? In one sense it can be used synonymously with higher-order theory of mind (see Miller, 2012). If there is a recursive loop of mental states (e.g. A thinks that B thinks) it is an advanced form of theory of mind, compared to a “simple” ascription of a mental state (e.g. A thinks that the chocolate is in the left drawer). As with the standard false-belief task, the second-order false-belief task (Perner & Wimmer, 1985) has been used extensively. In another sense advanced theory of mind may designate more complex social instances of mindreading compared to the false-belief task. Indeed to my knowledge (and supported by a quick search on Web of Science and Psycinfo), the term originates from the paper “An Advanced Test of Theory of Mind: Understanding of Story Characters' Thoughts and Feelings by Able Autistic, Mentally Handicapped, and Normal Children and Adults” by Happe (1994) which introduced the Strange Stories task described further below. The starting point of the study was the finding that individuals on the autism-spectrum with normal verbal intelligence may perform indistinguishably from neuro-typical subjects on false belief tasks while still displaying social handicaps. One explanation had been that these individuals solved the tasks using non-mentalistic strategies. To answer this question Happe devised a more naturalistic, contextually embedded measure of theory of mind. The success of this experiment instigated the development of other similar tasks, most of which were designed for use in the research of autism spectrum disorder. I will now describe later developments of mindreading falling under the terms higher-order or advanced theory of mind. Because this field lacks a comprehensive theory even more than basic theory of mind and was opened up pragmatically by new tasks with explanatory models lagging behind, I too will orient my review by these advanced measures.

One category of higher-order theory of mind measures can be labelled narrative tasks. These tasks usually consist of verbally, pictorial or audio-visually presentations of social narratives, often including statements that are not literally true, such as lies or sarcasm, or a faux pas. Most of these tasks were originally designed to probe performance in autism in more complex and realistic ways and in part to demonstrate an absence, not a delay in development as Miller (2009, 2012) points out. In the Strange Stories task (Happé, 1994), each item consists of a short story vignette including a nonliteral statement of a certain category (appearance-reality, persuasion, misunderstanding, sarcasm, double-bluff, contrary emotions, forgetting and white lie; e.g. blaming the dog for knocking over a vase). After a comprehension-question, the subject is asked to justify the story-character’s utterance (why-question), which usually requires addressing the speaker’s intent beyond the literal phrase. For many vignettes, answers to this question however can be scored as correct even when no reference to a mental state is made. Stories from Everyday Life (Kaland et al., 2002) is very similar to Strange Stories, slightly differing in the non-literary categories but more complex in terms of length and questions asked. The Faux Pas task (Baron-Cohen, O'Riodan, Stone, Jones, & Plaisted, 1999) follows the same rationale but is narrower in scope. Vignettes each contain a social faux pas in which a protagonist offends the feelings of another character apparently out of ignorance (e.g. disregarding a broken toy not remembering that it was the person he talks to that once gave him the present). Another task of social vignettes, Social Understanding Stories (Bosacki & Astington, 1999; Bosacki, 2000), differs in several notable points. First, it was designed for research on normally developing children. Second, it does not evolve around non-literary statements but "ordinary" social behaviour (e.g. two girls watching a new girl across the playground, nod to each other and approach the new girl who wonders what they want). Third, answers are not scored correct or incorrect but according to their complexity in terms of mental state understanding. Recent years have seen the emergence of multimedia based tasks to investigate theory of mind abilities in an even more naturalistic way. Analogue to the Reading the Mind in the Eyes task, a Reading the Mind in the Voice (Rutherford, Baron-Cohen, & Wheelwright, 2002; Golan, Baron-Cohen, Hill, & Rutherford, 2006) and Reading the Mind in Films task (Golan, Baron-Cohen, Hill, & Golan, 2006) has been developed. The former comprises spoken phrases which for themselves (i.e. when written) are ambiguous in regard to complex emotions of the speaker which are only conveyed through the voice. The latter consists of short scenes of commercial films in which complex social emotions have to be labelled. The Movie for the Assessment of Social Cognition (Dziobek et al., 2006) on the other hand was specifically created for the purpose of the task. It is a small movie split between several short scenes about a group of friends getting together. After each sequence the film is stopped and questions about the character's thoughts, intentions and feelings are posed.

Another post false-belief development is the understanding of interpretive diversity, the acknowledgment that one thing can mean two different things to two individuals (Miller, 2012, pp. 93-100). False-belief has been seen as an instance of interpretive diversity although the crucial difference is that in false-belief different agents have different information available to them leading to different interpretations of the same situation. Interpretive diversity however usually means perceiving or attributing differently in the face of objectively identical stimuli. On a deeper level interpretations of agents are always based on different information because of individual differences of the minds/brains of interpreters. One paradigm used to investigate interpretive diversity are ambiguous stimuli like reversible figures. Two famous examples are the Rubin vase, a black vase on white ground vs. two white faces in profile directed at each other on black ground and the duck-rabbit illusion, a drawing that can be seen as a duck or a rabbit.

Other forms of ambiguous stimuli are homophones (e.g. *mail-male*, or in German which has generally less homophones: *Meer-mehr*) or ambiguous instructions. Chapter 2.1.1.8 describes a visual perspective taking task (the Director's Task) used to assess theory of mind in adulthood which utilizes ambiguous instructions that can only be resolved by taking the perspective of another agent into account. Droodles are minimalistic abstract line drawings which do not themselves contain enough visual information to be unambiguously interpretable but have an explanatory caption that gives them meaning. Imagine or sketch a rectangle with two parallel vertical lines within and several circles between the lines. The caption could read: a giraffe passing a window. Similarly Droodles have been created that are unambiguous line drawings made ambiguous by covering large parts of it. Imagine a wedge on the left pointing right and another wedge on the bottom pointing up. When the whole picture is revealed it proves to be a ship coming to rescue a witch which has already drowned down to her pointed hat. In studies where children are shown the whole picture and have to predict the behaviour of another agent without that information, children of about 4 years usually succeed which is about the same age that standard false-belief tasks are passed (Ruffman, Olson, & Astington, 1991). In stark contrast, the understanding (measured through prediction or explanation) that two people usually form different beliefs (or trains of thought) based on the same ambiguous information (e.g. pictures, homophones, messages) or have different trains of thought regarding the same object is very limited in pre-schoolers and progresses considerably during the primary school years (Carpendale & Chandler, 1996; Eisbach, 2004). This finding has also been replicated with an ecologically more valid task where subjects had to make sense of conflicts between siblings about which they only had limited information (Ross, Rechia, & Carpendale, 2005). Going beyond acknowledgment that two people can interpret ambiguous situations differently is the question how inclusion of information about individual differences (e.g. personal attributes) affects interpretive diversity (Pillow, 1991; Pillow & Weed, 1995). A story character (A) is liked or disliked by other story characters. Character A then engages in damaging action, accidentally or with ambiguous intent. The child subject has to predict how the two other characters interpret A's behaviour. Children in second grade but not earlier typically take into account the information of individual differences potentially biasing people's interpretation.

The development of interpretive diversity is not necessarily of monotone growth however. Lagattuta, Sayfan & Blattman (2010) found that 6- to 7-year-olds over-interpret the diversity of interpretations. Subjects more often than younger or older participants judged that characters looking at snippets of pictures that unambiguously reveal the identity of the whole picture have different interpretations regarding the picture. A follow-up study that used similar occluded pictures but had the subjects rate the probability of different interpretations of naïve characters found this over-interpretive bias for 6- to 10-year-olds compared to preschool children and adults (Lagattuta, Sayfan & Harvey, 2014). The same study also provides further evidence that theory of mind continues to develop quantitatively, even when a qualitative understanding has been developed. All age groups, from preschool age to adulthood that had seen the full picture over-estimated the probability that naïve characters would guess the actual picture compared to a control group of participants that had not seen the full picture. This is an example of the phenomenon that one's perspective, knowledge or beliefs influence how we appraise perspectives, knowledge or beliefs of others which has been investigated under terms like egocentricity bias, curse of knowledge or hindsight-bias. An early study has shown that it is difficult even for adults to ignore privileged knowledge not shared by another individual when interacting with it (Camerer, Loewenstein, & Weber, 1989). More specifically this "curse of knowledge" has been used to quantify false-belief understanding in adults (Birch & Bloom, 2007). In this modified change-of-location false-belief task a girl puts her violin in a blue container. Depending on the experimental condition, in her absence the violin changes location to a red / violet / or unknown-to-the participant container. Subsequently the containers are spatially rearranged so that the red container is now in the same spot like the blue container before. Subjects have to rate in percentage how likely it is that the girl, upon her return, looks in each container first. In the condition with the red container, subjects rate the probability for the girl to look there significantly higher than in the unknown or the violet-container-condition. Knowledge of the real location of the violin influenced the subjects judgement, but only when the outcome was somewhat plausible (i.e. the girl looking in the same location even when the container was different). The size of this effect in adults has been called into question however (Ryskin & Brown-Schmidt, 2014). Relatedly a hindsight-bias for judging when a naïve observer would identify a pixelated or blurred object that is progressively clarified exists even in adults. In contrast to the over-interpretive diversity found in children, there is evidence for a false consensus effect in young adulthood. Young adults judged that naïve characters interpret am ambiguous stimulus more similarly compared to children or older adults (Dunning & Hayes, 1996; Lagattuta, Sayfan et al., 2010).

The recognition of the influence of concrete visual information (e.g. a picture) on knowledge and beliefs is a relatively simple example of interpretive diversity. Recognition and appreciation of individual differences in more abstract concepts like views on science, religion, politics or morality are achievements of late childhood and adolescence and are open to development even in adulthood (Hofer & Pintrich, 1997; Wright, 2012; Haidt, 2013). This reflective understanding or knowledge of mental activities is often referred to as meta-cognition (see Flavell, 1999 and Pillow, 2008 for reviews). The recognition of the existence of divergent beliefs is just the starting point. Thereafter acquiring and applying knowledge of how mental activities (e.g. judgments) are influenced and biased by external and internal factors is a potentially continuous and open-ended development. The perspective of knowledge, ability or competence is not everything there is to mental understanding still. Another aspect previously largely ignored is the motivation to engage in effortful mindreading in the first place (Carpenter, 2016). Motivation drives our behaviour and is a prerequisite for competence to manifest itself in performance. Without it, the ability lies dormant without effect. Furthermore without motivation to engage in mindreading, in the absence of social experience, the ability is unlikely to refine. This mind-reading motivation seems to be relatively stable over time, predicts depth of descriptions of others' perspective, amenability to different styles of persuasive arguments, behaviour in team contexts and leadership-styles (Carpenter, 2016).

The nomological network around theory of mind (related concepts that in many cases go back to before the term ToM was even coined) is too large to discuss in its entirety, even when focusing on aspects of prolonged development. For the sake of completeness I want to briefly summarize a list of Miller (2012, p.195) who states general themes related to theory of mind of what develops beyond the preschool years:

- Reduction in egocentricity-bias.
- Understanding of interpretive diversity and growing acknowledgment of the influence of individual differences (e.g. experiences, traits, relationships) on mentalizing.
- Understanding a wider range of mental states and increased flexibility, generalizability and efficiency of mental understanding.
- Advances in meta-cognition and growing knowledge about mental activities and epistemology.

2.1.1.7. Associations between tasks of advanced theory of mind.

There are only few studies that have used more than one task of advanced theory of mind and reported the correlations and results are mixed. The following short review on associations does not include findings on the second-order false-belief task since most children master it at the age of five or six (Miller, 2012). Furthermore following the rationale of the first-order false-belief task it is a task of pass or fail and it is not well suited to study advanced theory of mind in middle childhood and beyond which is of primary interest to this work.

In a study comparing adolescents with and without specific language impairment (SLI) Strange Stories correct mental answers were correlated with the Eyes task in both groups but association was higher (although not statistically tested) for adolescents with SLI than for typical developing subjects (Botting & Conti-Ramsden, 2008). Similarly in a study comparing adolescents with Asperger syndrome or high-functioning autism with typically developing adolescents Strange Stories and Stories from Everyday Life were significantly correlated in the autism spectrum disorder group but not in the control group (Kaland, Callesen, Moller-Nielsen, Mortensen, & Smith, 2008). The Eyes task was not correlated with Strange Stories or Stories from Everyday Life for either group. Size of associations between Strange Stories and the Eyes task was similar to the study by Botting & Conti-Ramsden (2008) however and small sample size might have led to the non-significant results. In another study with adults on the autism spectrum and neuro-typical subjects, Strange Stories' correct answers (correctly explaining the meaning behind a non-literary statement) but not Strange Stories' mental answers (answers including a mental as opposed to a physical justification) were significantly correlated with the Faux-pas test and the Eyes task which themselves were not significantly associated with each other (Spek, Scholte, & Van Berckelaer-Onnes, 2010). No correlations for sub-groups were reported. A study with middle-aged children on the autism spectrum and normal controls reported correlations between Strange Stories (mental and physical control stories), a cartoon task which asks subjects to explain why a cartoon was funny which requires attributing mental states and the Eyes task. They also provided correlations partialed for individual difference variables (age, intelligence and language-age), providing more meaningful information on task associations (Brent, Rios, Happé, & Charman, 2004). Surprisingly, performance for the autism spectrum group on the Eyes task was negatively correlated with Strange Stories mental but not physical stories when controlling for individual difference variables. The authors did not provide an explanation for this particular finding. For the typically developing group there was no significant association between the two tasks. This might be owed to the relatively small sample (n=20 per group) since the size of the uncorrected correlation with mental stories was not so small (.29; .18 for the partial correlation). However, correlations for the physical control stories were higher, speaking for general underlying factors between the tasks not similarity of advanced theory of mind measured. Associations with the mental cartoons were quite large for both mental Strange Stories and the Eyes task in the neuro-typical group and for mental Strange Stories in the ASD group. However again, for all these measures, correlations with the corresponding physical control items (Strange Stories and cartoons task) were of equal or even larger size. This suggests that these associations are largely based on shared non-specific task demands, not mentalizing ability. A study with middle-aged deaf children and adolescents found associations between Strange Stories mental answers, the Faux-pas task and a second-order false-belief task although only one vignette per task was used and separate results for native- and late-signers were not reported (Meristo & Hjelmquist, 2009). Concerning interpretive diversity, spontaneous alternating between ambiguous figures was significantly correlated with second-order false belief understanding in primary school age children (Mitroff, Sobel, & Gopnik, 2006). A significant association was also found between a composite variable of Droodles and ambiguous figures and first- and second-order false-belief tasks in a study controlling for age and language (Comay, 2010). Two other studies have found significant correlations between a second-order false belief task and the faux pas task (Banerjee, 2000; Qualter, Barlow, & Stylianou, 2011).

In contrast a large study investigating executive functions and theory of mind in healthy young adults did not find any correlations between either Strange Stories, Faux-pas or the Eyes task (Ahmed & Miller, 2011). Hayward (2011) extensively compared several advanced theory of mind tasks in typically developing middle-aged children and young adolescents. No significant correlations between Strange Stories correct answers, Strange Stories mental answers, Faux-pas test, the Eyes task, ambiguous figures or the Droodles task were found. Only the second-order false belief task was significantly but weakly correlated with the Droodles task and the Strange Stories correct answers.

Taken together the evidence concerning association between tasks of advanced theory of mind is inconclusive. There seems to be a tendency in the reviewed studies for clinical populations to show higher associations between tasks than normal populations although to my knowledge no study has systematically investigated this question (e.g. as a moderator in a meta-analysis). Interestingly evidence for associations between second-order false belief tasks and advanced measures of theory of mind is more solid than for associations among advanced measures of theory of mind themselves. This association however is only evident for early middle childhood since afterwards performance in second-order false belief tasks is near optimum.

2.1.1.8. Theory of mind across the lifespan.

2.1.1.8.1. A literature review of advanced theory of mind performance across the lifespan

As mentioned in the introduction of measures of advanced theory of mind, most research utilizing these tasks did not investigate age differences but rather group differences between clinical samples and normal controls with typically small samples and large age ranges, limiting the informational value regarding developmental issues. Miller (2009, 2012) reviewed the available evidence from non-clinical samples (mainly controls of clinical studies) of children and adolescents on advanced theory of mind tasks and concludes that most studies found an increase in performance with age (although see Hayward, 2011 for diverging findings). However, data ranging from middle-childhood to adolescence is often lumped together, thus obfuscating later development and possible ceiling effects in these tasks. Henry, Phillips, Ruffman, and Bailey (2013) compared age differences of adults in theory of mind and found a moderate effects with older adults performing worse than young adults. Data for some age groups and measures was sparse in these two reviews however. I conducted a restricted literature review to investigate the available evidence for age differences in advanced theory of mind across the life span more comprehensively for three widely used measures: strange stories (SS), faux pas task (FP) and the reading the mind in the eyes task (RMET). Literature search was done in March 2018 using the PsycINFO database which lists more than 2500 peer reviewed journals, over 4 million records with a temporal coverage from 1887 to the present. To keep the extended search in line with the first two reviews all searches were limited to publications before 2014. Exclusion criteria were: non-english language; publication types other than peer-reviewed articles (e.g. dissertation theses, book chapters); insufficient or ambiguous information about scoring procedures (e.g. paper of the original task was cited in the scoring section but mean scores indicated that more or less points were awarded for each item than described in the original paper); significant changes to the original or prevalent scoring procedures (e.g. -3 to +3 points instead of 0 to 2 awarded for strange stories vignettes or only binary detection questions counted for faux pas), drastic changes to the original or prevalent stimuli (e.g. video instead of pictorial stimuli, only one vignette used); no full text available; duplicate publications (i.e. same data reported in another paper identified through the search); no raw, mean or accuracy data reported; *other* (e.g. study type: not a controlled study, review, meta-analysis or task not given to the control group).

For strange stories search terms were: *strange stories* in *title* OR *abstract* OR *tests & measures*. This search returned 85 entries. Fifty-five of these studies were found not be eligible for inclusion: six were not in English, 13 because of publication type, three of unclear scoring, one because of scoring change, six studies that did not present data, one duplicate publication and 25 belonging to the other category. Twenty-five eligible studies were extracted from the two reviews (18 from Miller 2012, 7 from Henry et al., 2013) and one from the faux pas search (see below). After removing 12 duplicate findings, a total of 44 eligible studies remained.

For faux pas search terms were *faux pas* in *title* OR *abstract* OR *tests & measures*. In a preliminary search it became evident that a large number of returned studies did not include a healthy control sample so an additional restriction was included: *healthy* OR *control** OR *comparison subjects* OR *comparison group* in *abstracts*. This search returned 89 studies. Of these, 56 did not meet the inclusion criteria (language: 6, publication type: 6, stimuli change: 3, scoring unclear: 4, scoring changed: 10, no data given: 4, no full text available: 6, duplicate publication: 3, other: 14). Eight studies were extracted from Miller (2012), two from Henry et al. (2013) and four identified through papers returned in the strange stories search. After removing one duplicate finding, a total of 46 eligible studies remained.

For RMET search terms were (*rmet* OR *eyes task* OR *reading the mind in the eyes* in *title* OR *abstract* OR *tests & measures*) AND (*healthy* OR *control** OR *comparison subjects* OR *comparison group* in *abstracts*). Because studies with adults samples found through the searches for strange stories and faux pas were already abundant, age groups were limited to children and adolescents (AND (*children* OR *adolescents* in *age group*)). This search returned 30 studies. Fifteen of these did not meet the inclusion criteria (language: 2, publication type: 5, no data: 4, no full text available: 1, other: 3). Thirteen eligible studies were identified through the reviews and 30 through the previous strange stories and faux pas searches. After removing three duplicate findings, a total of 55 studies remained. A reference list of included studies can be found in Appendix B.

Adult samples were divided into three age groups (in analogue to PsycINFO age group terms) according to their mean age: young adults (18-40 years), middle-old adults (40-65 years) and older adults (65+ years). Accuracy scores of different healthy samples in the same study (e.g. male vs. female, different experimental groups at time 1) were merged to achieve more representative samples. Figures 1 to 3 show scatter plots of mean performance of normally developing samples from the studies extracted in the literature review for the Strange Stories, Faux Pas and the RMET. Noticeably, data from adolescence is sparse and completely lacking in the case of the Faux Pas. Growth in task performance for Strange Stories and Faux Pas is primarily visible at the transition from early to middle childhood and in the first half of middle childhood (approximately age six to nine).

For the strange stories task (Figure 1) there is a marked increase in performance through middle childhood until about age 12 when it reaches adult performance (*M*=84%). In adulthood, performance is stable between younger and middle-old adults with one study utilizing an unusual dichotomous scoring procedure skewing the distribution for middle adulthood (*M*=87%, *Mdn*=87% and *M*=82%, *Mdn*=85% respectively). In older adulthood a distinct drop is visible (*M*=73%, *Mdn*=70%) and performance is similar to middle childhood.

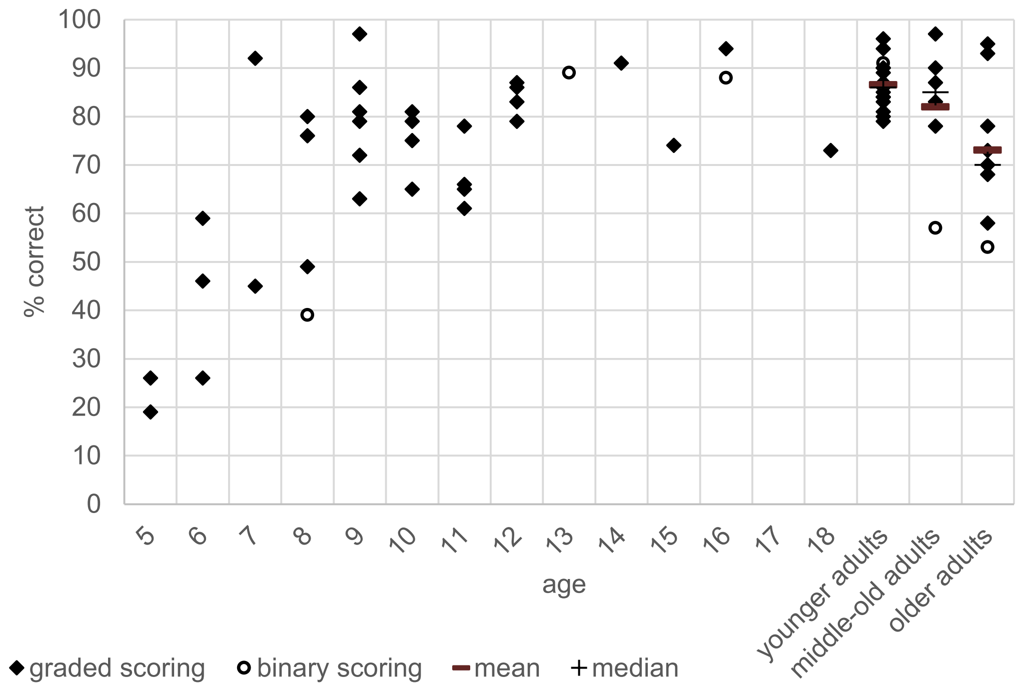

Figure 1. Strange Stories mean performance of normally developing samples.

Regarding the faux pas task (Figure 2) no study with an adolescent mean age group could be identified. There is a steady increase in performance during middle childhood. All included adult data is based on the graded scoring scheme of (or one similar to) Stone, Baron-Cohen and Knight (1998) while all but one child and adolescent data is based on the binary scoring scheme of (or one similar to) Baron-Cohen et al. (1999). Samples at age 7, 8, 9 and 10 that distinctly stand out in terms of high performance rates were from two studies (Banerjee, 2000, for ages 8 and 10; Filippova & Astington, 2008, for ages 7 and 9) that used only 2 vignettes played out with dolls and props (Banerjee, 2000). In adults performance is similar for younger (*M*=88%, *Mdn*=91%) and middle-old adults (*M*=86%, *Mdn*=91%) while for older adults there is a slight trend towards lower performances (*M*=83%, *Mdn*=84%). Performance between children and adults has to be compared with caution however, since child and adult studies used a different scoring scheme. For the child faux pas task, a dichotomous scoring scheme was used (i.e. all justification questions have to be answered correctly to score one point for each vignette) while for the adult faux pas task, points for the justification questions were summed. For adults the soring scheme implicitly acknowledges that social situations containing a faux pas can be partly understood and that this information is meaningful, while for children this assumption is not given.

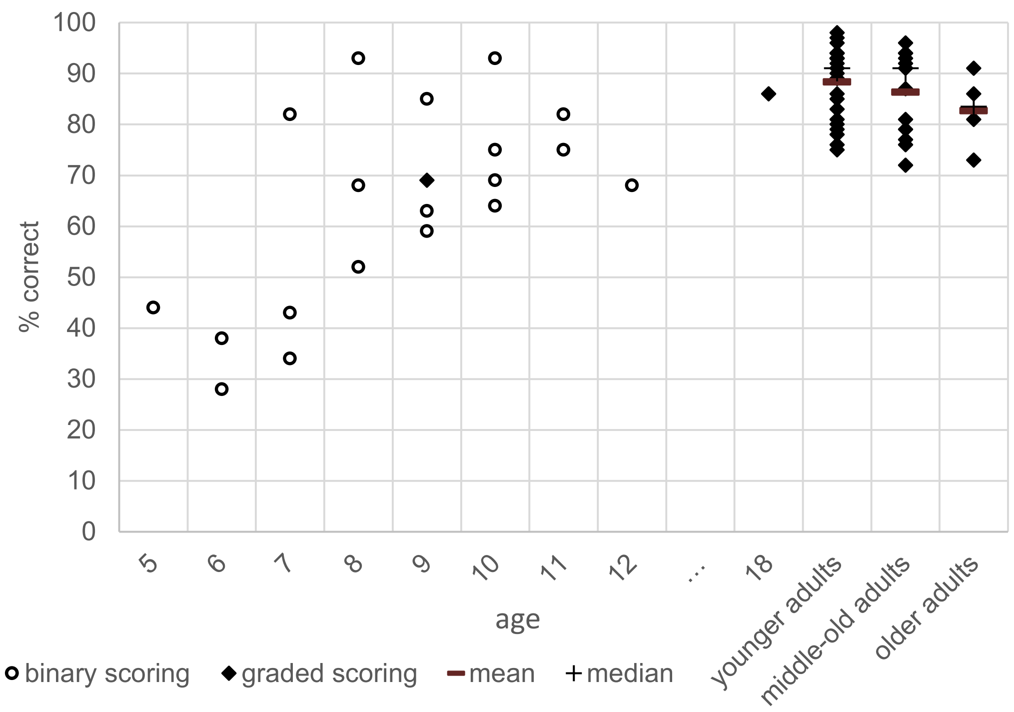

Figure 2. Faux Pas mean performance of normally developing samples.

For the Eyes task, developmental data is disheartening. No trend is visible from age 10 to 16 and performance in this age range is almost at adult level (*M*=70% vs. *M*=72% for young adults). In adults performance is more or less constant between early and middle adulthood (*M*=72%, *Mdn*=75% and *M*=74%, *Mdn*=74% respectively) while there is a slight decrease in older adulthood (*M*=69%, *Mdn*=67%).

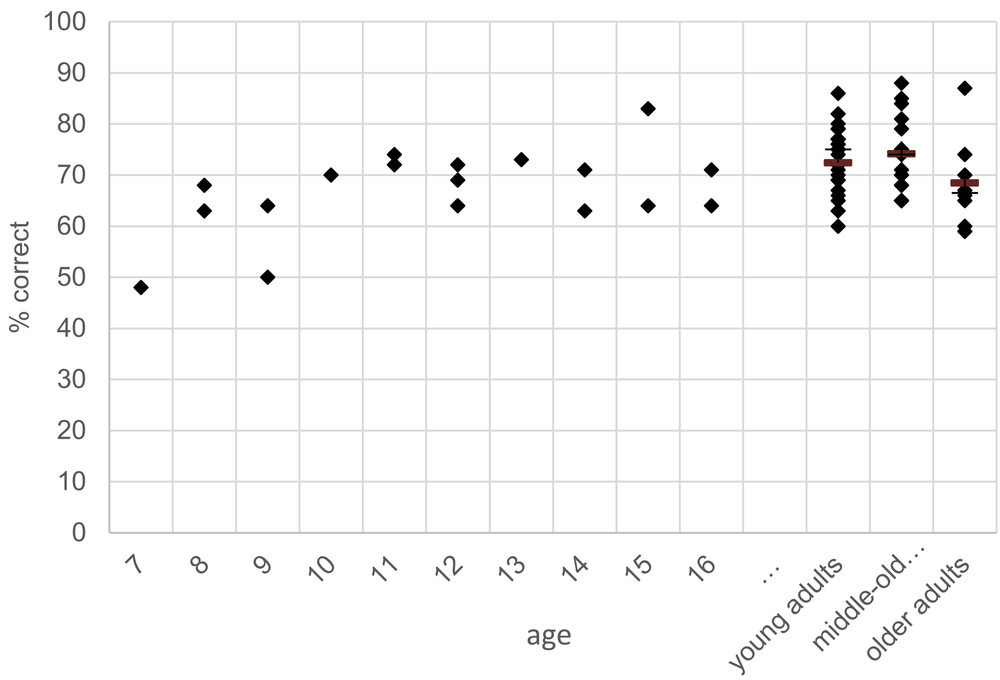

Figure 3: Eyes Task mean performance of normally developing samples.

Comparing different age groups originating from different studies of course involves the danger of drawing conclusions from heterogeneous samples. Comparing different age groups from the same study reduces such noise and within subject comparisons from longitudinal studies even more so. For the Faux Pas task, there are four studies identified by Miller (2012) comparing age groups. In three studies there was (virtually) no difference between 7 and 9 years (Filippova & Astington, 2008), 8 and 10 years (Banerjee, 2000) or even 8 and 12 years (Hayward, 2011). In one longitudinal study performance increased from 9 to 10 but not from 10 to 11 years (Banerjee, Watling, & Caputi, 2011). Taken together these findings suggest that for the Faux Pas task there is no further improvement in adolescents. For the Eyes task, there is sparse and contradictory evidence with one study showing progression across middle childhood (Baron-Cohen, Wheelwright, Spong, Scahill, & Lawson, 2001) and one larger study finding an almost flat trajectory (Hayward, 2011). Concerning the Strange Stories, two studies found only a very small increase between 7 and 9 years (Filippova & Astington, 2008) and 8 and 12 years (Hayward, 2011). Another study used a modified scoring scheme which only rewarded points for correct mental justifications (O'Hare, Bremner, Nash, Happe, & Pettigrew, 2009). They found continuing development between 5 and 12 years of age. More specifically a period of rapid development in the transition from early to middle childhood (5-7 years) was followed by a plateau between 7 and 8, another increase between 8 and 9, yet another plateau between 9 and 11 and a final substantial step forward between 11 and 12 years of age. Thus the traditional Strange Stories procedure does not seem to capture developmental trends beyond 7 or 8 years well, but a scoring which focuses on mental justifications does, at least through middle childhood.

Taken together existing measures of advanced theory of mind have not produced decisive evidence of developmental changes in non-clinical samples beyond middle childhood. This could mean that they are insensitive to later changes or that development is largely complete at this age. Brain imaging studies however point to a protracted development into adulthood. Brain activity can reveal differences in information processing even if behavioural tasks do not show variance. Blakemore (2010) reviewed developmental changes in the brain associated with social cognition and found evidence that activation of prefrontal cortex during mentalizing tasks is lower in adults compared to adolescents, even when equating performance across groups, which could indicate more efficient neural processing. Furthermore the review brought forth evidence for a shift of activity from anterior to posterior regions with age for thinking about intentions.

Recently new tasks, more complex or with a continuous response variable, have been devised that are suited to capture more subtle developmental than traditional tasks. Some of these findings, that demonstrate a continued development of social cognition, are reviewed below.

2.1.1.8.2. Measures from basic theory of mind research paradigms adapted for use in adolescence and adulthood.

One research strategy has been to just take first-order (Sue thinks) and second-order (Sue thinks that I think) recursive thinking and add loops thus creating third-, fourth or even fifth-order theory of mind tasks. Higher-order recursive loops of reasoning are difficult to handle even for older children but adults can handle chains of up to four items relatively well (Miller, 2012). The ecological validity of higher-order recursive loops is questionable however. Even complex examples of mentalizing usually do not involve changes of mental perspective beyond three (e.g. how will Sue react when she learns that I know that she really doesn't care about my ill relative). Recently new tasks, more complex or with a continuous response variable, have been devised that are suited to capture more subtle development than traditional tasks. Some of these findings, that demonstrate a continued development of social cognition, are reviewed below.

Most research on theory of mind beyond the preschool years introduces new aspects (social conventions, social emotions) and raises task complexity (multiple agents, context information) compared to basic aspects of theory of mind (e.g. intentionality, false-belief). An exception to this is visual perspective taking for which a task has been devised that captures variability even into adulthood. Like in most instances of theory of mind research in older children and beyond, the question is not if a person can pass a task or not but rather how well she can, in terms of accuracy and/or speed. The focus shifts from a dichotomous frame of reference (competence / no competence) to a continuous estimate of competence. In the director's task, (Keysar, Barr, Balin, & Brauner, 2000; Dumontheil, Apperly, & Blakemore, 2010) subjects have to move objects in an array or cupboard (in real-life or on a computer screen) as requested by a director standing on the other side of the array. Some objects are visible by both, the subject and the director, while some obfuscated to the director (e.g. by a back panel). In a director-condition subjects have to take into account the perspective of the director to resolve ambiguities (e.g. two balls of which one is obfuscated). In the no-director condition subjects are instructed to move certain objects but ignore objects that have a coloured back panel, thus resolving ambiguities via a specific rule held in memory. Thus the no-director condition removes the visual perspective taking aspect. In a study with the computerized version with several age groups from middle childhood to early adulthood, subjects' accuracy improved in the director condition between older adolescents and young adults but not in the no-director condition demonstrating that visual perspective taking improves even into adulthood (Dumontheil, Apperly, & Blakemore, 2010). The authors argue that while the competence to take other visual perspectives might be fully developed in childhood, the propensity to actually take the other perspective increases further.

A continuous measure of implicit egocentricity bias is the sandbox task (Sommerville, Bernstein, & Meltzoff, 2013). In a classic false belief change-of-location paradigm enacted in a sandbox subjects have to mark the spot where Sally will look for the object. To prevent subjects from fixating the original location, a visual search distraction task is administered. Dependent variable is the distance between the original location and the estimated location. In two experiments, 3- and 5-year-olds as well as young adults were biased by their own private knowledge about the object's actual location. Two control conditions were included to rule out effects of memory or attentional demands. In a no-false-belief condition, a second object was placed while the protagonist watched to match attentional demands. In a memory-condition, the subject, after the object had been moved, was asked where the protagonist put the object before he left. In another study (Bernstein, Thornton, & Sommerville, 2011), this task also revealed that this false belief bias was greater in middle (M = 56 years) and older adults (M = 68 year) compared to younger adults (M = 19 years) independently of neuropsychological functioning. In summary these new measures promise to capture individual differences in mentalizing in typically developing adults without relying on verbal skills to such a large amount as traditional tasks do. In addition tasks suitable for both young children and adults transform the perhaps artificial catecorial chasm between child and adult theory of mind into a dimensional slope.

2.1.1.8.3. Theory of mind in older age.

While studies addressing age-differences in theory of mind beyond childhood are scarce, old age has received more attention since Happé, Winner and Brownell's (1998) finding of improved Strange Stories performance in old age. Most studies since then, however, have found evidence that points in the other direction.

A recent meta-analysis investigated changes in theory of mind in old age (Henry, Phillips, Ruffman, & Bailey, 2013). The authors differentiated between six types of tasks (*stories, faux pas, videos, eyes, false belief-video* and *false belief-other*), three domains (affective, cognitive, mixed) and five modalities (verbal, visual-static, visual-dynamic, verbal and visual-static, or verbal and visual-dynamic). Age was compared in two groups with the younger group consisting of samples with a mean between 19 and 56 years and the older group of samples older than 65 years and the additional inclusion that mean age of groups in a study had to be 25 years apart for the study to be deemed eligible for inclusion. Effects of decline in old age were found in all types of tasks, domains and modalities with an overall association of -.41. Additionally, age effects on theory of mind conditions were larger than effects on control conditions (e.g. physical vignettes) for all but the *eyes* task and *videos* which were both only based on one study. Notably, for *stories*, *faux pas* and *false belief-video* task categories effect sizes of age decline were twice as large for the theory of mind conditions compared to the control conditions. The authors argue that the comparison with control conditions make a decline due to linguistic demands unlikely although executive functions might still play a role due to the difficulty to match mental and control conditions in this regard. A qualitative literature review conducted around the same time as the aforementioned meta-analysis comes to the same conclusion that theory of mind, evidenced by different measures, is indeed impaired in older age (Moran, 2013). The author also reviewed the evidence of mediations by crystallized and fluid intelligence as well as executive functions. He concludes that in the Reading the Mind in the Eyes task impairment has more often than not been shown to be independent from general cognitive decline and executive functions but related to emotion recognition. The case with inferring mental states from video stimuli is similar with the exception that there is evidence for dependence on inhibitory processes in some tasks and emotion recognition abilities. Regarding first-order theory of mind (including Strange Stories) age-differences often disappear when matching young and older adults on fluid intelligence and executive functions, suggesting that they underlie age-related declines. Good vocabulary (and crystallized intelligence) seems to have a protective effect against age-related declines. Almost all studies used discrete measures of theory of mind however. On a continuous change-of-location task however, age differences were independent from executive functions and crystallized intelligence. There is a dearth of evidence regarding second-order tasks. A recent study (Duval, Piolino, Bejanin, Eustache, & Desgranges, 2011) compared younger (*M* = 24 years), middle-aged (*M* = 53 years) and older (*M* = 70 years) adults in subjective and objective measures of cognitive (attribution of intention, first- and second-order false belief), affective (Reading the Mind in the Eyes) and composite measures (Tom's Taste) of theory of mind. Age differences were found in all objective but not the subjective measures, in line with deficiencies in self-evaluation in other domains. In the attribution of intention task, the second-order false belief task and the tom's taste task, older adults were worse than younger and middle aged adults. In the first-order false belief task and the recognition of complex emotions, older adults were worse only than younger adults. Interestingly there is evidence that the theory of mind impairment in autism spectrum disorder becomes smaller in older age when compared with neuro-typical individuals (Lever & Geurts, 2016). Taken together, in older age there seems to be a decline in mentalizing to a various degree depending on specific abilities.

2.1.2. Development and relevance of executive functions.

"The executive functions consist of those capacities that enable a person to engage successfully in independent, purposive, self-serving behaviour" (Lezak, Howieson, Loring, Hannay & Fisher, 2004, p.35).

> Executive functions (EFs; also called executive control or cognitive control) refer to a family of top-down mental processes needed when you have to concentrate and pay attention, when going on automatic or relying on instinct or intuition would be ill-advised, insufficient, or impossible. (Diamond, 2013, p.136).

These are only two, more pragmatic definitions of the umbrella term executive functions by noted scholars. Already more than twenty years ago Eslinger (1996) mentioned 33 definitions. But long before attempts were made to find definitions and models of the diverse processes and capacities subsumed by the term executive functions came the detailed case descriptions of patients with frontal lobe injuries –most notably Phineas Gage (Harlow, 1848; Macmillan, 2000) – that instigated the field. There has been a long standing debate on whether executive functions are in essence unitary or distinct (Teuber, 1972) which has largely been resolved to acknowledge existence of at least partly separable sub-functions (as I will discuss a bit later after a short digression).

The trend to fractionate executive functions mirrors a general development in psychological and neuropsychological science and assessment practice. Formalized neuropsychological assessment was driven by the need for the large scale examination of the cognitive ramifications of brain damaged during World War II (Russell, 2012). Brain damage was largely seen as a unitary phenomenon during the 30s, 40s and into the 50s of the 20th century supported by experiments of Karl S. Lashley (1880-1958). This view was reflected in the dominance of single function tests aimed to assess the presence or level of "organicity" of a patient like the Trail Making Test (U.S. War Department, 1944; Armitage, 1946). Later localizationism, driven by findings of neurosurgeons Wilder Penfield (1891-1976) and Theodore Rasmussen (1910–2002) became dominant. The endeavour of localizing brain damage with psychological assessment techniques however proved largely futile (Klebanoff, 1954) not least because of the realisation that patients with different lesion sites may exhibit similar impairments and vice versa (Lezak et al., 2004). Clinical neuropsychology eventually more ore less departed from the claim to localize brain damage and focused on refining the assessment of differentiated cognitive functions.

Early cognitive frameworks that instantiated executive control as a rather uniform capacity are the central executive of Baddeley's (1986) working memory model and the supervisory attentional system (Norman & Shallice, 1986). The study of dissociations (lesion *X* disrupts performance on task *I* but spares performance on task *J*) and double dissociations (in addition to the former observation: lesion *Y* disrupts performance on task *J* but spares performance on task *I*) in brain-damaged patients has shed light on the divisibility of executive functions (e.g. Shallice & Burgess, 1991) both behaviourally and anatomically. With scientific progress driven by lesion studies and brain imaging techniques, it became clear that abilities falling under the umbrella term of executive functions are served by distributed networks of different prefrontal sites, including posterior (mainly parietal) and subcortical regions (Collette, Hogge, Salmon, & van der Linden, 2006; Stuss & Alexander, 2007). A conceptualization of executive functions rooted in clinical work stems from Lezak et al. (2004) who distinguish four high level concepts of executive functions: volition, planning, purposive action and effective performance. Volition encompasses the forming of intentions, initiating of activity and self-awareness. Planning requires the identification and organization of steps to achieve a goal, to hold sequential and hierarchical ideas in working memory, inhibit impulses to act prematurely and overall sustained attention. After forming an intention and developing a plan, purposive action refers to the programming of activity which itself needs initiation and maintaining, switching between different intermediary steps and stopping the action sequence once the goal has been reached. Effective performance is achieved through monitoring and self-correction.

Efforts have been made to break down the conglomerate of executive functions into underlying core components. One framework supported by latent variable analysis that has gained widespread popularity in basic and applied science postulates three separable but interrelated core executive functions, the unity-but-diversity view (Miyake et al., 2000). These three core functions are: updating and monitoring of information (working memory), mental set shifting (cognitive flexibility) and the inhibition of prepotent responses and interfering stimuli (Miyake et al., 2000; Diamond, 2013). In recent years the independence of inhibition has been called into question however (Miyake & Friedmann, 2012; Munakata et al., 2011).

Executive functions start developing in infancy und undergo changes until old age (Diamond, 2013; Zelazo, Craik, & Booth, 2004). Compared to the large progress made in early and middle childhood (Welsh, 1991; Levin, 1991), developmental trajectories flatten subsequently but continue to show improvement into late childhood and adolescence depending on the subtype (Huizinga, Dolan, & van der Molen, 2006). Additionally there is behavioural and neuroimaging evidence that executive functions and prefrontal functional neural systems are less differentiated in younger than in older children (Isquith, Gioia, & Espy, 2004; Tsujimoto, 2008). In an attempt to ameliorate the problem of task impurity (task performance relies on more than one cognitive function to different degrees) authors have used latent variable analysis (e.g. principal component analysis) to approximate the hypothetical constructs of executive functions (e.g. Anderson, Anderson, Northam, Jacobs, & Catroppa, 2001; Brocki & Bohlin, 2004; Huizinga et al., 2006). *Inhibition* is believed to mature earlier than other executive functions at around 11 (Brocki & Bohlin, 2004; Huizinga et al., 2006) but has also been shown to develop until 14 years (Luna, Garver, Urban, Lazar, & Sweeney, 2004). For *shifting / cognitive flexibility* some research has found no change during adolescence (Anderson et al., 2001) while other studies indicate maturation around 15 years (Luna et al., 2004; Huizinga et al., 2006). Even so, improvement in reaction time on task-switch trials has been shown to progress into early adulthood (Cepeda, Kramer, & Gonzalez de Sather, 2001). *Working memory* shows a protracted development into late adolescence / early adulthood (Luna et al., 2004; Huizinga et al., 2006) although different aspects of working memory seem to mature at a different pace (Luciana, Conklin, Hooper, & Yarger, 2005). Lastly, *planning* (often operationalized with the Tower of London test), constituting a complex, higher level executive function which is influenced by the three aforementioned core functions, also shows development into late adolescence / early adulthood (Anderson et al., 2001; Huizinga et al., 2006).

As certain as development of executive functions continuous beyond the preschool years, the data on developmental trajectories or ages of maturation is inconclusive. This heterogeneity is at least in part due to different task complexities and scoring methods and response modality of the tasks used to operationalize the different executive functions. In summary, inhibition shows striking improvement during the preschool years and matures first. Shifting, working memory and planning emerge during the preschool years but the largest improvement takes place later with continuous development through adolescence. (Best, Miller, & Jones, 2009; Best & Miller, 2010). However, the question of functional maturation does not simply seem to be one of linear growth. There is evidence that the role of cognitive functions and their relation to each other changes during development. In a study with young children, the best predictor for problem solving was inhibition in children aged 2 years 8 months to 4 years but working memory in children aged 4 to 6 years (Senn, Espy, & Kaufmann, 2004). Another study (Huizinga & van der Molen, 2007) investigated developmental change of set-shifting and set-maintenance performance on the Wisconsin Card Sorting Test. In a principal component analysis, a single factor for set-shifting and set-maintenance emerged which was best explained by shifting in 11-year-olds, shifting and working memory in 15-year-olds and working memory in 21-year-olds.

Executive functioning contributes to health outcomes, is disturbed in many psychiatric conditions and is thought to play a role in the development and expression of a range of important abilities. Positive relations have been found with math and literacy skills in preschool children (Blair & Razza, 2007), longitudinally to overall academic achievement (Willoughby, Blair, Wirth, & Greenberg, 2012) to mathematics and English ability in school-age and to mathematics and science attainment in adolescence (Gathercole, Pickering, Knight, & Stegmann, 2004; for a review see Raghubar, Barnes, & Hecht, 2010). In regard to psychopathology alterations of some executive functions have been found for example in major depressive disorder (Taylor Tavares et al., 2007), obsessive compulsive disorder (Penades et al., 2007) and schizophrenia (Barch, 2005; Barch & Ceaser, 2012). Children with inhibition problems are more likely to be a bully or victim of bullying (Verlinden et al., 2013). Inhibition and sequencing ability predicts externalizing and internalizing behaviour longitudinally (Riggs, Blair, & Greenberg, 2003). There seem to be alterations in sensitivity to reward and punishment in conduct disorder, particularly the early-onset variant (Fairchild et al., 2009). Probably the most impressive evidence underlining the significance of self-regulation, a concept closely related to executive functions (Hofmann, Schmeichel, & Baddeley, 2012), comes from a longitudinal life-course study of 1000 children spanning over 30 years (Moffitt et al., 2011). Self-control in childhood predicted adult physical health, substance dependence, wealth and criminal convictions even after controlling for low intelligence and family socio-economic-status.

2.1.3. Interplay between theory of mind and executive functions.

Various accounts have been put forth to explain the link between executive functions and theory of mind. In this section I will briefly summarize the most popular ones.

Expression accounts (e.g. Leslie, German, & Polizzi, 2005; Baillargeon et al. 2010) try to explain the temporal gap between early competence on implicit tasks of social cognition and explicit theory of mind tasks. According to them and based on earlier studies (Mitchell & Lacoheé, 1991; Russel, Mauthner, Sharpe, & Tidswell, 1991) cognitive demands that (explicit) theory of mind tasks pose limit performance of the underlying competence. The large meta-analysis from Wellman et al. (2001) did show that one form of false-belief task modification (removing the object in question from the scene) had a small facilitating effect on performance. The authors argue however that this did not increase performance of young children above chance levels which is to be expected from an expression account. Furthermore Chinese children have been shown to outperform peers from the U.S. in executive functioning while not displaying superior theory of mind (Sabbagh Xu, Carlson, Moses, & Lee, 2006). Therefore it seems unlikely for executive functions as a limiting factor being the main constituent of the link between theory of mind and executive functions.

Conceptual-overlap or domain-general accounts posit that domain general cognitive abilities underlie the association between theory of mind and executive function tasks. According to the cognitive complexity and control theory (Frye, Zelazo, Palfai; 1995; Zelazo, Müller, Frye, Marcovitch, 2003), the ability to formulate and understand complex, embedded, conditional rules (e.g. if-if-then) is required in theory of mind (e.g. false-belief tasks) as well as executive functions tasks (e.g. Dimensional Change Card Sort test, DCCS; Zelazo, 2006) and explains the found association. Another conceptual overlap account, the re-description hypothesis (Kloo, Perner, & Giritzer, 2010), explains the relation in terms of requiring to understand that things can be described differently under different situations. These two theories however do not easily predict associations with executive function tasks that do not include complex conditional rules (e.g. simple inhibitory tasks like the Bear/Dragon task or planning tasks like the Tower of London) yet which are regularly found (Devine & Hughes, 2014).

Emergence accounts posit a deeper, functional connection between the development of executive functions and theory of mind. One competence enables or at least promotes the development (i.e. emergence) of the other. Perner and Lang (1999, 2000) first proposed that theory of mind facilitates meta-representational control. Most proponents of a functional connection however favour an account in which executive functions are a prerequisite for the emergence of theory of mind (Moses, 2001). A meta-analysis (Devine & Hughes, 2014) found evidence from longitudinal studies that early executive functions predict later false-belief understanding more strongly than vice versa, favouring the emergence account. One large, prospective study offers particular convincing evidence for this causal direction. In a sample of 226 children, executive function performance at 3 and 4 years predicted theory of mind at age 4 and 5 respectively, controlling for receptive language ability, sex and income-to-needs (Marcovitch et al., 2015). Executive functions have been shown to predict theory of mind even earlier. In one study, a composite executive function score at 24 months (T1) predicted the composite theory of mind score at 39 months (T2) after controlling for age, sex, verbal ability, maternal education, composite theory of mind score at T1 and parents' rated internal state language production at T1 (Carlson, Mandell, & Williams, 2004). In another, similar longitudinal study principal component analysis of executive measures suggested not one but two factors spatial working memory and conflict inhibition (Muller, Liebermann-Finestone, Carpendale, Hammond, & Bibok, 2012). Conflict inhibition (but not working memory) at age 2 and 3 predicted the aggregate theory of mind score at age 3 and 4, respectively, after controlling for age, verbal ability and prior theory of mind performance. Conversely in middle childhood there is evidence that working memory but not inhibition predicts performance in measures of advanced theory of mind (Lecce, Bianco, Devine, & Hughes, 2017). Interrestingly these disparate age findings are reminiscent of the study by Senn et al. (2004) that found problem solving to be more reliant on inhibition in toddlers and on working memory in pre-schoolers. Importantly, in the aforementioned studies, earlier theory of mind did not predict later executive function performance. Hughes & Ensor (2007) found evidence for bidirectional influences but a stronger effect for executive functions facilitating theory of mind. Another interesting approach to study development is a micro-genetic design. In this research method, behaviour is observed repeatedly and at short intervals, to better capture the process of change. A study following this approach found that improvement in inhibition skills preceded improvement in false belief understanding (Flynn, O'Malley, & Wood, 2004).

It should be noted that these different accounts do not necessarily preclude each other however. The main explanatory factor may be emergence with task demands playing a limited role in explicit false-belief understanding and some domains of executive functions (e.g. cognitive flexibility as measured with the Dimensional Change Card Sort test, Zelazo, 2006) sharing a larger common ground with theory of mind than others. To my way of thinking, it is important to remember that these conceptual views are located on a high level of abstraction and that the behaviour (including thoughts) subsumed under the term executive functions or theory of mind is realized by overlapping neural networks that share communalities yet also display considerable individual differences between members of our species, resulting in the fuzzy relationship between theories and data.

For second-order false belief reasoning the evidence regarding direction and mode of influence is dearth. There are however reasonable theoretical arguments to assume that the direction of influence goes primarily from executive functions and language to second-order theory of mind. Concerning emergence vs. expression accounts, the late appearance of second-order theory of mind in development as well as findings from older adults that show deficiencies in executive functions and theory of mind speak for an expression account as Miller (2009) points out.

2.1.4. A new measure of advanced theory of mind: Flexibility and Automaticity of Social Cognition (FASC).

Despite the prominence of false-belief tasks and the focus on first-appearance of mental state reasoning, several tasks to measure advanced theory of mind have been developed and used repeatedly in research (see chapter 2.1.1.6). However, the relation between these heterogeneous tasks has not been demonstrated in one sample until recently (see chapter 2.1.1.7). Hayward (2011) reviewed the literature on common advanced theory of mind tasks like second-order theory of mind, ambiguous figures, interpretive droodles, strange stories, faux-pas, Reading the Mind in the Eyes and found research on typically developing populations lacking. A comparison of the performance of 112 children between 7 and 13 years, controlling for verbal ability and inhibitory control, revealed that these tasks, against expectation, did not capture developmental variation and were largely unrelated. A lack of or very low associations between some of these tasks has been shown in several other studies before as well (Filippova and Astington, 2008; Brent et al., 2004). Some studies that have found some support for an association (Banerjee, 2000; Mitroff et al., 2006; Kaland, Callesen, Moller-Nielsen, Mortensen, & Smith, 2008) did not control for language, which has been found to be related to theory of mind performance, even in non-verbal tasks like the Reading the Mind in the Eyes task (Peterson & Miller, 2012). Another important question for which evidence is sparse is how sensitive these measures are to age differences in adolescence and beyond. Most previous studies only used the above mentioned tasks to investigate differences between normal and clinical populations (e.g. in the context of autism) and did not look at age differences in the control group. As discussed in chapter 2.1.1.8.1, data from the few studies that did present scores from different age groups of typically developing subjects, even if not analysing them statistically, paint a mixed picture (Miller, 2012).

According to Hayward, there are two factors commonly associated with theory of mind tasks yet frequently not controlled for, which may explain the contradictory results regarding inter-task correlations: the first factor is the presence or absence of language, the second factor is the degree of ambiguity (e.g. use or lack of social scripts) in the task material (henceforth referred to as *ambiguity*). To investigate the influence of these two variables and capture developmental change beyond middle childhood and variability in typically developing samples, Hayward et al. (2016) developed the FASC-task. This task consists of several vignettes depicting social situations similar to the Strange Stories (Happe, 1994) or Bosacki's social understanding stories (Bosacki & Astington, 1999; Bosacki, 2000). In contrast to these measures however, most information is conveyed visually in a cartoon-strip style. As mentioned, each vignette is defined by two categories: language (*verbal* or *nonverbal*) and ambiguity (*unambiguous* or *ambiguous*). Eight cartoons depicting social situations were created by Hayward with the comic web-tool Bitstrips (www.bitstrips.com; personal communication, February 09, 2012). Of these vignettes, one *verbal-* and one *nonverbal-ambiguous* vignette was based on the Kenny/Mark and Nancy/Margie story from Bosacki (2000) respectively, one *verbal-* and one *nonverbal-unambiguous* vignette was based on the Aunt Jane and the John item from Happe (1994) respectively and four were new. Subjects look at the cartoon for as long as they want and after having the vignette removed respond to the open-ended question: "Explain why the boy/girl behaves the way he/she does in the story". This prompt is repeated until no further response is made. Responses are captured on audio for later transcription. Dependent variables can be tailored to the specific research question but primarily consist of number of (unique) responses with/without mental references, number of (unique) internal state terms, overall response time, time taken after the prompt to initiate the first response and ratios between internal state terms and response time.

2.1.5. Goal of study 1 – Evaluating FASC Vienna as a tool to assess flexibility and automaticity of social cognition across the lifespan.

Goal of study one was to investigate age differences in mental state attribution across the lifespan with a new procedure (FASC) to assess social cognition developed by Hayward et al. (personal communication, February 09, 2012). To expand the number of item material to use for pre-post designs, we (Sprung et al., 2012) developed the FASC Vienna (henceforth only referred to as FASC) consisting of new vignettes based on the methodology of Hayward et al. (2016). Furthermore the influence of verbal cues and degree of ambiguity in the social vignettes on outcome variables of flexibility and automaticity of social cognition was of interest. Finally, relations to executive functions and verbal ability were to be explored.

2.2. Methods

2.2.1. Participants.

Forty subjects were recruited for each age group (children, adolescents, adults, older adults) for a total of 160. Children were recruited from second and third grades (7-9 years) of two primary schools and adolescents from third and fourth grades (12-14 years) of one secondary school in the state of Lower-Austria. Permission from the state education authority had to be acquired prior to obtaining informed consent from the parents. Adults between 25 and 45 were recruited by word-of-mouth. Older adults over 70 years were recruited from two retirement homes and additionally by word-of-mouth. Informed consent was obtained from all adults and older adults. Age characteristics of the sample are presented in Table 1.

Table 1
Age characteristics of the sample of study 1 in years

Age group	*n*	Min	Max	*M*	*SD*
Children	40	7.20	9.60	8.13	.57
Adolescents	40	12.13	14.06	13.15	.56
Adults	40	25.00	45.48	31.62	6.94
Older adults	40	69.76	91.85	79.40	6.23
Total	160	7.20	91.85	33.08	28.60

Gender among children and adults was almost evenly distributed ($\chi^2(1)=0.10$, $p=.752$ and $\chi^2(1)=0.00$, $p=1.00$, respectively). Among adolescents and older adults however females were over-represented ($\chi^2(1)=12.10$, $p=.001$ and $\chi^2(1)=9.26$, $p=.002$, respectively). In the adolescent-group this was due to a limited assent of males to participate in the study while the skewness in older adults can be partially explained by the natural demographic with women outliving men by 5.3 years on average (Statistik Austria, 2016). Table 2 shows the gender distribution of the sample. The effect of this uneven gender distribution were analysed and if present accounted for (e.g. through weighing and randomly selecting sub-samples; see section

2.2.4). It should be noted here however that only few variables were affected by gender differences.

Table 2
Gender distribution of the sample of study 1

Age group	*N*	female	male
Children	40	21	19
Adolescents	40	31	9
Adults	40	19	21
Older adults	40	30	10
Total	160	101	59

2.2.2. Procedure.

Data acquisition took place from September till November 2012 and was split between two dates because of the length which would have put too much strain on children and older adults. We aimed for a time interval of one week between the two test dates which could not be upheld in every case because of logistic and private reasons concerning schools and subjects. For three cases (2 adults, one older adult) the time interval was excessively short or large (58, 37 and 0 days respectively). Thus these cases were eliminated from analyses. For the remaining 157 subjects, modulus and mean was 7 days, with a range of 3 to 19 days, with 96% of intervals falling between 4 and 10 days. Tests were administered by four master students working on the project after having trained administration of the tasks. To ensure the FASC-procedure was followed as closely as possible, we obtained a sample video from Elisabeth Hayward, the author of the original FASC, showing her administering the task for training purposes. For details on the FASC-procedure see section 2.2.3.1. Tests of executive function were computerized and included a training phase thus maximizing administration objectivity. Clinical questionnaires were filled out by subjects and significant others on their own, except for youth self-report which was conducted as an interview (see section 2.2.3.6 for details). In the TUCA, items and questions are clearly listed on the test forms and can be read off the sheet. The other-report questionnaires were given the subjects to be filled out by a friend or relative and returned on the second test date.

Given the open ended nature of FASCs response format, task completion can take very long, depending on the richness of the subject's responses. To counter detrimental effects on motivation which might have led to ever shorter answers as the task drags on, the pool of vignettes (16) was split in half and administered over both test dates (8 vignettes, two for each category). Order of administration of vignettes (the same order was used at both times) was pseudo-randomized under constraints with the program Mix (Van Casteren & Davis, 2006). First the four FASC categories (verbal unambiguous, non-verbal unambiguous, verbal ambiguous, non-verbal ambiguous) were each given a digit code (1 to 4) and permutated resulting in 24 possible combinations of four-digit strings (e.g. 3421). One vignette per category per test date was determined to be in the first iteration and the other in the second (e.g. 3421 3421). This list of 24 combinations was then randomized under the following restraints: balancing categories presented first across the sample was deemed to be of particular importance since the subject would have no experience of the detailed task demands, in contrast to following turns. Thus the category being presented first was the repeated every 5th subject. As an additionally constraint to randomization, a category at 2nd to 4th position should repeat itself every 2nd to 7th permutation (subject). The randomization succeeded as position one and four was equally frequent for the four categories, while positions two and three differed only marginally in each age group (verbal ambiguous vignettes were placed 2nd for 41 subjects and 3rd for 39 subjects and vice versa for non-verbal ambiguous vignettes). Order of measures was fixed and as follows (see the following chapter for a detailed description of each task):

First test date:

1. Mini Mental Status Examination (only for older adults)
2. Flexibility and Automaticity of Social Cognition (part 1)
3. Flanker Inhibitory Control and Attention Test (NIH Toolbox Cognition Battery)
4. Dimensional Change Card Sort (NIHTB-CB)
5. Youth / Adult / Older-Adult Self Report
6. Child / Adult / Older-Adult Behavior Checklist
7. Selected vignettes from the Test of Understanding of Cognitive Activities (only for children; Sprung et al., 2012)

Second test date:

8. Flexibility and Automaticity of Social Cognition (part 2)
9. Digit Span from WISC-IV and WAIS
10. Vocabulary from WISC-IV and WAIS

2.2.3. Measures.

2.2.3.1. FASC Vienna: Flexibility and Automaticity of Social Cognition.

As mentioned before, new vignettes were created to use with the FASC procedure. All cartoons were conceived and created by master students in the project (Kerstin Ganglmayer, Julia Neudorfer, Ann-Kathrin Schmidt) with input from the author with the web-tool Bitstrips (www.bitstrips.com). For a verbal description of the social vignettes see chapter 2.2.3.1.2 to 2.2.3.1.5, for example vignettes see Appendix A. The general procedure for the FASC has already been described in chapter 2.1.4, but for the sake of clarity the most important points will be repeated here before describing the specifics of this study. During the procedure, subjects are handed a cartoon on paper and are told to look at the cartoon thoroughly for as long as they want and to indicate when they are finished looking. The experimenter then takes the comic away and asks "Explain why the boy/girl behaves the way he/she does in the story". After the subject has finished her response, the experimenter asks if she can think of another reason and after each response repeats the question until the subject signals that she can't think of another reason. The whole task was captured on audio with the open source audio recording and editing program Audacity (www.audacityteam.org) or iPhone's Voice Memo app to allow for exact transcription and to avoid influencing the subject's behaviour during the task by taking time or notes. For flexibility the outcome variables were: mean number of *unique mental state justifications* (MSJ) and mean number of *unique internal state terms* (IST). For automaticity/efficiency of social cognition the main variables of interest were mean *initial response time* (IRT) and *internal state terms ratio* (ISTr). The next section explains the variables in more detail.

2.2.3.1.1. Outcome Variables.

This section explains the outcome parameters of the FASC as defined in this study. Coding of mental state justifications and internal state terms was done in consensus by four graduate students under guidance of this author after training on sample data. Values for time variables were determined by zooming in on the audio track in the editing software Audacity, visually determining the acoustic beginning and secession of the targets overall response and reading out the timestamp of the selection in milliseconds.

Total responses: This is the total number of mental and non-mental justifications (i.e. explanations) for the character's behaviour in the cartoon, given by the respondent. There were several possibilities for a justification to be recognized as distinct. The response was counted if it was separated from preceding or succeeding responses by the experimenter's prompt. Utterances were also counted as a response if they were separated by words indicating the beginning of a unique justification such as "or", or "it is also possible that".

Mental state justifications: This is the total number of mental justifications given by the respondent, that is, responses that contain a reference to a mental state related to the prompt (e.g. not when talking about a personal experience or the state of the world in general as occasionally encountered with subjects from the older adults sub-group). Separate variables with and without repeated justifications were computed. Justifications were counted as repetitions when their meaning was analogous to a prior justification given even when worded slightly different (e.g. "She wanted it. It's because that's what she wanted".). If a different mental state term was used however, even when carrying the same meaning, the justification was counted as unique. Repetitions were extremely rare: 157 subjects produced 3848 justifications of which only 53 (1.38%) were counted as repetitions. A total of 2509 vignettes were responded to by the subjects of which only 45 (1.79%) responses contained repetitions. 121 subjects made no repetitions at all, 27 made one, seven made two, one made three and one made nine repetitions. Mean repetitions per subject and vignette where 0.021 (*SD* = 0.056). As a measure of flexibility only unique justifications are relevant thus from now on for sake of brevity unique mental state justifications are referred to as mental state justifications.

Unique internal state terms: The total number of unique internal state terms used in response to a vignette with repeated occurrences only counting once. This is a measure of flexible use of mental state language. Scoring followed the German coding-scheme by Klann-Delius (1998) which is based on the scheme by Bretherton & Beeghly (1982). Words regardless of class (e.g. noun, verb) were coded as internal state terms if they fell into one of the following categories: cognition (e.g. think), emotion (e.g. happy), volition (e.g. want), ability (e.g. try), obligation/permission (e.g. has to, may), morality (e.g. good as well-behaved) and physiology (e.g. hear, pain).

Overall response time: This is the time from the beginning of the first word of the first response to the end of the last word of the last response.

Initial response time (of mental justifications): This is the time from the end of the experimenter-prompt to the onset of the subject's response. Since not response speed in general but speed in attributing mental states was of interest, only responses including at least one mental justification were analysed. However response time still always referred to the first justification – mental or non-mental – since it could not be ascertained whether two responses might have been present in the mind of the subject when responding with one justification being queued. Indeed some participants gave several justifications immediately consecutive without pausing or having to be prompted.

Internal state term ratio: This is the sum of overall response times for each vignette divided through the sum of internal state terms used in response to a vignette, resulting in an inverse measure for the frequency of internal state terms use. To operationalize automaticity of social cognition, recurrences were counted to measure how heavily language was loaded with references to internal and mental states overall. A person referring to other people as *thinking* multiple times while not using any other mental state term might not display a varied mental state language but still displays a language more characterized by mental states than another person only mentioning the term *think* once in a statement of similar length. The decision to use the ratio of sums and not the mean of ratios (per vignette) was based on following reasoning: vignettes that did not receive an internal state term in their response would have had to be logically excluded from analysis (the division by zero is not defined) thus loosing valuable information (i.e. subjects responding but not using a single internal state term).

2.2.3.1.2. Description of verbal-unambiguous vignettes.

Visit: Susanne sits on her sofa looking tired after a hard day of work. A Woman that knows her knocks at the door. She tells Susanne that she was in the neighbourhood and hopes that she's not disturbing her. Susanne answers that she is happy to see her and offers her coffee. Prompt: Explain why Susanne behaves the way she does in the story.

Eric playing: Eric's Mother tells Eric to tidy up his room. Eric complies but when his mother's gone he keeps on playing. Eric hears his mother returning and sits down at his desk. When his mother remarks that he has not cleaned his room he says he had to do schoolwork for tomorrow. Prompt: Explain why Eric behaves the way he does in the story.

Birthday-dinner at the restaurant: A woman gets something to drink for three children celebrating birthday. When the woman is gone, one boy tells the other (Tom) that he likes coke and asks him what he likes. Tom says he doesn't like coke. The woman returns and tells them she has gotten coke for everybody and asks them if they like it. Otherwise she would get something else. The other boy cheers while Tom smiles half-heartedly and says he likes coke. Prompt: Explain why the boy behaves the way he does in the story.

Cake: Marie discovers the last piece of a cake in the kitchen. Just then Marie's mother enters the room with the neighbour. The Mother asks the neighbour if she wants a cake. The neighbour tells her that she'd gladly have it if nobody else wants the last piece. Marie looks at the last piece unhappily and hands it to the neighbour with a faltering smile. *Prompt:* Explain why Marie behaves the way she does in the story.

2.2.3.1.3. Description of nonverbal unambiguous vignettes.

Restaurant: A man, a woman and a boy are eating at a restaurant and are served by the waiter. When the waiter is gone the woman accidentally drops a bottle and breaks it. When the waiter returns with a broom the woman angrily points at the boy. Prompt: Explain why the woman behaves the way she does in the story.

Birthday present: A group of young people is celebrating. A girl holds a teddy-bear in her hands and drops it to the floor. A boy enters and gives the girl a present. The girl is thinking (indicated by a thought bubble) about a doll. She turns away, opens the present and when she discovers that the present is another teddy-bear looks disappointed. She turns to the boy again who is smiling and smiles too. Prompt: Explain why the girl behaves the way she does in the story.

Playing: Two children, a girl and a boy are playing in the same room next to each other. The girl is making a tower out of plastic dishes while the boy is playing with a toy train. When the girl leaves the room to get a woman the boy tosses over the tower. The girl returns and cries. Then the woman arrives, the girl points at the boy who is shrugging his shoulders and putting on an innocent face. Prompt: Explain why the boy behaves the way he does in the story.

Skirt: Two women walk on the street. One points at a violet skirt in a shop window, looks at her friend and makes sort of a disgusted face. Later they are sitting at a bar. When she sees a woman wearing the violet skirt she makes a gesture toward it smiling, while her friend is making a thumbs up and smiling too. Prompt: Explain why the two women behave the way they do in the story.

2.2.3.1.4. Description of verbal ambiguous vignettes.

Telephone: Anna is calling her friend Sarah and asking her if she wants to meet her today. Sarah, standing in her room looking sad, tells her that she cannot meet her because she is sick. Anna says that’s a shame and tells her to get well soon. On the next picture Anna sees someone looking like Sarah walking past her house and asks herself (thought bubble) if that is Sarah she is seeing. Prompt: Explain why Sarah behaves the way she does in the story.

Party: A couple of young adults are having a party in what looks like a club. Two women enter the room. One woman in a black dress (Maria) is leaning towards another woman and talking to her with her hand at her mouth as if she was whispering. Next Maria is calling Karin, one of the newly arrived women. They stretch out their arms for a welcome hug and Maria is saying to Karin that she is glad she came. Prompt: Explain why Maria behaves the way she does in the story.

Phone Call: In this vignette the verbal information is presented through thought bubbles since the characters are alone. Kathi is sitting in her room bored and thinking what to do. The idea comes to her mind that she could call Anna and ask her if she has time and she starts smiling. Anna is watching TV when the phone rings. She is thinking “the phone is ringing again” and move her hand towards the phone. But then she lets the phone keep ringing while sitting at her sofa with crossed arm. Prompt: Explain why Anna behaves the way she does in the story.

Restaurant visit: Elisa and Leo are in an elegant restaurant, sitting at a table with food and drinks. Elisa says that it looks delicious. Leo agrees and wishes bon appétit! Elisa realizes she does not have a fork. Leo says he will tell the waiter. He then calls the waiter who is standing not far away. The waiter walks past the couple. Elisa and Leo are looking upset and Leo asks Elisa what is wrong with the waiter. Prompt: Explain why the waiter behaves the way he does in the story.

2.2.3.1.5. Description of nonverbal ambiguous vignettes.

Skateboard: Two boys are skateboarding on the street. An old man skates by on his board in the distance and does a trick jump landing successfully. They look at each other wide-eyed while the man continues skateboarding. The boys then smile and wave at the old man who is looking in their direction. Prompt: Explain why the boys behave the way they do in the story.

Computer-class: Children are sitting in a class with computers on desks. There is one free desk. All computers are running except one at which a girl is sitting. Two boys enter the room. They look at the girl who is looking at her desk-neighbour looking distressed and shrugging her shoulders. The boys get closer standing next to the free desk which is situated next to the distressed girl and exchange whispers. Prompt: Explain why the boys behave the way they do in the story.

Swimming pool: There is a swimming pool with a lawn in front of it. A blond boy is handing a beach ball to a brunette boy. One girl is inside the pool and a second girl is walking towards a third girl at the edge of the pool. The blond boy is stretching his hands upwards while the brunette boy is preparing to throw the ball. The ball goes past the blond boy and towards the girl in the pool who looks scared, the brunette boy makes a triumphant gesture. On the last picture the ball is swimming in the pool next to the girl who still looks a little bit unnerved. The two boys stand close to each other with the blond boy looking shyly and the brunette boy laying his hand on him. The second and third girl at the edge of the pool continue standing there throughout the story. Prompt: Explain why the boys behave the way they do in the story.

Not eating: There is a children’s party with an adult woman standing behind a table with a cake. The woman calls out and starts handing out plates with cake. Finally everyone is standing there in groups of twos or threes with cake in hand except a boy and a girl standing side by side. Prompt: Explain why the boy and the girl behave the way they do in the story.

2.2.3.2. Executive Functions.

2.2.3.2.1. NIH Toolbox Dimensional Change Card Sort Test.

The Dimensional Change Card Sort Test (DCCS) (Zelazo, 2006), measures cognitive flexibility, one of the core constructs of executive function (Miyake et al., 2000; Diamond, 2013). Subjects are presented two target pictures that vary along two dimensions (i.e. shape and colour). A series of analogue test pictures appear and have to be matched to the target pictures according to one dimension (e.g. shape). After a while the target dimension for sorting changes (e.g. colour). So if the left target picture is a white rabbit and the right target picture is a brown boat and the matching rule is colour, and a brown rabbit appears as test picture, the subject has to press the right button to match the brown rabbit to the brown boat. In switch trials the required dimension to match changes in comparison to the last trial which places demands on cognitive flexibility. Accuracy and response times are transformed into vector-scores ranging from 0 to 5. If a subject achieves less than 80% correct, the accuracy vector equals the total score. If accuracy is 80% correct or higher, the response times vector is added to the accuracy vector for the total score to better differentiate in high performance ranges. Participants in this study completed a pre-release version of the DCCS (J. Anderson, personal communication, June 20, 2012) to be included in the NIH Toolbox Cognitive Battery (Weintraub et al., 2013) running on E-Prime® 2.0 (Psychology Software Tools, Inc., 2010) because the finished Toolbox (www.nihtoolbox.org) had not been released at that point of time. The measure is recommended for ages 3 to 85 years (Weintraub et al., 2013), thus it lends itself nicely for investigations of diverse age groups.

2.2.3.2.2. NIH Toolbox Flanker Inhibitory Control and Attention Test.

This task measures both inhibitory control and selective attention. It is based on the Eriksen Flanker Task from the Attention Network Test (Rueda et al., 2004). The subject is presented a central target arrow that can point left or right and two flanking arrows on each side that can all point in the same direction as the target arrow (congruent) or in the opposite direction (incongruent). The participant has to focus on the central arrow and press a corresponding key while ignoring the distracting, flanking arrows. For subjects ages 3-7, fish facing left or right, instead of arrows are presented. If the child participant achieves an accuracy of 90% or higher, additional trials with arrows are presented. Scoring is identical to the DCCS (see above) and recommended age is also 3 to 85 years. As with the DCCS above the measure was obtained as a pre-release version and running on E-Prime® 2.0 (Psychology Software Tools, Inc., 2010).

2.2.3.2.3. Digit Span Forward / Backwards.

Subjects completed the Digit Span subtest of the German version of the Wechsler Adult Intelligence Scale III (WAIS-III; Aster, Neubauer, & Horn, 2006) or the German version of the Wechsler Intelligence Scale for Children IV (WISC-IV; Petermann & Petermann, 2011), depending on their age. Digit Span is a measure of auditory short term memory, forming of series, concentration and working memory and consists of two sub-tasks: Digit Span forward and Digit Span backwards. Digit span backwards in particular requires rearranging information in working memory. The participant is read a series of numbers in one-second intervals and has to repeat the string afterwards in the same order or backwards.

2.2.3.3. Verbal intelligence / vocabulary.

Subjects completed the Vocabulary subtest of the German version WAIS-III or the WISC-IV, depending on their age. In Vocabulary, subjects are presented pictures of objects and words and have to explain what that object/word is. This test assesses word knowledge and concept formation. Furthermore it measures general knowledge, language development (particularly in children), learning aptitude and long term memory. Vocabulary correlates strongly with the verbal intelligence score (between .63 and .87 in the age-stratified norm population between 6 and 12 years) and the general intelligence quotient (between .61 and .77 in the age-stratified norm population between 6 and 16 years) (Petermann & Petermann, 2011, pp. 166-176) and thus is often used as a rough estimate of verbal intelligence or even intelligence in general. Subjects older than 89 years were compared with the age norm for ages 85-89.

2.2.3.4. Mini Mental Status Examination.

The Mini Mental Status Examination (MMSE) is a screening for cognitive impairment (Folstein, Folstein, & McHugh, 1975) and is conducted as an interview. The clinician assesses orientation, attention, memory, language and visuo-construction through a series of short questions and prompts. It was completed by older adults to screen for cognitive impairment which was an exclusion criterion.

2.2.3.5. Metacognition.

The Test of Understanding of Cognitive Activities (TUCA; Sprung et al., 2012; Maier, 2012) is an experimental task to assess metacognitive abilities in children (see Flavell, Green, & Flavell, 1995 for a review). It is a collection of existing tasks integrated into a coherent story to probe understanding of mental phenomena like the stream of consciousness (Flavell, Green, & Flavell, 1993) or limited controllability of mental activities (Flavell, Green, &

Flavell, 1998). Only subjects in the child age group completed the TUCA. Results are not relevant to this study and reported elsewhere (Neudorfer, 2014).

2.2.3.6. Clinical symptoms.

To assess clinical symptoms similarly in all age groups the Achenbach System of Empirically Based Assessment (ASEBA) was used. It consists of broad-band rating scales to assess clinical symptoms. Originally developed for children and adolescents, there are versions for preschool-aged children to older adults available today. For each scale, except for the preschool-age, a form for ratings obtained by parents, caregivers or spouses and a form for self-ratings is available. The school-age form and the adult form consist of following syndrome scales: anxious/depressed, withdrawn, somatic complaints, thought problems, attention problems, aggressive behaviour and rule-breaking behaviour. The first three scales make up the internalizing syndrome group while the last two scales make up the externalizing syndrome group. The school-age form also includes a social problems scale. The older-adults forms do not have the internalization / externalization syndrome groups and differ more strongly in terms of syndrome scales to reflect the important impact of neuro-cognitive and functional impairments in the life of older people. The syndrome scales of the older-adult forms are as follows: anxious/depressed, worries, somatic complaints, functional impairment, memory/cognition problems, thought problems, irritable/disinhibited. The total problem score in all forms is computed by summing all symptom-items. Therefore the total problem score will be the main score for cross-age analyses. For the Teacher Report Form (TRF), Youth Self-Report (YSR), Adult Behavior Checklist (ABCL) and Adults Self-Report (ASR) existing German translations were used (Arbeitsgruppe Deutsche Child Behavior Checklist, 1993, 1998, 2009, 2009, respectively). For the Older-Adult Behavior Checklist (OABCL) and the Older-Adult Self-Report (OASR) no German Translation was available. Most items of the older adult forms are identical with the adult forms and available translations were assumed unchanged. The author and two graduate students translated the remaining items. Originally the Youth Self-Report is intended for children aged 11 years and older but several studies support the validity of this measure in children younger than 11 years old (Kolko & Kazdin, 2003; Yeh & Weisz, 2001; Ebesutani, Bernstein, Martinez, Chorpita, & Weisz, 2011).

2.2.4. Data inspection and processing.

Data was evaluated for non-normality by visual inspection, calculating z-standardized skew and kurtosis, as well as with Shapiro-Wilk tests. Outliers were identified with box-plots and extreme outliers (pertaining to two-times the inter-quartile range) replaced by the next non-extreme-outlier value + one unit. This reduced skew and kurtosis to acceptable levels in most cases. Still, for FASC dimensions (e.g. verbal-ambiguous) distributions within groups deviated from normality for most factor levels with right skewed and platykurtic distributions (see Table C 1 to Table C 5 in Appendix C for details). Concerning robustness of ANOVA skew has little effect on power and error-rate when tests are two-tailed while platykurtic distributions make the test too conservative (Glass, Peckham, & Sanders, 1972), which can be deemed acceptable. Nevertheless parametric analyses were followed up with non-parametric tests to increase confidence in the finding. A conservative data analysis strategy was used where results were only deemed significant when both parametric and non-parametric analyses were significant. Because of the skewness displayed by the FASC variables, equality of variances was examined with a modified Levene's test using the median instead of the mean (Brown & Forsythe, 1974). An exception was the log-transformed mean initial response time which displayed asymptotic normality so the mean could be used. Additionally Hartley's F, the ratio of the largest and smallest variance across groups, was computed and compared to critical values (Hartley, 1950). However, when sample sizes are (nearly) equal, ANOVA is quite robust against heterogeneity of variance (Glass et al., 1972). Follow-up analyses (e.g. pairwise comparisons) however are not. Hence the ANOVA was still run when variances showed heterogeneity but post-hoc tests were not based on the pooled error term of the ANOVA but only on the groups of the specific comparison (Howell, 2002). Furthermore, main effects of age were backed by conducting additional Welch-ANOVAs and Games-Howell post-hoc tests.

2.3. Results

2.3.1. Descriptives.

Table 3

Descriptive statistics of associated variables over all age groups

Measure	*M*	*SD*
Vocabulary (r)	37.75	13.29
Vocabulary (S)	9.92	2.61
Digit Span (r)	14.85	3.19
Digit Span (S)	10.17	2.20
DCCS	7.23	2.05
DCCS (IQ)	103.53	15.81
Flanker	8.12	1.59
Flanker (IQ)	104.52	17.25
ASEBA self (r)	31.08	18.44
ASEBA self (T)	47.39	9.72
ASEBA other (r)	19.32	14.55
ASEBA other (T)	46.29	9.87

Note. r = raw score, S = Wechsler scaled score (*M*=10, *SD*=6), IQ = IQ-normed score (*M*=100, *SD*=15), T = T-score (*M*=50, *SD*=10); for details on the calculation of the DCCS and Flanker scores see 2.2.3.2

Table 4

Descriptive statistics of associated variables for separate age groups

	Children		Adolescents		Adults		Older Adults	
Measures	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Vocabulary (r)	23.30	5.33	41.80	5.87	50.50	10.60	35.97	12.16
Vocabulary (S)	9.05	2.24	9.20	1.80	10.95	3.22	10.54	2.56
Digit Span (r)	12.60	1.97	15.47	2.34	17.45	3.19	14.00	3.00
Digit Span (S)	10.13	1.95	9.45	2.04	10.34	2.56	10.77	2.08
DCCS	6.10	1.34	8.70	0.58	8.97	0.97	5.19	1.81
DCCS (IQ)	105.52	17.41	113.44	6.94	105.94	10.81	88.99	14.90
Flanker	7.38	1.04	9.25	0.62	9.40	0.63	6.42	1.42
Flanker (IQ)	107.81	16.51	112.85	11.31	109.35	10.31	87.48	17.58
ASEBA self (r)	22.82	13.76	30.14	16.99	35.37	17.44	36.03	22.03
ASEBA self (T)	42.08	8.30	46.39	9.61	48.66	7.33	52.38	10.61
ASEBA other (r)	13.94	10.90	12.44	12.45	22.00	12.16	30.55	16.31
ASEBA other (T)	43.08	9.27	42.50	12.62	47.84	6.74	52.69	6.46

Note. r = raw score, S = Wechsler scaled score (*M*=10, *SD*=6), IQ = IQ-normed score (*M*=100, *SD*=15), T = T-score (*M*=50, *SD*=10); for details on the calculation of the DCCS and Flanker scores see 2.2.3.2

Performance of older adults on the Mini Mental State Examination (MMSE) indicated normal cognition (>=24; Tombaugh & McIntyre, 1992) for all subjects (*Mdn* = 28 [25, 30]).

Table 5
Descriptive statistics for FASC outcome variables

	Overall		Children		Adolescents		Adults		Older Adults	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Flexibility										
MSJ	1.46	0.66	0.87	0.35	1.34	0.34	2.09	0.70	1.56	0.54
MSJ v-ua	1.52	0.62	1.05	0.38	1.40	0.40	1.95	0.71	1.71	0.55
MSJ v-a	1.64	0.78	0.98	0.50	1.49	0.52	2.38	0.74	1.75	0.60
MSJ nv-ua	1.34	0.71	0.83	0.46	1.24	0.48	1.97	0.77	1.37	0.61
MSJ nv-a	1.38	0.87	0.75	0.55	1.29	0.57	2.10	1.00	1.44	0.73
uIST	3.47	2.47	1.58	1.06	3.03	1.19	5.82	3.16	3.58	1.80
uST v-ua	4.13	2.62	2.09	1.36	3.59	1.38	6.28	3.07	4.69	2.37
uIST v-a	3.36	2.24	1.55	1.09	2.98	1.33	5.35	2.65	3.67	1.75
uIST nv-ua	3.37	2.83	1.40	0.98	3.03	1.35	6.14	3.85	3.06	1.92
uIST nv-a	2.96	2.46	1.31	1.16	2.47	1.09	5.28	3.34	2.90	1.71
Automaticity										
IRT	3.57	2.87	2.38	1.58	2.23	1.02	4.57	2.06	5.19	4.37
IRT v-ua	3.19	2.90	2.41	2.08	1.88	1.08	4.29	2.81	4.26	4.05
IRT v-a	3.82	4.10	2.28	1.98	2.52	1.52	4.49	2.66	6.10	6.74
IRT nv-ua	3.70	4.27	2.29	2.21	1.95	1.48	4.36	3.30	6.28	6.68
IRT nv-a	3.59	3.02	2.41	1.83	2.48	2.07	5.04	2.89	4.37	3.95
IRT log	3.45	0.29	3.31	0.25	3.30	0.22	3.62	0.19	3.60	0.33
IRT log v-ua	3.37	0.34	3.26	0.33	3.20	0.26	3.56	0.27	3.47	0.37
IRT log v-a	3.43	0.35	3.25	0.27	3.32	0.28	3.58	0.28	3.59	0.41
IRT log nv-ua	3.39	0.39	3.21	0.36	3.19	0.29	3.53	0.32	3.62	0.39
IRT log nv-a	3.42	0.35	3.26	0.35	3.28	0.31	3.62	0.29	3.51	0.34
IST-ratio	6.83	2.98	8.76	4.48	5.64	1.76	5.81	1.97	7.07	1.54
IST-ratio v-ua	5.40	2.00	6.47	2.53	4.65	1.67	4.85	1.47	5.63	1.65
IST-ratio v-a	6.70	2.89	8.00	4.18	6.05	2.18	5.88	1.80	6.84	2.36
IST-ratio nv-ua	8.23	4.40	11.06	5.90	6.53	2.71	6.48	3.67	8.95	3.00
IST-ratio nv-a	7.96	4.24	10.42	6.78	6.33	2.40	6.80	2.55	8.57	2.82

Note. MSJ = mean mental state justifications, uIST = mean unique internal state terms, IRT = mean initial response time (sec.), IST-ratio = ratio of sum of overall response time (sec.) to sum of IST, v-ua = verbal unambiguous, v-a = verbal ambiguous, nv-ua = non-verbal unambiguous, nv-a = non-verbal ambiguous

2.3.2. Gender differences.

Gender differences were not a focus of this research. However, because of unequal distribution of gender in the adolescents and older-adults age groups, outcome variables were subjected to independent samples Welch's t-tests to investigate to what extent results from different age groups could be compared (for all results see Table C 6 in the Appendix). For adolescents, boys and girls differed on (z-transformed) Flanker scores ($t(32.73)=-3.73$, $p=.001$, $r=.55$), Flanker T-norms ($t(31.36)=-3.44$, $p=.002$, $r=.52$), (z-transformed) ASEBA other-report raw scores ($t(9.01)=-2.98$, $p=.016$, $r=.70$) and ASEBA other-report T-norms ($t(18.09)=-3.64$, $p=.002$, $r=.65$). For older adults, gender differences for overall (z-transformed) internal state terms ratio ($t(16.64)=-2.61$, $p=.018$, $r=.53$), and nonverbal-unambiguous (z-transformed) internal state terms ratio ($t(13.60)=-2.91$, $p=.012$, $r=.66$)

emerged. Additionally log-transformed mean initial response time for nonverbal-unambiguous vignettes was different in males and females ($t(14.34)=-2.68$, $p=.018$, $r=.60$). To account for these differences I decided to analyse variables in which gender differences were found for adolescents or older adults using weighed cases. Unfortunately SPSS's command "weigh by" rounds to the next integer, resulting in an inflated sample size in this data scenario (since weights for female subjects are rounded up to 1). Therefore, after assigning the weights, a sub-sample of female subjects was randomly selected to hold constant the sample size while obtaining a balanced gender weight ratio (details of the sub-sampling are described under the corresponding section of analyses).

2.3.3. Correlative analyses.

Because distribution of most variables proved to differ from normality quite strongly, associations between FASC-variables and other constructs of interest (vocabulary, executive functions, and clinical symptoms) were investigated with Spearman rank-correlations (Table 6). Mind that variables of automaticity of social cognition (IRT; ISTr) have to be interpreted inversely because they are based on time variables (i.e. small values of ISTr – small latencies between internal state terms in the responses - indicate a high automaticity of social cognition). Gender was highly unevenly distributed in adolescents and prior t-test revealed age-differences in this age group on Flanker and ASEBA other-report variables. Consequently weights for male (n=9) and female (n=31) subjects were computed (2.22 and 0.65, rounded by SPSS to 2 and 1 respectively). Therefore, after assigning the weights, a sub-sample of female adolescents was randomly selected (22 out of 31) to hold constant the sample size (22*1 females + 9*2 males = 40) while obtaining a closed to balanced gender ratio.

Table 6
Rank-correlations (Spearman) between FASC and constructs of interest

Measures	n	MSJ	IST	IRT	ISTr
Vocabulary (r)	157	.61**	.59**	.30**	−.25**
Vocabulary (S)	157	.37**	.32**	.30**	.02
Digit Span (r)	157	.38**	.40**	.12	−.25**
Digit Span (S)	157	.08	.08	.11	.02
DCCS	157	.38**	.41**	−.05	−.44**
DCCS (IQ)	157	−.05	−.03	−.29**	−.31**
Flanker	156	.30**	.34**	−.02	−.31**
Flanker (IQ)	156	−.08	−.04	−.19*	−.12
ASEBA self (T)	152	.26**	.30**	.17*	.00
ASEBA other (T)	137	.10	.06	.13	−.05

Note. r = raw score, S = Wechsler scaled score (M=10, SD=6), IQ = IQ-normed score (M=100, SD=15), T = T-score (M=50, SD=10)

There is a striking pattern with executive function raw scores showing significant correlations with almost all FASC variables which disappear when looking at executive function norms (computing correlations with sub-sampling and gender-weighting only slightly changed some effect sizes and yielded the same significant results; see Table C 7). This most likely reflects group differences where children achieve low scores in both executive function raw scores and FASC variables and adults scoring high in both areas. As this might simply mirror general cognitive development and does not specifically answer if flexibility or automaticity of social cognition is related to verbal intelligence, executive function or clinical symptoms, relative performance, controlling for age, has to be looked at. Age normed scores (as routinely used in psychological tests) are an elegant way to do this but there are no norms for FASC available yet. A partial correlation can account for common variance shared with age but assumes a linear relationship which descriptive statistics did not indicate. To look at the association between relative performance in vocabulary, executive functions, clinical symptoms and FASC, ignoring differences between age groups, variables were z-standardized for each age group separately and correlations with the pooled z-variables computed (Table 7).

Table 7
Correlations between FASC and constructs of interest, z-standardized within each age group, pooled across age groups

Measures	n	MSJ (z)	uIST (z)	IRT (z)	ISTr (z)
Vocabulary (z)	157	.29**	.23**	.15	.01
Digit Span (z)	157	.11	.14	−.01	−.04
DCCS (z)	157	.15	.19*	−.17*	−.33**
Flanker (z)	156	.16*	.16	−.07	−.03
ASEBA self (z)	152	.17*	.22**	−.01	−.02
ASEBA other (z)	137	−.04	−.09	−.14	−.09

Note. MSJ = mean mental state justifications, uIST = mean unique internal state terms, IRT = mean initial response time, ISTr = ratio of sum of overall response time to sum of IST

A similar pattern to the correlations with the norm variables emerged for vocabulary, DCCS, and self-rating of ASEBA with smaller effect sizes throughout. Interestingly, additional positive correlations emerged between the DCCS and internal state terms and Flanker and mental state justifications. All correlations with initial reaction time except with the DCCS vanished. Computing correlations with sub-sampling and gender-weighting only slightly changed some effect sizes and yielded the same significant results (see Table C 8) except for the association between z-transformed internal state terms and DCCS score which just became significant ($r = .16$, $p = .049$).

To investigate a possible effect of the two FASC-dimensions language and ambiguity on correlations with vocabulary, executive functions and clinical symptoms, the four initial

FASC-categories were merged into new variables for each outcome measure (e.g. mental state justifications, internal state terms, etc.): verbal (unambiguous and ambiguous), non-verbal (unambiguous and ambiguous), unambiguous (verbal and non-verbal) and ambiguous (verbal and non-verbal) cartoons. Separate correlations were computed for these variables and correlations were compared by means of confidence intervals of differences (Zou, 2007) between categories (verbal / non-verbal, unambiguous / ambiguous) utilizing the web-version of the R-package *Cocor* (Diedenhofen & Musch, 2015). All comparisons were non-significant (i.e. the confidence intervals included zero) meaning that correlations between constructs of interest and FASC variables did not differ in size for verbal vs. non-verbal or unambiguous vs. ambiguous cartoons (see Table 8).

Table 8
Confidence intervals (95%) of differences between correlations

	Vocabulary (z) [a]		DCCS (z) [a]		Flanker (z) [b]		ASEBA self (z) [c]	
Variables	lower	upper	lower	upper	lower	upper	lower	upper
MSJ (z) language	−.064	.148	-	-	−.061	.157	−.111	.111
MSJ (z) ambiguity	−.126	.062	-	-	−.007	.186	−.033	.162
uIST (z) language	−.039	.095	−.132	.004	-	-	−.017	.120
uIST (z) ambiguity	−.165	.091	−.154	.105	-	-	−.245	.017
IRT (z) language	-	-	−.155	.112	-	-	-	-
IRT (z) ambiguity	-	-	−.057	.223	-	-	-	-
ISTr (z) language	-	-	−.119	.125	-	-	-	-
ISTr (z) ambiguity	-	-	−.142	.074	-	-	-	-

Note. Confidence intervals (95%; Zou, 2007) for differences of correlations between constructs of interest and FASC-variables of verbal − non-verbal (language) and unambiguous − ambiguous (ambiguity) cartoons. Comparisons were only made if the correlations over all FASC-categories were significant (see previous Table 7) so some fields are left empty.
[a] n=157, [b] n=156, [c] n=152

To ascertain if the pattern of associations was consistent across age groups, significant correlations were subsequently investigated group-wise (see Table 9). Results of correlative analyses for separate age groups show that the positive association between vocabulary and mental state justifications is significant for adults and older adults while internal state terms are significant for older adults only. However, associations were positive for all age groups and reached trend level for adults (zIST: p=.054) and adolescents (zMSJ: p=.052). The pattern of association for the DCCS with positive correlations for variables of flexibility and negative correlations with variables of automaticity (interpretable as a positive association between cognitive flexibility and automaticity of social cognition) is carried by different age groups. While only adults show a medium to large positive correlation between cognitive flexibility and flexibility of social cognition, negative correlations with internal state term ratio are more consistent across age groups with adolescents and older adults displaying significant correlations and children reaching trend level (zISTr: *p*=.063). The flanker task is only significantly correlated with mental state justifications in the adult age group. The only notable correlations regarding ASEBA self-report are with mental state justifications and internal state terms for children. Computing correlations with sub-sampling and gender-weighting for adolescents only slightly changed some effect sizes and yielded the same significant results (see Table C 9).

Table 9
Correlations between FASC-variables and constructs of interest, z-standardized within each age group

Measures	n	MSJ (z)	uIST (z)	IRTm (z)	ISTr (z)
Vocabulary (z)					
Children	40	.15	.07	−.08	.02
Adolescents	40	.31	.11	.22	.25
Adults	38	.37*	.32	.34*	−.14
Older Adults	39	.34**	.43**	.11	−.12
DCCS (z)					
Children	40	.16	.17	−.14	−.30
Adolescents	40	−.04	.13	−.23	−.49**
Adults	38	.38*	.34*	−.03	−.20
Older Adults	39	.13	.12	−.28	−.32*
Flanker (z)					
Children	40	.19	.19	.03	.13
Adolescents	40	.15	.15	−.03	−.13
Adults	38	.32*	.15	−.13	−.02
Older Adults	38	−.01	.13	−.15	−.08
ASEBA self (z)					
Children	39	.45**	.48**	.24	.12
Adolescents	36	.08	.12	.16	−.10
Adults	38	−.03	.07	−.25	−.05
Older Adults	39	.17	.20	−.20	−.05

Note. MSJ = mean mental state justifications, uIST = mean unique internal state terms, IRT = mean initial response time, ISTr = ratio of sum of overall response time to sum of IST

2.3.4. Flexibility and automaticity of social cognition across the life span.

2.3.4.1. Data analysis.

Performance in social cognition was expected to differ between age groups. Therefore, significant main effects of age group and interaction effects including age group were followed up with planned contrasts, comparing each age group with the previous (i.e. repeated contrasts). Significant two- and three-way interactions were decomposed with simple interaction and second-order simple effects analyses utilizing the GLM procedure, LMATRIX and MMATRIX subcommands, as laid out in Howell & Lacroix (2012). Because the log of 0 is not defined, response times that had been coded with 0 milliseconds (because the subject responded before the experimenter had finished the prompt) were recoded to 1 milliseconds before taking the log of 10 which conveniently scaled these earliest responses to 0 again.

2.3.4.2. Unique mental state justifications.

Median based Levene's test of equality of error variances was significant for all within-subject factor combinations (language, ambiguity) variances indicating heterogeneity across age groups ($F(3, 153) = 3.95$, $p = .010$; $F(3, 153) = 2.92$, $p = .036$; $F(3, 153) = 3.87$, $p = .011$; $F(3, 153) = 5.05$, $p = .002$; for verbal- unambiguous, verbal-ambiguous, nonverbal-unambiguous, nonverbal-ambiguous respectively). Variance ratio (Hartley, 1950) was also above the recommended criterion of about 2.5 for most factor-combinations (verbal unambiguous = 3.56, verbal ambiguous = 2.37, nonverbal unambiguous = 2.80, nonverbal ambiguous = 2.77).

2.3.4.2.1. Main effects.

There was a significant main effect of age group on the number of unique mental state justifications produced ($F(3, 153) = 33.4$, $p < .001$). Planned contrasts comparing each age group with the previous were all significant (Bonferroni corrected alpha value of .017) with adolescents providing more mental state justifications than children ($D = −.45$, 95% CI [−.69, −.22], $p = .002$), adults achieving a higher score than adolescents ($D = −.74$, 95% CI [−.98, −0.50], $p < .001$) and a decline from adults to older adults ($D = .53$, 95% CI [.29, .77], $p < .001$), resulting in an inverted u-shape across age groups. Welch's ANOVA comparing age groups confirmed the initial ANOVA ($F_W(3, 81.8) = 37.9$, $p < .001$) and results of post-hoc multiple comparisons (Games-Howell) were in line with previous contrasts (children vs. adolescents: $t(80) = 6.00$, $p < .001$, $r = .56$; adolescents vs. adults: $t(78) = 6.01$, $p < .001$, $r = .56$; adults vs. older adults: $t(77) = 3.71$, $p = .002$, $r = .40$). There was a significant main effect of language ($F(1, 153) = 47.6$, $p < .001$) with more responses generated for verbal than for nonverbal vignettes. There was also a significant main effect of ambiguity ($F(1, 153) = 7.66$, $p = .006$) with more responses generated for ambiguous compared to unambiguous vignettes.

In the nonparametric analysis the Kruskal-Wallis test for the main effect of age group was significant ($\chi^2(3) = 72.58$, $p < .000$). Follow-up U-tests (Bonferroni corrected alpha value of .017) again confirmed results from parametric analysis that adolescents ($Mdn = 1.34$) produced more mental state justifications than children ($Mdn = 0.88$, $U = 313.50$, $z = −4.69$, $p < .001$, $r = −.52$), adults ($Mdn = 1.91$) more than adolescents ($U = 253.50$, $z = −5.07$, $p < .001$, $r = −.57$) and older adults ($Mdn = 1.44$) fewer than adults ($U = 426.00$, $z = −3.21$, $p = .001$, $r = −.37$). Significant main effects for language and ambiguity were also confirmed through Wilcoxon signed-rank tests. Verbal cartoons elicited more mental state justifications than nonverbal cartoons ($Mdn_{\text{verb}} = 1.50$, $Mdn_{\text{nverb}} = 1.25$, $T = 1988$, $p < .000$, $r = -.49$) and ambiguous cartoons more than unambiguous ones ($Mdn_{\text{amb}} = 1.38$, $Mdn_{\text{nonamb}} = 1.38$, T = 3578, p = .025, r = −.18) despite being characterized by the same median.

2.3.4.2.2. Two-way interaction effects.

There were no significant interactions between language and either age group ($F(3, 153) = 1.82$, $p = .147$) or ambiguity ($F(3, 153) = 2.97$, $p = .087$). There was a significant interaction effect between ambiguity and age group ($F(1, 153) = 6.19$, $p = .001$). Pairwise comparisons (Bonferroni corrected alpha value of .013) suggested that only adults produced more mental state justifications for ambiguous than for unambiguous vignettes (children: M_{unamb}= 0.94, SE_{unamb}= 0.06, M_{amb} = 0.87, SE_{amb} = 0.08, $t(39) = 1.70$, $p = .096$, $r = .26$; *adolescents*: M_{unamb}= 1.32, SE_{unamb}= 0.06, M_{amb} = 1.39, SE_{amb} = 0.08, $t(39) = -1.23$, $p = .226$, $r = .19$; adults: M_{unamb}= 1.96, SE_{unamb}= 0.11, M_{amb} = 2.24, SE_{amb} = 0.13, $t(37) = -3.37$, $p = .002$, $r = .48$; older adults: M_{unamb}= 1.54, SE_{unamb}= 0.08, M_{amb} = 1.59, SE_{amb} = 0.10, $t(38) = -1.15$, $p = .256$, $r = .18$; also see Figure 4). Repeated contrasts (Bonferroni corrected alpha value of .013) revealed that the difference of mental state justifications produced for ambiguous vs. unambiguous vignettes was not significantly different between children and adolescents (D = .28, 95% CI [−.04, 0.60], $p = .086$) but larger for adults than for adolescents (D = .42, 95% CI [.10, 0.75], $p = .012$) and older adults (D = −.44, 95% CI [−.77, −0.12], $p = .008$).

Significant parametric analyses were again repeated with non-parametric tests to ascertain confidence in the findings. A difference variable of mental state justifications from ambiguous and unambiguous vignettes (unambiguous − ambiguous) was computed. The Kruskal-Wallis test of overall age differences was significant ($\chi^2(3) = 14.28$, $p = .003$). Pairwise comparisons (Bonferroni corrected alpha value of .013) between unambiguous and ambiguous vignettes for each age group with the Wilcoxon signed-rank test were consistent with results from the t-tests (children: Mdn_{unamb}= 0.88, Mdn_{amb} = 0.88, $T = 175.50$, $p = .094$, $r = -.26$; adolescents: Mdn_{unamb}= 1.25, Mdn_{amb} = 1.38, $T = 223.0$, $p = .199$, $r = -.260$ adults: Mdn_{unamb}= 1.81, Mdn_{amb} = 2.13, $T = 116.5$, $p = .003$, $r = -.48$; older adults: Mdn_{unamb}= 1.50, Mdn_{amb} = 1.50, $T = 256.0$, $p = .222$, $r = -.20$). Follow-up Mann-Whitney U-tests (Bonferroni corrected alpha value of .017) comparing successive age groups on the difference-variable contradicted parametric contrasts in that no comparison reached significance after Bonferroni correction (children vs adolescents: Mdn_{children} = 0.13, $Mdn_{\text{adolescents}}$ = −0.13, $U = 570.00$, $z = -2.23$, $p = .026$, $r = -.25$; adolescents vs. adults: Mdn_{adults} = −0.25, $U = 571.50$, $z = -1.89$, $p = .058$, $r = -.21$; adults vs. older adults; $Mdn_{\text{older-adults}}$ = −0.13, $U = 538.50$, $z = -2.08$, $p = .038$, $r = -.24$).

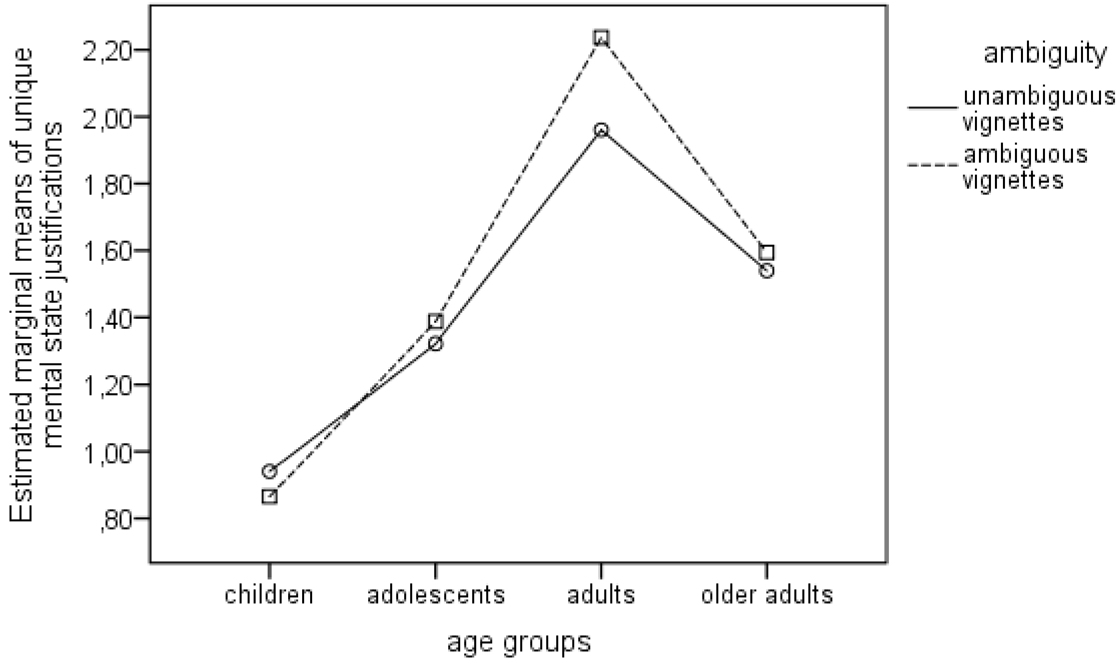

Figure 4. Two-way interaction effect of mixed ANOVA between age groups and ambiguity on mental state justifications

2.3.4.2.3. Three-way interaction effect.

The interaction between language, ambiguity and age group was non-significant ($F(3, 153) = 2.35, p = .074$).

2.3.4.3. Unique internal state terms.

Variances across age groups showed considerable heterogeneity as revealed by median based Levene's test ($F(3, 153) = 7.72, p < .001$; $F(3, 153) = 6.53, p < .001$; $F(3, 153) = 12.14, p < .001$; $F(3, 153) = 13.01, p < .001$; for verbal-unambiguous, verbal-ambiguous, nonverbal-unambiguous, nonverbal-ambiguous respectively). Variance ratio (Hartley, 1950) was also above the recommended criterion of about 2.5 for most factor-combinations (verbal unambiguous = 5.12, verbal ambiguous = 5.93, nonverbal unambiguous = 15.43, nonverbal ambiguous = 9.31).

2.3.4.3.1. Main effects.

There was a significant main effect of age group on the number of unique mental state terms used ($F(3, 153) = 31.59, p < .001$). Repeated contrasts (Bonferroni corrected alpha value of .017) mirrored the inverted u-shape of unique mental state justifications across age groups with an increase from children to adolescents ($D = -1.43$, 95% CI [−2.28, −0.58], p=.001) and adolescents to adults ($D = -2.74$, 95% CI [−3.61, −1.88], $p < .001$) and a decrease from adults to older adults ($D = 2.19$, 95% CI [1.32, 3.01], $p < .001$). Welch's

ANOVA comparing age groups confirmed the initial ANOVA ($F_W(3, 80.61) = 30.61$, $p < .001$) and results of post-hoc multiple comparisons (Games-Howell) were in line with previous contrasts (children vs. adolescents: $t(80) = 5.76$, $p < .001$, $r = .54$; adolescents vs. adults: $t(78) = 5.11$, $p < .001$, $r = .50$; adults vs. older adults: $t(77) = 3.81$, $p = .002$, $r = .39$). The main effect of language was also significant with descriptive statistics showing more unique mental state terms being produced for verbal than nonverbal vignettes ($F(1, 153) = 67.51$, $p < .001$). There was also a significant main effect for ambiguity where responses for unambiguous vignettes contained significantly more unique mental state terms than responses to ambiguous vignettes ($F(1, 153) = 62.66$, $p < .001$).

Analysis of main effects of age group was repeated with the nonparametric Kruskal-Wallis test and significant ($\chi^2(3) = 67.81$, $p < .000$). Follow-up U-tests (Bonferroni corrected alpha value of .017) confirmed that adolescents (*Mdn* = 2.75) used more internal state terms than children (*Mdn* = 1.28, $U = 277.00$, $z = -5.03$, $p < .001$, $r = -.56$), adults (*Mdn* = 4.44) more than adolescents ($U = 283.00$, $z = -4.77$, $p < .001$, $r = -.54$) and older adults (*Mdn* = 3.19) fewer than adults ($U = 406.50$, $z = -3.41$, $p = .001$, $r = -.39$). Significant main effects for language and ambiguity were also confirmed through Wilcoxon signed-rank tests. Verbal cartoons elicited more internal state terms than nonverbal cartoons ($Mdn_{\text{verb}} = 3.25$, $Mdn_{\text{nverb}} = 2.63$, $T = 1727$, $p < .001$, $r = -.58$) and unambiguous cartoons more than ambiguous ones ($Mdn_{\text{unamb}} = 3.25$, $Mdn_{\text{amb}} = 2.75$, $T = 1795$, $p < .001$, $r = -.58$).

2.3.4.3.2. Two-way interaction effects.

Two-way interactions between language and age group ($F(3, 153) = 10.37$, $p < .001$), as well as language and ambiguity ($F(1, 153) = 7.92$, $p = .006$) were significant. The interaction between ambiguity and age group was non-significant ($F(1, 153) = 2.45$, $p = .066$). For the Language x Age group interaction (see Figure 5), pairwise comparisons (Bonferroni corrected alpha value of .013) indicated that all age groups except adults produced fewer internal state terms in response to nonverbal compared to verbal vignettes (children: $M_{verb} = 1.82$, $SE_{verb} = 0.18$, $M_{nverb} = 1.35$, $SE_{nverb} = 0.16$, $t(39) = 5.74$, $p < .001$, $r = .68$; adolescents: $M_{verb} = 3.29$, $SE_{verb} = 0.20$, $M_{nverb} = 2.75$, $SE_{nverb} = 0.18$, $t(39) = 5.64$, $p < .001$, $r = .67$; adults: $M_{verb} = 5.82$, $SE_{verb} = 0.44$, $M_{nverb} = 5.71$, $SE_{nverb} = 0.57$, $t(37) = 0.48$, $p = .635$, $r = .08$; older adults: $M_{verb} = 4.18$, $SE_{verb} = 0.31$, $M_{nverb} = 2.98$, $SE_{nverb} = 0.28$, $t(38) = 9.05$, $p < .001$, $r = .83$). Repeated contrasts (Bonferroni corrected alpha value of .017) revealed that the difference between verbal and non-verbal cartoons was larger for older adults ($D = -2.19$, 95% CI [−2.98, −1.39], $p < .001$) compared to adults while there was no significant difference for children vs. adolescents (*D*=−.14, 95% CI [−0.92, −0.63], *p* = .715) or adolescents vs adults (*D* = 8.64, 95% CI [0.78, 1.65], *p* = .031).

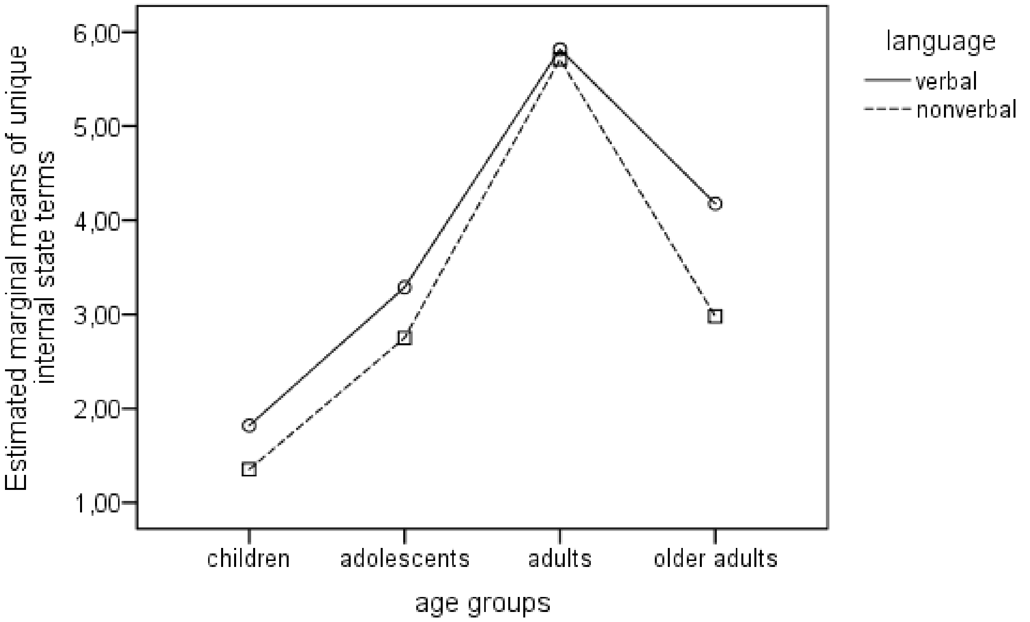

Figure 5. Two-way interaction effect of mixed ANOVA between age groups and language on internal state terms

For the Language x Ambiguity interaction, the difference between unambiguous and ambiguous vignettes was larger in verbal than in nonverbal vignettes, as can be seen in Figure 6.

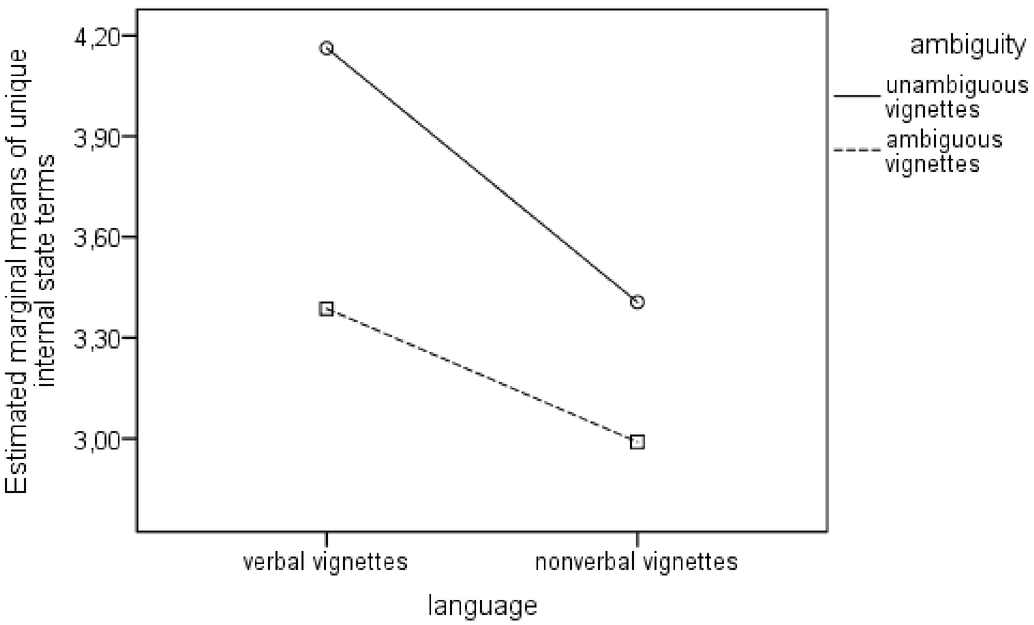

Figure 6. Two-way interaction effect of mixed ANOVA between ambiguity and language on internal state terms

To test the two-way interaction between language and age group non-parametrically, a variable of the difference of internal state terms between verbal and nonverbal vignettes (verbal – nonverbal) was computed and subjected to a Kruskal-Wallis-test which confirmed that age groups differed significantly regarding internal state terms between verbal and nonverbal vignettes ($\chi^2(3) = 24.81, p < .001$). Pairwise comparisons (Bonferroni corrected alpha value of .013) between verbal and nonverbal vignettes for each age group with the Wilcoxon signed-rank test were in line with results from the t-tests (children: $Mdn_{\text{verb}} = 1.63$, $Mdn_{\text{nverb}} = 1.31$, $T = 35.50$, $p < .001$, $r = -.71$; adolescents: $Mdn_{\text{verb}} = 2.94$, $Mdn_{\text{amb}} = 2.75$, $T = 73.50$, $p < .001$, $r = -.68$ adults: $Mdn_{\text{verb}} = 4.81$, $Mdn_{\text{nverb}} = 4.50$, $T = 320.50$, $p = .468$, $r = -.12$; older adults: $Mdn_{\text{verb}} = 3.88$, $Mdn_{\text{nverb}} = 2.63$, $T = 11.50$, $p < .001$, $r = -.83$). Follow-up Mann-Whitney U-tests (Bonferroni corrected alpha value of .017) comparing consecutive age groups again confirmed the result from parametric contrasts for children ($Mdn = 0.38$) vs. adolescents ($Mdn = 0.44$, $U = 749.50$, $z = -0.49$, $p = .626$, $r = -.05$), adolescents vs. adults ($Mdn = 0.38$, $U = 610.00$, $z = -1.50$, $p = .134$, $r = -.17$) and adults vs. older adults ($Mdn = 1.13$, $U = 353.00$, $z = -3.96$, $p < .000$, $r = -.45$). To test for the Language x Ambiguity interaction, two difference variables between unambiguous and ambiguous were computed, one for verbal and one for nonverbal vignettes (i.e. verbal-unambiguous – verbal-ambiguous; nonverbal-unambiguous – nonverbal-ambiguous). A Wilcoxon signed-rank test lent support to the parametric result that the difference in internal state terms between unambiguous and ambiguous vignettes was larger for verbal than for nonverbal cartoons ($Mdn_{\text{vs–va}} = 0.50$, $Mdn_{\text{ns–na}} = 0.25$, $T = 3615$, $p = .003$, $r = -.24$).

2.3.4.3.3. Three-way interaction effects.

The three-way interaction between age group, language and ambiguity was non-significant ($F(3, 153) = 2.22$, $p = .088$).

2.3.4.4. Initial response time.

Eight subjects (five children, one adolescent, two older adults) were not included in the analysis because they did not contribute mental justifications in at least one category (hence no initial response time was computed), resulting in a total sample size of 149. Gender was highly unevenly distributed in older adults and prior t-test revealed age-differences in this age group on the initial response time variable. Consequently weights for male (n=9) and female (n=28) subjects were computed (2.06 and 0.66, rounded to 2 and 1 by SPSS respectively). Therefore, after assigning the weights, a sub-sample of female older adults was randomly selected (18 out of 28) to hold constant the sample size of older adults (18*1 females + 9*2 males) while obtaining a balanced gender ratio.

Variances of log-transformed mean initial response times did not differ significantly across age groups ($F(3, 150) = 1.60$, $p = .191$; $F(3, 150) = 2.28$, $p = .081$; $F(3, 148) = 1.16$, $p = .328$; $F(3, 144) = 0.62$, $p = .606$; for verbal-unambiguous, verbal-ambiguous, nonverbal-unambiguous, nonverbal-ambiguous respectively). Variance ratio (Hartley, 1950) was below the recommended criterion of about 2.5 for all factor-combinations (verbal unambiguous = 1.93, verbal ambiguous = 1.96, nonverbal unambiguous = 1.75, nonverbal ambiguous = 1.45).

2.3.4.4.1. Main effects.

There was a significant main effect of age group on the lg10-transformed mean response times ($F(3, 144) = 25.61$, $p < .001$). Planned repeated contrasts comparing consecutive age groups (Bonferroni corrected alpha value of .017) showed that adolescents did not differ from children ($D = -0.01$, 95% CI [−0.12, 0.10], $p = .918$), adults were slower in responding than adolescents ($D = -0.32$, 95% CI [−0.43, −0.21], $p < .001$) and older adults did not differ from adults ($D = -0.04$, 95% CI [−0.14, 0.07], $p = .526$). There was also a significant main effect of ambiguity ($F(1, 144) = 7.10$, $p = .009$) with slower responses for ambiguous vignettes. There was no significant main effect for language ($F(1, 144) = 0.70$, $p = .405$).

2.3.4.4.2. Two-way interaction effects.

There was no significant two-way interaction (Language x Age group: $F(3, 144) = 2.28$, $p = .082$; Ambiguity x Age group: $F(3, 144) = 1.37$, $p = .253$; language x scripts: $F(1, 144) = 3.46$, $p = .065$).

2.3.4.4.3. Three-way interaction effects.

The interaction between age group, language and ambiguity was significant ($F(3, 144) = 6.01$, $p = .001$). Figure 7 and Figure 8 suggest an interaction for nonverbal items and ambiguity between adults and older adults. Follow up simple-interaction effects (Bonferroni corrected alpha value of .008) indicated that for verbal items there was no significant interaction between ambiguity and age group ($D = 0.08$, 95% CI [−0.07, 0.22], $p = .316$; $D = -0.10$, 95% CI [−0.24, 0.05], $p = .193$; $D = 0.15$, 95% CI [0.003, 0.30], $p = .046$; repeated contrasts for age groups). At the nonverbal level however, the interaction between ambiguity and adults vs. older adults was significant (contrasts for children vs. adolescents: $D = 0.03$, 95% CI [−0.12, 0.19], $p = .660$; contrasts for adolescents vs. adults: $D = 0.01$, 95% CI [−0.15, −0.16], $p = .940$; adults vs. older adults: $D = -0.28$, 95% CI [−0.43, −0.12], $p = .001$). Second-order simple effects (Bonferroni corrected alpha value of .025) revealed that older adults were slower in responding to nonverbal-unambiguous ($D = -0.22$, 95% CI [−0.38,

−0.07], p = .006) but not to nonverbal-ambiguous (D = 0.03, 95% CI [−0.09, 0.20], p = .461) vignettes.

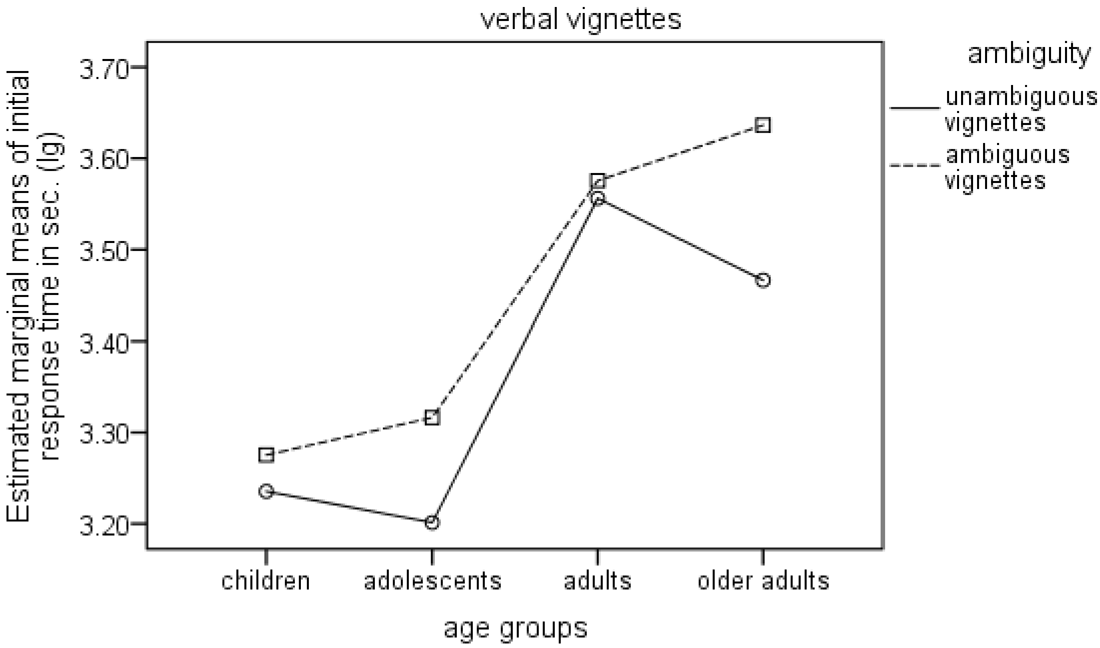

Figure 7. Three-way interaction effect of mixed ANOVA between age groups and ambiguity on initial response time for verbal vignettes

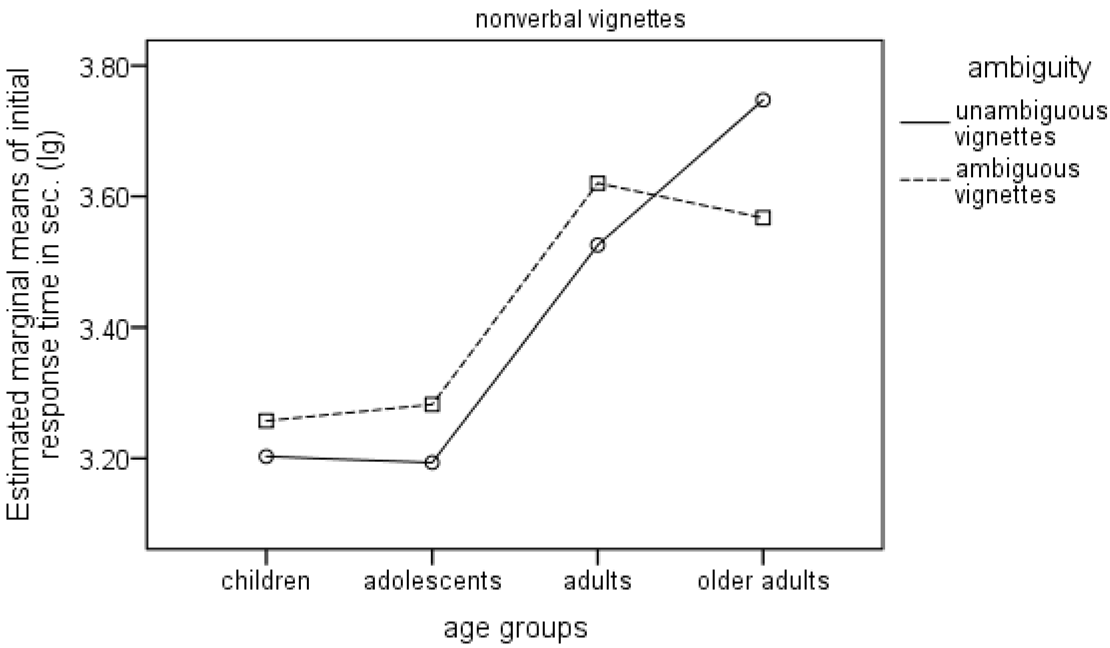

Figure 8. Three-way interaction effect of mixed ANOVA between age groups and ambiguity on initial response time for nonverbal vignettes

2.3.4.5. Internal state term ratio.

In the analysis of the internal state term ratio, the sample size was reduced from 157 to 150 because five children and two older adults did not contribute responses containing internal state terms and therefore no ratio could be computed. Since initial t-test had revealed gender differences for older adults, analysis was conducted using the same gender-weighting and random sampling as described in the results section for initial response time.

Median based Levene's test was significant for all within-subject factor combinations (language, ambiguity) indicating serious heterogeneity of variances ($F(3, 150) = 4.20$, $p = .007$; $F(3, 150) = 6.85$, $p < .001$; $F(3, 148) = 7.02$, $p < .001$; $F(3, 145) = 11.63$, $p < .001$; for verbal-unambiguous, verbal-ambiguous, nonverbal-unambiguous, nonverbal-ambiguous respectively). Variance ratio (Hartley, 1950) was also well above the recommended criterion of about 2.5 for all factor-combinations (verbal unambiguous = 2.95, verbal ambiguous =5.41, nonverbal unambiguous = 4.74, nonverbal ambiguous = 7.96).

2.3.4.5.1. Main effects.

All main effects were significant (age group: $F(3, 145) = 11.54$, $p < .001$; language: $F(1, 146) = 94.66$, $p < .001$; ambiguity: $F(3, 146) = 14.03$, $p < .001$) with smaller time intervals between internal state terms for verbal compared to nonverbal and for unambiguous compared to ambiguous vignettes. Planned repeated contrasts (Bonferroni corrected alpha value of .017) comparing consecutive age groups revealed that mean time between internal state terms was significantly lower for adolescents compared to children ($D = 2974$, 95% CI [1783, 4164], $p < .001$), and higher for older adults compared to adults ($D = -1867$, 95% CI [−3064, −671], $p = .002$) but that there was no significant difference between adolescents and adults ($D = -116$, 95% CI [−1281, 1049], $p = .844$). Welch's ANOVA comparing age groups confirmed the initial ANOVA ($F(3, 81.75) = 10.76$, $p < .001$) and results of post-hoc multiple comparisons (Games-Howell) were in line with previous contrasts (children vs. adolescents: $t(80) = 4.09$, $p = .001$, $r = .42$; adolescents vs. adults: $t(78) = 0.40$, $p < .979$, $r = .05$; adults vs. older adults: $t(74) = 3.58$, $p < .003$, $r = .38$).

In the ensuing non-parametrical re-analysis, the Kruskal-Wallis test for age groups was significant ($\chi^2(3) = 24.05$, $p < .000$). Follow-up U-tests (Bonferroni corrected alpha value of .017) confirmed each contrast: internal state term ratio was smaller for adolescents ($Mdn = 5458$) than children ($Mdn = 7499$, $U = 443.00$, $z = -3.44$, $p = .001$, $r = -.38$) and smaller for adults ($Mdn = 5077$) than for older adults ($Mdn = 7388$, $U = 365.00$, $z = -3.45$, $p = .001$, $r = -.40$) while adolescents and adults did not differ significantly ($U = 743.50$, $z = -0.17$, $p = .869$, $r = -.02$). Significant main effects for language and ambiguity were also confirmed through Wilcoxon signed-rank tests. For verbal cartoons internal state term ratio was smaller than for nonverbal ones ($Mdn_{\text{verb}} = 5565$, $Mdn_{\text{nverb}} = 7263$, $T = 1290$, $p < .000$, $r = -.68$) and it was smaller for unambiguous than for ambiguous cartoons (Mdn_{unamb}= 6319, Mdn_{amb} = 6993, T = 3771, p < .000, r = −.32) again supporting initial parametric analyses.

2.3.4.5.2. Two-way interaction effects.

There was a significant two-way interaction for Language x Age group (F(3, 145) = 6.48, p < .001). Pairwise comparisons (Bonferroni corrected alpha value of .013) revealed that all age groups used internal state terms significantly sparser in nonverbal than in verbal vignettes (children: M_{verb} = 7397, SE_{verb} = 483, M_{nverb} = 11059, SE_{nverb} = 977, t(37) = −5.25, p < .001, r = .65; *adolescents*: M_{verb} = 5346, SE_{verb} = 271, M_{nverb} = 6426, SE_{nverb} = 375, t(39) = −5.08, p < .001, r = .63; adults: M_{verb} = 5364, SE_{verb} = 229, M_{nverb} = 6641, SE_{nverb} = 452, t(37) = −3.75, p = .001, r = .52; older adults: M_{verb} = 6311, SE_{verb} = 296, M_{nverb} = 9429, SE_{nverb} = 480, t(35) = −6.64, p < .001, r = .74). Repeated contrasts (Bonferroni corrected alpha value of .017) revealed that the difference between verbal and nonverbal vignettes was larger in children than in adolescents (D = −4070, 95% CI [−6531, −1608], p = .001) and older adults than in adults (D = 3682, 95% CI [1208, 6156], p = .004) but there was no significant difference for adolescents vs adults (D = 394, 95% CI [−2015, 2803], p = .747; also see Figure 9).

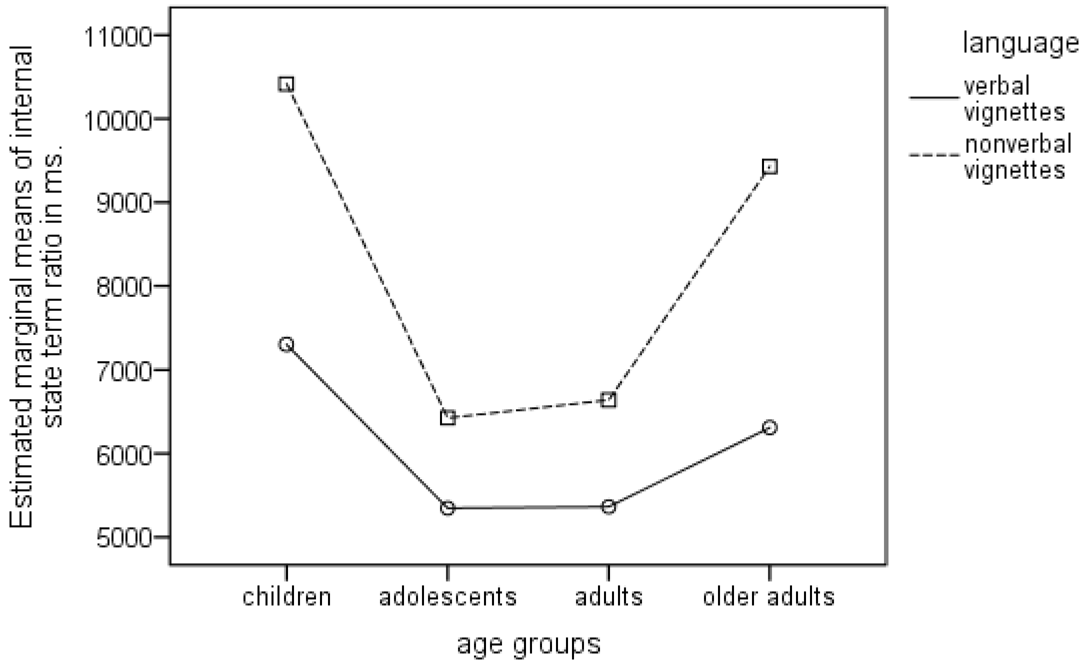

Figure 9. Two-way interaction effect of mixed ANOVA between age groups and language on internal state term ratio

Furthermore there was a significant interaction between language and ambiguity (F(1, 145) = 20.57, p < .001) with a larger difference between unambiguous and ambiguous vignettes for verbal than for nonverbal vignettes (see Figure 10). Pairwise comparisons

(Bonferroni corrected alpha value of .025) between unambiguous and ambiguous were only significant for verbal ($M_{verbal\text{-}unamb}$ = 5378, $SE_{verbal\text{-}unamb}$ = 162, $M_{v\text{-}a}$ = 6756, $SE_{v\text{-}a}$ = 238, $t(153) = -7.01$, $p < .001$, $r = .49$) but not for nonverbal vignettes ($M_{n\text{-}s}$ = 8227, $SE_{n\text{-}s}$ = 351, $M_{n\text{-}a}$ = 8060, $SE_{n\text{-}a}$ = 352, $t(148) = .31$, $p = .534$, $r = .05$). The interaction between ambiguity and age group was non-significant ($F(3, 145) = 0.29$, $p = .836$).

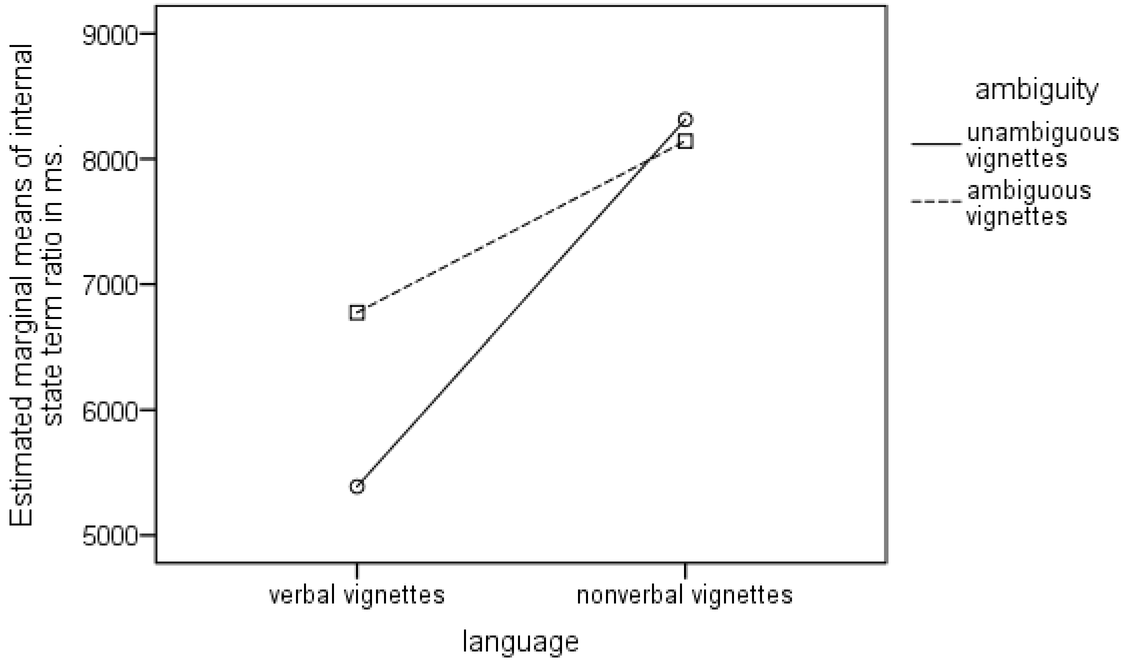

Figure 10. Two-way interaction effect of mixed ANOVA between language and ambiguity on internal state term ratio

To test non-parametrically whether differences between verbal and nonverbal vignettes varied across age, a difference variable (verbal – nonverbal) was computed. The Kruskal-Wallis test for age groups was significant ($\chi^2(3) = 14.00$, $p = .003$). Pairwise comparisons (Bonferroni corrected alpha value of .013) between verbal and nonverbal vignettes for each age group with the Wilcoxon signed-rank test confirmed all significant differences (children: Mdn_{verb} = 6501, Mdn_{nverb} = 10019, T = 78.00, $p < .001$, $r = -.69$; adolescents: Mdn_{verb} = 5080, Mdn_{nverb} = 6094, T = 109.00, $p < .001$, $r = -.64$ adults: Mdn_{verb} = 5099, Mdn_{nverb} = 6003, T = 168.00, $p < .001$, $r = -.48$; older adults: Mdn_{verb} = 5647, Mdn_{nverb} = 8761, T = 17.00, $p < .001$, $r = -.83$). Mann-Whitney U-tests (Bonferroni corrected alpha value of .017) comparing successive age groups were consistent with the parametric analysis insofar as the difference between verbal and nonverbal vignettes was larger for older adults ($Mdn = -2663$) than for adults ($Mdn = -987$, U = 403.00, $z = -3.04$, $p = .002$, $r = -.35$) and non-significant between adolescents ($Mdn = -1251$) and adults (U = 717.50, $z = -0.43$, p = .671, $r = -.05$). In contrast to the parametric analysis however, the comparison for children

($Mdn = -1974$) and adolescents did not reach significance ($U = 560.00$, $z = -2.00$, $p = .046$, $r = -.23$). To follow up the Language x Ambiguity interaction non-parametrically, two difference variables between unambiguous and ambiguous were computed, one for verbal and one for nonverbal vignettes (i.e. verbal-unambiguous – verbal-ambiguous; nonverbal-unambiguous – nonverbal-ambiguous). A Wilcoxon signed-rank test confirmed that the difference in internal state terms ratio between unambiguous and ambiguous vignettes was larger for verbal than for nonverbal cartoons ($Mdn_{vs-va} = -1007$, $Mdn_{ns-na} = -65$, $T = 3208$, $p < .001$, $r = -.37$). Wilcoxon signed-rank tests (Bonferroni corrected alpha value of .025) supported results from t-tests that internal state term ratio was only different for verbal unambiguous vs verbal ambiguous vignettes ($Mdn_{vs} = 5091$, $Mdn_{va} = 6200$, $T = 2383.00$, $p < .001$, $r = -.52$) but not for nonverbal unambiguous vs nonverbal ambiguous vignettes ($Mdn_{ns} = 7006$, $Mdn_{na} = 7199$, $T = 5456.00$, $p = .803$, $r = -.02$).

2.3.4.5.3. Three-way interaction effects.

There was no significant three-way interaction between age group, language and ambiguity was non-significant ($F(3, 145) = 0.96$, $p = .415$).

2.3.4.6. Summary.

Variables of flexibility of social cognition (number of unique mental state justifications, number of unique internal state terms) showed an increase from children to adolescents to adults and a decrease from adults to older adults. Verbal compared to nonverbal cartoons elicited more mental state justifications over all age groups and more unique internal state terms in children, adolescents and older adults. Unambiguous vignettes overall elicited more unique internal state terms. This effect was larger for verbal than for nonverbal vignettes. Ambiguous vignettes elicited more mental state justifications over the whole sample but pairwise comparison was only significant for adults. However, repeated contrasts did not indicate a significant difference for adults compared to adolescents or older adults in nonparametric analysis.

In terms of automaticity, initial response time was characterized by a markedly slower response in adults compared to adolescents while children compared to adolescents as well as adults compared to older adults did not differ. Responses to unambiguous vignettes were faster in average than responses to ambiguous vignettes. The internal state term ratio, which is the average time interval between internal state terms used and thus represents an inverse measure of frequency per time (smaller values indicating a more frequent internal state term use per time interval), showed a relationship between age groups similar to the variables of flexibility. In other words, internal state terms relative to response length were more frequent in adolescents than in children and also more so in adults than in older adults but there was no significant difference between adolescents and adults. As with the total number of unique internal state terms, relative internal state term use (as indicated by the internal state term ratio) was less frequent in response to nonverbal than to verbal cartoons in all age groups. The effect was larger in children than in adolescents however. Internal state term ratio was smaller in unambiguous vignettes than in ambiguous ones but only for verbal cartoons. Table 10 shows a visual representation of the coherence between parametric and nonparametric results.

Table 10
Summary and comparison of parametric and nonparametric results

		Ag			L	A	Ag x L			Ag x S			LxA	Agx LxS
		c-ad	ad-a	a-oa			c-ad	ad-a	a-oa	c-ad	ad-a	a-oa		
MSJ	P													
	NP													
uIST	P													
	NP													
IRT	P													
ISTr	P													
	NP													

Note. Greyed out fields indicate a significant result. Dark-grey narrow bars code omnibus age results while the three separate cells below it code follow-up contrasts of children vs. adolescents (c-ad), adolescents vs. adults (ad-a), adults vs. older adults (a-oa). MSJ: mean mental state justifications, uIST: mean unique internal state terms, IRT: mean initial response time, ISTr: ratio of sum of overall response time to sum of IST; Ag: main effect of age, L: main effect of language, A: main effect of ambiguity; other columns show interaction effects

2.4. Discussion

2.4.1. Flexibility of social cognition over the lifespan.

Flexibility of social cognition (both number of mental state justifications and internal state terms) followed an inverted u-shape across age groups showing an increase from childhood to adolescence and adulthood and a decline in older adulthood. In respect to children and adolescents, this replicates findings from Hayward et al. (2016) in regard to mental justifications while they did not find the expected age effect for mental state terms found in this study. Similar developmental trajectories with an increase in performance in childhood and a decrease in old adulthood have been found for hindsight bias (Bernstein, Erdfelder, Meltzoff, Peria, & Loftus, 2011) and executive functions (Zelazo et al. 2004). In the study about hindsight bias however, visual hindsight bias significantly declined from 3 to

5 years, did not change until young adulthood and only significantly differed again between young and older adults. In the present study, there was a significant difference in social cognitive flexibility between children and adolescents suggesting that the task captures even later developments in mental understanding. The results suggest that flexibility of social cognition continues to develop from childhood through adolescence into young to middle adulthood but also shows a decline in old adulthood.

2.4.2. Automaticity of social cognition over the lifespan.

Internal state term ratios (how frequently, relative to the length of responses, internal state terms are used when talking about social situations) showed a similar relationship across age groups like variables of flexibility with adolescents displaying more automaticity compared to children and adults compared to older adults and no difference between adolescents and adults. Interestingly Hayward et al. (2016) found a difference between children and adolescents for a related variable of automaticity, the ratio of response time to mental state justifications. They did not find a difference between younger and older adolescents. This indicates that automaticity of social cognition seems to peak in early adolescence, earlier than flexibility. The decline in old adulthood could also be (in part) explained with executive task demands. To reach a high score in the internal state term ratio, one has to stay on task and elaborate the reasons for the behaviour of the characters in the cartoon.

The second variable of automaticity yielded unexpected results. Children and adolescents were quicker to respond after the prompt than adults and older adults. If initial response time was a measure of automaticity of social cognition one would expect a decrease at least until adulthood. A possible explanation is that this variable is more indicative of response styles than of mental understanding. There was no instruction to respond as fast as possible so different task strategies might have been employed by the subjects. While some participants might have prepared an initial response before answering, others might have started answering right away and formulated their response as they spoke. In this vein the larger delay between prompt and response in adults and older adults compared to children and adolescents might mirror a more deliberate response strategy. Remotely similar effects have been observed for switch-costs in cognitive flexibility tasks. Shifting (i.e. switching) between different tasks or mental sets regularly incurs a cost in response speed or accuracy in comparison to non-shift trials termed the switch-cost. In a study by Davidson et al. (2006) the switch cost in accuracy declined between 9- and 13-year-olds, while the switch cost in response speed actually increased between 6-year-of-age and adulthood. The authors reason that a speed-accuracy trade-off is in effect, with older children as well as adults compromising their response times in order to ensure high accuracy. In contrast, Hayward and colleagues (2016) found a significant difference for 12- compared to 8-year-olds and a marginally significant difference for 16- compared to 12-year-olds to respond faster to FASC prompts. They did not investigate adult performance however. Further studies need to explore influencing factors to the initial response time. For example, instructions could be experimentally manipulated to examine the influence of explicit prompts to start responding as soon as possibly on different age groups.

Taken together, the internal state term ratio was developmentally sensitive while the initial response time used with the present mode of instruction might not be the best variable to capture automaticity of social cognition.

2.4.3. Role of verbal information.

Perhaps the most robust finding concerning item characteristics was that language seems to provide powerful cues to promote flexibility and automaticity of social cognition. That is, cartoons with verbal content (e.g. dialogue, short descriptions) elicited a higher number of mental justifications containing a larger variety of internal state terms, as well as more frequent use of internal state terms. Adults however produced just as much internal state terms for nonverbal as for verbal cartoons, supporting the assumption of peak performance regarding the dependence of mentalizing on verbal cues in this age group.This replicates findings of Hayward et al. (2016) who also found a similar effect for language and internal state terms and mental justifications. Concerning speed of initial responses there was no influence of verbal information, corroborating findings from Hayward et al. (2016). Language however did increase automaticity in terms of relative frequency of internal state term use.

2.4.4. Role of ambiguity.

The case was more complex for the dimension of ambiguity however. Cartoons depicting ambiguous social situations (e.g. due to lack of contextual information) elicited more unique mental justifications than vignettes following a social script (e.g. white lie). Evidently the constructed cartoons were successful to differ in terms of ambiguity and this ambiguity opened up the possibility for a variety of interpretations reflected in the responses (i.e. interpretive diversity). Notably, this effect was only significant for adults suggesting a protracted development of social cognition into adulthood. This is in line with Hayward et al. (2016) who also did not find an effect of ambiguity in children and adolescents. At the same time, this interpretive diversity did not translate in an equally rich mental state talk. Mental justifications for ambiguous vignettes contained fewer unique internal state terms, again a replication of findings from Hayward et al. (2016). One possibility is that the cartoons based on social scripts or customs are more typical, subjects had more experience with them and thus could talk about associated mental states more vividly. Research in forensic psychology has shown that made-up accounts are less rich in details than true stories (Undeutsch, 1967; Sporer, 2004). The ambiguous situations might also have produced a higher cognitive load, which reduces performances of effortful long-term-memory retrieval (Moscovitch, 1994).

Initial responses to ambiguous vignettes were slower than responses to unambiguous vignettes (again coinciding with results from Hayward et al., 2016). Internal state terms were also sparser during the responses (i.e. internal state term ratio). This can be seen as evidence that vignettes constructed to be ambiguous indeed were more equivocal and thus subjects took longer to formulate mental justifications rich in mental state terms.

2.4.5. Theory of mind, verbal ability and executive functions.

Theory of mind and social cognition in general have been linked to other constructs like verbal competence and executive functions.

Unsurprisingly, verbal comprehension was most strongly associated with flexibility of social cognition. Verbal abilities and theory of mind have been consistently linked. Interestingly the size of the correlation was very similar to the one found in the meta-analysis by Milligan et al. (2007) that looked at false-belief reasoning.

Cognitive flexibility displayed a medium strength association with automaticity of social cognition and a small strength association with flexibility of social cognition over all age groups. The correlation with the internal state term ratio was fairly consistent although strongest for adolescents and not significant for adults and only marginally for children. Regarding flexibility of social cognition the medium to strong correlation for adults stood out, while for the other age groups it was nonsignificant and only of small size.

The finding that cognitive flexibility is associated with FASC variables of social cognition is in line with other recent studies. Henning (2001) found that 3- to 6-year-old children's performance on the DCCS related to their overall performance on the ToM scale but that this relation was specific to those ToM tasks that tap children's understanding of epistemic states such as knowledge access, diverse beliefs, and false beliefs regarding content and location. In a study methodologically similar to this work, cognitive flexibility (also measured with a version of the DCCS) but not inhibition or working memory predicted social understanding (as assessed with the strange stories task) in 7- to 12-year-old children (Bock, Gallaway and Hund, 2015). Interestingly we also did not find a significant correlation between inhibition (flanker task) or working memory (digit span) and the FASC: The link between cognitive flexibility and social cognition has also been found using more distantly related constructs and methodologically dissimilar tasks. Kocsis-Bogar, Kotulla, Maier, Voracek and Hennig-Fast (2017) found that cognitive flexibility (measured via the Trail Making Test B) was associated with theory of mind performance (operationalized with a video based measure of social cognition) in healthy adult subjects both low and high on schizotypy. In another study, flexibility but not inhibition was related to external emotion understanding (inferring situational causes of emotions) in preschool children (Martins, Osorio, Verissimo, and Martins, 2016). The consistency of this link over a range of different methods for both cognitive flexibility and socio-emotional understanding supports the validity of the FASC as a measure of social cognition.

The heterogeneous results in this study raise several questions however: First, why is a measure of cognitive flexibility more strongly associated with automaticity than flexibility of social cognition? One has to consider that the internal state term ratio, the variable of automaticity most strongly related to the DCCS is not independent of flexible mental attributions. Indeed this variable is based on the total response time divided by the total number of mental state terms used and correlated with the number of unique mental state terms used. Furthermore the DCCS score is partly based on response speed thus both variables share a common methodological variance through this speed component. On the other hand the studies reviewed above that found a link between cognitive flexibility and theory of mind did not use continuous measures of flexible mental state attribution but normative measures of mentalizing accuracy. The internal state term ratio might be more related to this accurate attribution of mental states than the number of mental state justifications or unique mental state terms generated as possible explanations for the social vignettes. As no measure of mentalizing accuracy was used, this hypothesis has to be tested in in the future.

Second, why was the DCCS only significantly correlated to flexibility of social cognition in adults? The possibility of idiosyncrasies or a cohort effect in the adult sample cannot be ruled out. Another possibility is that the relationship between executive functions and social cognition changes across the lifespan. The adult sample might have recruited their cognitive flexibility resources more strongly in the service of the task. Alternatively, the link between cognitive flexibility and automaticity of social cognition might originate more early in development and express itself only later. As we have seen in chapter 2.1.3, there is sound evidence that development of executive functions facilitates (and precedes) advances in theory of mind. Unfortunately the design of this study cannot answer this question and further studies have to investigate if there are indeed differences in the link between flexibility of social cognition and cognitive flexibility across lifespan and if so elucidate the causes.

While many questions remain open, this study adds findings of a novel methodology to the existing body of research that links executive functions and theory of mind.

2.4.6. Theory of mind & clinical symptoms.

Self-report of clinical symptoms was positively associated with flexibility of social cognition but only in children. While high competence in mentalizing has mostly been linked with mental health (and vice versa) one has to bear in mind that the parent-report was unrelated to clinical symptoms. Although unfaithful responses of care-givers cannot be ruled out, this finding might be better explained by another theory of children's development of mental understanding. A prerequisite to report experiences and behaviour associated with symptoms like social shyness, anxiety or aggression is the development of mental representations (i.e. concepts) and the capacity to think about thinking (i.e. meta-cognition) to be able to become aware of and think about these experiences and behaviours and subsequently verbalize them. Thus self-report of clinical symptoms is usually not taken before middle childhood. Indeed, meta-cognitive knowledge about thinking has been implicated in children reporting negative intrusive thoughts. In a study seven months after hurricane Katrina, 5- to 8-year-old children were surveyed about their knowledge about the mind and their self-reported intrusive thoughts (Sprung, 2008). The meta-cognitive tasks used, probed children's understanding of the stream of consciousness (e.g. that you cannot-not think), mental uncontrollability (e.g. a sudden loud nose makes you involuntarily think about the origin of that noise), intrusive thoughts and belief based emotions. Children's self-reports of negative intrusive thoughts were predicted by their overall score of understanding of the mind. A follow up study replicated these findings (Sprung & Harris, 2010).

2.4.7. What does FASC measure.

Conceptually, the process involved in the FASC procedure might best be described as a form of social cognitive fluency or social divergent thinking in analogue to the existing neuropsychological constructs of verbal, figural or idea fluency (Lezak et al. 2004). The response format is open ended and the task requires to find multiple solutions to an open-ended problem – that is, explaining the behaviour of the characters in the social vignettes. Performance in divergent thinking has been shown to peak before age 40 and decline thereafter remaining relatively stable (Massimiliano, 2015). Similarly flexible and fluent social cognitive problem solving shows an inverted u-shape peaking in the 40s and 50s (Heidrich & Denney, 1994). In the alternative uses task (Guilford, 1967) new and nonconventional uses of an object have to be devised. Perchtold et al., 2018 found that generating alternative appraisals for anger-evoking events recruited a network largely overlapping with an adapted version of the alternative uses task in healthy subjects, hinting at similar processing demands. In the means-end social cognitive problem solving task (Platt & Spivack, 1975) subjects are presented with the beginning of a story vignette describing a person having a problem and the ending where the problem has been resolved with the task of filling in the how the person got there. In a study with schizophrenic patients examining the relationship between this task and various tasks of neurocognitive function, social cognitive problem solving was related to idea fluency (Yamashita, Mizuno, Nemoto, & Kashima, 2005). I hypothesize that performance on the FASC variables depends on perspective taking and mentalizing abilities, social experience, divergent thinking and verbal skills.

There was no explicit instruction to explain the behaviour of the characters in terms of mental states. Thus it is risky to equal the actual mental state talk happening in the FASC responses to an actual ability to describe social situations in mental terms. Standard prompts might be compared with prompts in which participants are explicitly asked to explain the character's behaviours in terms of mental states. Trait-wise, mind-reading motivation has been recently proposed as an important concept to complement mentalizing abilities (Carpenter, 2016). Mind-reading motivation refers to "an individual difference in the extent to which a person is willing to effortfully engage in understanding the perspectives and mental states of other people" (p. 358). Without knowing how performance in the FASC relates to established measures of mind reading accuracy it is premature to explain FASC performance in terms of ability as the influence of mind-reading motivation is unknown. Cautious evidence that FASC responses might indeed capture more normative aspects of mind reading can be found in the literature about mental state talk. Hughes and Dunn (1998) investigated relations between mental state talk in children's dyadic play sessions, theory of mind and emotion understanding in a longitudinal study at three time points at ages 3 years 11 months, 4 years 6 months and 5 years. Mental state talk at time points one and two were related to different measures of understanding mind and emotion 7 and 13 months later at time points two and three. Mental state language use has also been shown to be moderately correlated with theory of mind in primary school children (Grazzani & Ornaghi, 2012).

More broadly, motivation is seen as a *conditio sine qua non* for most competencies to express in performance. Test-taking motivation, for example, is an important caveat issued in undergraduate courses of psychological assessment, one exceptionally difficult to account for however. In self report questionnaires in personality assessment there are sometimes so called scales of verisimilitude that try to safeguard against unfaithful response styles (e.g. social desirability) and there are a few objective personality tests that measure achievement motivation as a trait. In a single test of aptitude however, there is usually no way to objectively tell if a test taker could have finished the task a few seconds earlier if he had put more effort into it, barring severe cases of non-compliance or aggravation. Prompts to work as fast and accurate as possible, or praise and encouragement during practice trials have to suffice. In the practice of achievement assessment, the context in which the examination takes place is weighed in. Taking two extreme examples, in job aptitude testing, a high motivation in test takers can be assumed naturally, while in neuropsychological assessment in the context of insurance issues, the chance to encounter tendencies to aggravate or even simulate impairment is relatively high. Subsequently, through analysis of performance patterns over several tasks and the use of tests of malingering, such tendencies can be detected by an experienced psychologist. These methods are inadequate however for detecting cases in which a test-taker does not have the intention to aggravate and just performs below optimum in a single task for lack of motivational variables. In subjects taking part in scientific studies, motivation to comply and cooperate at their best is often taken for granted. Motives involved in participating may include receiving treatment, curiosity, receiving monetary or course-credit compensation and prosocial sentiments to help the researcher personally or scientific progress at large, only to name a few that come to mind. To my knowledge there is surprisingly little research on these motives and their effect on study results. Assuming that motivation influences response behaviour in the FASC procedure, the actual responses would only equal (or better approximate) the ability to describe social situations, if effort was equally distributed among subjects with high and low abilities, something which was not tested for. Finding a way to manipulate task motivation would help untangle motivation's influence on the outcome variables. Alternatively, the level of motivation to participate and motives to do so could be surveyed prior to the examination. Additionally post-experiment subjects are sometimes asked to rate to which degree they gave their best in the task. In summary, motivation, both the inclination to mentalize (e.g. as a trait) and task-specific (i.e. state) seem to be neglected subjects in social cognition research and may help better understand what tasks like the FASC actually measure.

2.4.8. Limitations and outlook.

Internal state term language use is a valuable methodological approach in itself as studies – mostly with toddlers and young children – have shown (e.g. Bretherton & Beeghly, 1982; MacWhinney & Snow, 1985). However to better understand how FASC performance relates to theory of mind, it should be compared with established tasks (e.g. strange stories, faux pas, etc.) as a next step. Similarly investigating how internal state talk in response to FASC vignettes relates to internal state talk in everyday situations would inform ecological validity. Furthermore the associations with other constructs that have been revealed (e.g. cognitive flexibility, self-reported clinical symptoms) should be replicated with different tasks and relations to further abilities examined in a multi-trait-multi-method approach. A usual but important demand is the replication with a different sample. This is especially important when comparing age groups to safeguard against idiosyncrasies of the different sub-samples collected. A longitudinal study (obviously only applicable in a more narrow age range) would provide even more thorough evidence concerning developmental trends. Looking further into the future, advances in speech recognition might dramatically facilitate the analysis of mental state language and automate the scoring of FASC.

A strength and weakness at the same time is FASC's lack of narrow definitional boundaries concerning the vignette's narratives. Potentially every imaginable social situation might be the subject of an item and its open-ended response format. Although in this study an entire set of new items has been used compared to the original study by Hayward et al. (2016), the large amount of replicated results suggests the validity of the item material and the flexibility of the FASC procedure. Still, it might prove fruitful to systematically collect social situations, typically encountered in different age groups and construct item sets, tailored to each age group. This might be particularly important for children whose scope of social experience is arguably more different than that of adolescents and adults. This approach would take into account the importance of social experience in the development of social cognition that has been suggested (Carpendale & Lewis, 2004; Fernyhough, 2008). On the contrary this approach would potentially complicate comparability across age groups and require careful examination of relative item difficulty. Alternatively a set of items might be constructed and thoroughly validated that consists of social situations equally familiar to all age groups. Systematic manipulation of the degree of familiarity might even be used to further investigate the influence of social experience on social understanding.

2.4.9. Conclusion.

This study demonstrated profound age differences in flexible social cognitive reasoning and mental state talk from childhood till old age in response to social cartoon vignettes. In particular mental justifications and absolute, as well as relative mental state term frequency increased from primary school age over adolescence to adulthood and showed a decline in old age. Additionally the presence of verbal cues and the degree of ambiguity in the items had an effect on these variables. Verbal vignettes and ambiguous vignettes elicited more mental justifications and mental state terms than non-verbal and unambiguous vignettes. Mental justifications showed an interaction for language and age while mental state terms produced an interaction for ambiguity and age. Flexibility of social cognition was related to verbal comprehension while cognitive flexibility was associated with automaticity of social cognition. In children only, flexibility of social cognition was related to self-reported clinical symptoms. The FASC seems to be a promising continuous measure to capture flexible social cognitive reasoning. This notion is further supported by the high degree of replicated results of the original study by Hayward et al. (2016) despite using an entirely new set of stimuli. However, the conceptualisation of automaticity particularly concerning initial response time needs further elaboration. Also much more work needs to be done to establish its reliability, construct validity and incremental validity in comparison with other tasks of social cognition. As with other measures utilizing an open-ended response format, the influence of task- and mindreading-motivation is a possible confounder that needs further consideration.

3. Study 2: Computerized Task of Recognizing and Understanding Emotions

3.1. Introduction

3.1.1. Emotion understanding.

The ability to perceive emotions, to understand their antecedents and consequences and malleability in oneself and others is a crucial ability for navigating our social world. Emotion understanding encompasses at least nine different components (Pons, Harris & DeRosnay, 2004): 1) recognizing facial expressions, 2) external causes of emotions (e.g. the loss of something cherished causes sadness), 3) desire- and 4) belief-based emotions (e.g. people's mental states influence the emotions felt in a situation), 5) memories as triggers of emotions, 6) emotion regulation, 7) hiding emotions (e.g. facial expressions of emotions can be manipulated to not match the corresponding feeling) 8) mixed emotions (e.g. some situations can elicit more than one emotion) and 9) moral emotions (e.g. breaking social norms can make you feel sad). Furthermore research has shown that these nine components can be grouped along difficulty and developmental acquisition into three superordinate categories: external emotion understanding is concerned with the recognition of emotions in the face and relating external events emotions (in the presence or as memory cues) to emotions (components one, two and five). The mental level is about understanding the influence of mental states on emotions as well as the distinction between real and apparent emotions (components three, four and seven). Understanding the malleability and variability of emotions and the influence of social and moral expectancies on feelings is called reflective emotion understanding (components six, eight and nine).

Some form of implicit understanding of emotions (e.g. processing of facial expressions) and their behaviour guiding function develops much earlier and presumably has an innate basis (Nelson, 1987). Yet when defined more narrowly, namely treating emotions as an object of explicit knowledge, emotion understanding develops after the second year of life. The present work focuses on explicit emotion understanding, acknowledging that: first, it does not tell the whole story about emotion understanding and second, that the distinction between explicit and implicit processing may in many cases be an artificial one since all behaviour rests on brain activity that, on a subjective level, can have the properties of qualia (i.e. subjective conscious experience) and propositional attitudes to a varying degree.

Furthermore there has been a longstanding debate about the universality and discreteness of emotion categories. The view that emotions are basic, discrete entities that are given by nature (i.e. natural kinds), non-arbitrarily grouped and identifiable by a set of typical or even defining autonomic, muscular and hormonal responses and distinct causal mechanisms (e.g. a specific brain circuit) follows a long research tradition, can be considered scientific mainstream doctrine and is reflected in folk theories of emotions (e.g. Darwin, 1872; Ekman & Friesen, 1971; Ekman, 1992; Ekman & Cordaro, 2011; Izard, 1977; Panksepp, 1992; Panksepp & Watt, 2011; Plutchik, 1980). The notion that a small set of basic emotions form the building blocks of our rich affective life has not remained unchallenged however (e.g. Ortony & Turner, 1990; also see the following debate: Ekman, 1992; Panksepp, 1992; Turner & Ortony, 1992; Barrett, 2006a). It has further been shown that emotions assessed through self-reports are multidimensional, heterogeneous experiences, statistically better conceptualized in a geometric space, generated by the more elemental properties valence and arousal than by separable factors (Russell, 1980; Russell & Barrett, 1999; Barrett, 2006a). These two qualities form the *core affect*, which is described as the most basic affective feeling and its corresponding neurophysiological changes (Russell & Barrett, 1999; Russell, 2003). The core affect can be seen as information about the external world, transcoded into internal representations (also see Damasio's concept of somatic markers; Damasio, 1994) influencing homeostasis and informing the organism of possibly positive or detrimental environmental events. According to the conceptual act model of emotion (Barrett, 2006b), the interpretation of the core affect, utilizing conceptual knowledge, generates emotion. This account has been further elaborated into the theory of constructed emotion (Barrett, 2017). The debate is complicated by researchers using the terms emotion, feeling and affect differently and at times interchangeably and by focussing on different levels of measurement (e.g. facial expressions, brain activity, subjective experience) while often equating these proxy-measures with the construct emotion itself. Accordingly, Ledoux (2015), argues to view emotional behaviour (e.g. defensive behaviour) and conscious feelings as not synonymous in the sense of not being reducible to the same brain systems. Survival circuit behaviours and motive states are important, but not the only ingredients to subjective feelings and may or may not be accompanied by conscious feelings in humans. An extensive account has been laid out against the natural kinds / basic emotions view (Barrett, 2006; also see responses and commentary: Panksepp, 2007; Izard, 2007; Barrett et al., 2007). Meta-analytic evidence has been presented and interpreted in favour of one or the other position (Vytal & Hamann, 2010; Lench, Flores, & Bench, 2011; Lindquist, Wager, Kober, Bliss-Moreau, & Barrett, 2012) with a lively discussion ensuing (Lindquist, Siegel, Quigley, & Barrett, 2013; Lench, Bench, & Flores, 2013). Although posing a valid philosophical question, for the purpose of the present work this distinction is of little relevance since it is concerned with emotions as concepts that in dyadic interactions are usually given discrete labels, whatever the origin of this concept.

Whether a mother talks to her child about being angry or sad because a taxon of the emotion experience is hardwired into the brain at birth and later given a name, or because the taxon has been shaped through her own social experience and social conventions does not change the fact that it will help the child understand emotions which a wealth of studies show as I will review later.

3.1.2. Concepts related to emotion understanding.

In the course of the last four decades or so different researchers have established several similar, partly overlapping concepts related to understanding of emotions. Most relevant for the present work is the concept of emotion understanding as described above, but for the purpose of orientation and relating the present work to findings from other studies a short overview of related concepts is provided first.

3.1.2.1. Emotion knowledge.

Emotion knowledge is a concept from developmental psychology (Izard, 1971) and encompasses the ability to recognize and label emotions in facial expressions, behaviour and social situations (Trentacosta & Fine, 2010). Thus it can be seen as largely covering the external level of emotion understanding. Emotion knowledge is usually measured with tasks probing emotion recognition in pictures of facial expressions and labelling emotions in short descriptions of social situations and social behaviours. One such task is the Assessment of Children's Emotion Skills (ACES; Schultz, Trentacosta, Izard, Leaf, & Mostow, 2004). Thus in the present work, when using the term emotion knowledge in reporting study findings, tasks were used where subjects have to verbally and/or nonverbally identify facial emotional expressions as well as a task where social situations and/or social behaviours (depicted by vignettes or enacted with puppets) had to be labelled. If only recognition or emotional situation knowledge was assessed it is explicitly mentioned instead of using the term emotion knowledge.

3.1.2.2. Emotional competence.

Emotional competence (Saarni, 1999) consists of six interrelated domains where the first five are considered procedural and the last one is declarative: (1) The subjective experience of emotions, (2) the expression of emotions, (3) the recognizing of emotional expressions, (4) the control of emotional expressions, (5) the regulation of the subjective experience of emotions and (6) the understanding of the antecedents and consequences of emotions (Pons, de Rosnay, Andersen, & Cuisinier, 2010). Viewed in this light, emotion understanding can be seen as part of emotional competence largely pertaining to point three to six.

3.1.2.3. Emotional intelligence.

In the literature of emotional intelligence at least two competing concepts can be distinguished. One account is narrower and views emotional intelligence as a cognitive ability like other intelligences (termed *ability account*), while the other is broad and contains descriptions also found in other psychological constructs (e.g. personality traits), thus marking it a *mixed account* (Mayer, Roberts, & Barsade, 2008). The differences of these models are also reflected in the way they are typically assessed. While the ability account of emotional intelligence is usually assessed with performance based measures (e.g. MSCEIT; Mayer, Salovey, & Caruso, 2002) the mixed account relies on self- and other-reports (e.g. EQ-I; Bar-On, 1997). In line with their conceptualizations, they differ strongly in their overlap with personality variables. In a comparison of three measures of emotional intelligence (Brackett & Mayer, 2003) for example, the Emotional Quotient Inventory shared 75% of its variance with the big five personality dimensions and correlated with each factor (up to r=-.57 with neuroticism) questioning the incremental validity of this broad conceptualization. In contrast, the MSCEIT shared 38% of variance with the big five and correlated significantly with only two dimensions (highest r=.28 with agreeableness).

3.1.3. Development of emotion understanding.

Emotion understanding and theory of mind should not be viewed as separate abilities as both are concerned with thinking about mental states. In the literature however, both competences are often treated separately. Recognizing the fact that no universally accepted definition exists for the terms theory of mind, mentalizing, mental understanding or emotion understanding, the term emotion understanding is both more narrow and broader than theory of mind. It is narrower in so far as theory of mind concerns a diverse range of mental states (e.g. intentions, beliefs, desires or emotions). It is broader than theory of mind (at least as commonly conceptualized by theory-theorists) because it encompasses not only conceptual understanding but also relatively implicit understanding like recognizing emotions from facial expressions, tone of voice, body postures or gestures. In the following paragraph I will shortly summarize the development of emotion understanding as conceptualized by the test of emotion understanding (Pons et al., 2004). I then turn to the nine components of emotion understanding and review them more broadly.

Between 3 and 4 years, young children learn to recognize and label emotion expressions. At about the same age, they can link emotional states to external causes and acknowledge people's desires and goals when judging the emotional state of others. Starting at the same time but stretching over a longer time span (3-6 years), children develop the understanding of how emotions are affected by memory (e.g. that a reminding cue can trigger an emotion). Later (4-5 years) they start to take into account the influence of a person's beliefs and expectations on its actions. Yet the attribution of a person's emotion based on its beliefs lags behind for around one year. From 4-6 years, children acquire knowledge about appearance and reality of expressed emotions. Emotion expressions can be altered to not reveal the emotion felt. Emotion regulation develops throughout childhood and beyond. Younger children (6-8 years) predominantly use behavioural strategies (e.g. distraction) while children from 8 years on start to use cognitive strategies (e.g. positive distraction, reattribution) to influence their emotions. At that time, children also start to acknowledge the existence of mixed or ambivalent emotions, that is, multiple and even opposing emotions can be experienced in a situation depending on the elicited associations. Also around that time, the understanding that the breaking of social norms can cause feelings of guilt and sadness arises.

3.1.3.1. Emotion recognition.

An early procedure to study children's explicit (i.e. verbally mediated) facial affect recognition that has been utilized widely is the affect labelling task by Denham (1986). Children have to expressively and receptively identify four drawn faces expressing feeling happy, sad, angry, or scared. In the original study introducing the task, which only reported an aggregate score combining facial affect labelling and affective perspective taking (see next section), 2- to 3-year-olds already showed performance clearly above chance. In a longitudinal study also reporting the affective aggregate score, children at mean age 3 years 11 months achieved a score of 37 out of 50 and at age 4 years 6 months they scored 43 points (Hughes & Dunn, 1998). In another study reporting separate scores for the facial affect labelling task, a sample of 3- to 4-year-olds (mean age 4,16) scored almost perfectly suggesting that prototypically expressed emotions (at least happiness, sadness, anger and fear) are reliably recognized by most 4-year-olds (Cutting & Dunn, 1999). It is important to note however, that there was considerable inter-individual variance in performance in the studies above, a fact generalizable to the development of emotion understanding as a whole (Dunn, Brown, & Beardsall, 1991; Pons et al., 2003, 2004; Pons & Harris, 2005). Importantly, development of facial affect recognition is not finished in preschool and continues through childhood and into adolescence (Herba & Phillips, 2004). There is also evidence that the ability to recognize emotion-expressions in faces does not stay constant throughout adulthood but decreases to a variable degree depending on the emotion. A recent meta-analysis (Ruffman, Henry, Livinstone, & Phillips, 2008) investigated age related differences in emotion recognition in adulthood. Mean age of the included studies ranged between 19 and

30 years for the younger age group and 65 and 77 years for the older age group. For age differences in facial emotion recognition medium effect sizes for anger, sadness and fear and small effect sizes for surprise and happiness were found. A large cross-sectional study (Mill, Allik, Realo, & Valk, 2009) compared smaller age groups with the youngest ranging from 14 to 20 and the oldest from 61 to 84. While performance for all emotions dropped markedly in the oldest age group, the ability to recognize sadness and anger started to decline almost linearly from 30 years onward. The ability to identify if a face displays an emotion at all (i.e. recognition of neutral expressions) did not show a decline even in the oldest age group.

Deficits of facial emotion recognition have been found in a range of child and adolescent populations with psychiatric disorders, both externalizing and internalizing, such as schizophrenia, mood disorders, anxiety and neurotic disorders, eating disorders, ADHD and conduct disorders (Collin, Bindra, Raju, Gillberg, & Minnis, 2013). The mechanisms for these associations are not yet well understood.

The short overview so far has only addressed explicit emotion recognition assessed through verbal labelling. Because assessment of emotion recognition lends itself more toward implicit measures than higher levels of emotion understanding I want to briefly turn to this venue of research. The analysis of an infant's ability to discriminate and possibly recognize emotion expressions (as with most early competencies) is made difficult through the lack of verbal abilities and largely relies on habituation paradigms in young infants and social referencing paradigms later. Nelson (1987) reviewed the literature on facial expression recognition in the first two years of life. Infants younger than 4 months can discriminate facial expressions in the same face most likely based on simple perceptual features. From 4 to 7 months, infants start to distinguish some form of facial affect categories in exemplars varying on several dimensions (e.g. age, sex) although the exact taxonomy of these categories is not clear yet. Walker-Andrews (1997) in her review comes to a similar conclusion but subsumes that for unimodal facial expressions (i.e. without accompanying vocalisations) feature information is preferably used for discrimination (e.g. toothy vs. non-toothy expressions) but that they are able to extract affect information from varying faces if feature information is not salient. By 7 months, infants show preference for two stimuli on different modalities that are congruent in affect (i.e. face and voice) called intermodal matching (Walker-Andrews, 1997; Widen & Russel, 2008). How rich the conclusions derived from this evidence based on looking time data should be is subject to personal speculation. Minimally and from a cautious viewpoint, findings of facial expression discrimination based on habituation experiments only demonstrate that infants are able to extract patterns of perceptual features (e.g. facial features) which have been formed by repeated exposure in an environment associated with the fulfilment of basic needs (i.e. mother). Assessing the appropriateness of responses to affect expressions (e.g. via social referencing) is another, perhaps more valid, way to assess infants' ability to extract meaning from facial expressions. Affect expressions indeed seem to influence the behaviour of infants from 8 months onward (Nelson, 1987; Walker-Andrews, 1997; Campos, Thein, & Owen, 2003) although results are often ambiguous in terms of interpretation for which two examples are described (see Nelson, 1987 and Walker-Andrews, 1997 for more examples). In a study with 12 to 18-month-old infants (Klinnert, 1984), mother's reaction to a new toy (smiling vs. neutral vs. fearful) had an influence on how close the infant moved to the mother but not on approach behaviour towards the toy. In several experiments (Sorce, Emde, Campos, and Klinnert, 1985) with a visual cliff that elicited no clear avoidance, 12-month-old infants did not cross the cliff when the mother looked fearful but a majority crossed when she looked happy. Similar avoidance effects were shown for angry and sad expressions however. Furthermore, in a control experiment however that in contrast featured a cliff without a drop-off, very little referencing and no influence of affect expressions was observed and almost all infants crossed the visual cliff. Nelson (1987) in his review highlighted other limitations: a substantial portion of children in the reviewed experiments did not reference at all or failed to respond "appropriately" to this information which, taken together, exceeded 50% for some expressions analysed. Unfortunately most studies did not compare emotion categories of the same valence. According to Widen & Russel (2008) the data from those that did is more consistent with the view that prior to 3 years, children "understand" emotions more in terms of the two dimensions valence (pleasant-unpleasant) and arousal (low-high) which form four broad categories. For example in the study by Sorce et al. (1985), not only fearful but also angry and sad expression deterred infants from crossing the cliff, although to a different degree. Apart from these ambiguous results, the interpretation that appropriate responses to facial expressions indicate understanding of them is not unanimously accepted. As Flavell (1999) exemplifies, an infant would not have to associate a mother's anxious expression with the subjective feeling to avoid an object. It would suffice to associate it with aversive consequences which is presumably a more basic process thus offering a more parsimonious explanation to the phenomenon. Similarly, Saarni & Harris (1989) argue, that "the capacity to distinguish among different expressions of emotion is not tantamount to the possession of a concept of any of the discriminated emotions" (p. 5). Thus the question of what constitutes "understanding", whether it begins with perceptual pattern recognition or with concept formation is open to debate and perhaps the term is too broad to be useful in scientific discourse apart from introductions.

3.1.3.2. External causes of emotions.

External causes for emotions are more easily identified for positive than for negative emotions (Borke, 1971). In fact in this study, 3-year-old children were already quite adept in linking situational descriptions to positive emotions. Being afraid was reliably identified from the age of 4.5 onwards while situations causing sadness were a little harder to understand (approximately age 5.5 and older). For identifying anger-eliciting situations the developmental trend was less clear. Five-year olds more reliably identified anger-eliciting situations compared 4-year-old but still less than 50% of subjects responded correctly to both story vignettes beyond the age of five. In a naturalistic study experimenters observed children playing on the playground of a day-care facility (Fabes, Eisenberg, Nyman, & Michealieu, 1991). When a situation involving expressed emotions emerged they rated the emotion and the situational causes. Whenever another child that was not involved in the situation witnessed the expressed emotions of the other child, the observer immediately interviewed it afterwards and asked which emotion the child had experienced and why. For social situations (involving another child or adult) they distinguished between physical (e.g. pushing), material (e.g. sharing a toy), control (e.g. guardian requesting something), nonverbal (e.g. ignoring) and verbal (e.g. joke telling) emotion contexts. Correspondence of the child's and the observer's evaluation of the situational antecedents increased from 3.5 to 4.5 and 5.5 year (correspondence of .67, .71 and .85 respectively) but non-significantly, possibly due to the small sample size (n=12, 15 and 16 respectively).

A standard procedure that has been used to investigate children's understanding of external causes of emotions is the affective perspective taking task from Denham (1986). Puppets with blank faces act out emotion-eliciting situations with vocal and visual affective cues and subjects have to attach the corresponding emotional face. While this task arguably captures important aspects of contextual emotion comprehension, the understanding of external causes is mingled with the recognition of emotions in the voice and non-facial body language. In a task not suffering from this limitation, devised by Denham, Zoller and Couchoud (1994), subjects are presented puppets with facial expression and have to come up with multiple explanations what made the puppet feel this way. In a large longitudinal study (O'Brien et al., 2011), there was a significant increase in both affective perspective taking and knowledge of emotion causes between 3 and 4 years.

3.1.3.3. Reminder of emotions.

A similar timeline emerges for the understanding of how thinking about past events or experiences cues present emotions and how emotional intensity wanes over time. Six-year-olds and to some extent 4-year-olds predict that emotions wane in intensity over time and that thinking or forgetting about the eliciting event modulates this process (Harris, Guz, Lipian & Man-Shu, 1985). In the preschool years children gradually link present emotions to past events and start to explain them in mental terms. Lagattuta and Wellman (2001) presented stories to children between 3 and 5 years about an event eliciting a positive or negative emotion in a story character. After many days have passed in the story the character faces a cue for the old event (e.g. same person/animal or object) which re-invokes the old feeling but at the same time is related to a new positive or negative context. The stories thus present four possible combinations: matched or mismatched positive or negative emotions. In a negative mismatch story, for example, a cue for a negative past event which elicits the negative emotion is paired with a new positive event (e.g. a dog that in the past chased away a rabbit is now advancing very friendly). Across all story types, 5- compared to 3-year-olds more often referred to prior experiences when explaining current emotions and 5- compared to 4-year-olds gave more mental references and cognitive cueing explanations. Only in 5-year-olds more than half of the responses were prior experience- and thinking explanations while cognitive cueing explanations were given only for only half of the stories, even in 5-year-olds. There was a large difference between story-type however. In the positive emotion mismatch and the emotion match conditions only about half of the 5-year-olds referred to prior experiences and thinking and only about a third also included cognitive cueing in their explanations. Only about a third of the 3- and 4-year-olds included references to prior experiences and only few included explanations based on thinking or cognitive cueing. In the negative emotion-mismatch condition the majority of 3-, 4- and 5-year-old or older subjects made references to prior experiences, thinking and cognitive cueing (respectively). In a second experiment, the majority of 7- but not 5-year-olds or younger children gave thinking and cognitive cueing explanations in the emotion match condition. In both experiments, adults gave rich explanations referring to prior experiences, thinking and the specific cognitive cue in nearly all instances. In summary, significant changes in taking past events into account when attributing current emotions occur around the age of five. Cues of negative past events are seen as stronger influencers on the experience of current positive situations than vice versa while acknowledging the influence of past same-valence events on current experience (a form of interpretive diversity) develops even later.

3.1.3.4. Desire-based emotions.

People may experience different emotions about the same situation because they have different desires. Understanding the link between desires and emotional reactions develops between 3 and 5 years (Harris, Johnson, Hutton, Andrews, & Cooke, 1989, Pons, Lawson, Harris, & de Rosnay, 2003; Pons et al., 2004). The emotional consequences of a match or mismatch between a desire and reality (e.g. get what one wants) are comprehended more easily (age 2 or 3) than if the evaluation depends not on reality but a belief about reality which is not understood until about age 5 (Hadwin & Perner, 1991, Wellman & Woolley, 1990). There is evidence however that even 18-month-olds can infer desires from affect reactions in some contexts (e.g. food preferences). In a study (Repacholi &Gopnik, 1997), 14- and 18-month-old children observed an experimenter tasting two different kinds of food and reacting with a positive affect or with disgust. When the experimenter later requested food, 18- but not 14-month-olds were more likely to give him the one he had previously shown preference for, even if it was the food they themselves did not prefer.

3.1.3.5. Belief-based emotions.

False-Belief understanding (as measured with unexpected location or content tasks) develops between 3 and 5 years (Wellman et al., 2001). Attribution of belief-based emotions however has been consistently shown to lag behind (false-) belief understanding 1 or 2 years (Harris, Johnson, Hutton, Andrews, & Cooke, 1989; Hadwin & Perner, 1991; Bradmetz & Schneider, 1999; De Rosnay, Pons, Harris, & Morrell, 2004; Harris, deRosnay, & Ronfard, 2014). This could be at least in part because it requires second order ascription of mental states (an emotion based on a belief). Understanding of belief-based emotions starts at 4 but even many 7-year-olds have problems to attribute the correct emotion based on a false-belief (Bradmetz & Schneider, 1999). A recent study also found this lag for attributing belief-based emotions of self (Bender, Pons, Harris, & de Rosnay, 2011). While passing-rates for false-beliefs significantly increased between 5 and 6, correct attribution of false-belief-based emotions did not. Other studies found that the passing rate for correctly attributing emotions based on false-beliefs changed from minority to majority from 4 to 6 (Harris et al., 1989) or 5 to 7 years (Pons et al., 2004). Taken together these studies suggest that important changes in understanding belief-based emotions take place in the transition from preschool to primary school age.

3.1.3.6. Hiding Emotions.

In an early study (Saarni, 1979), 10-year-olds compared to 8- or 6-year-olds referred more often spontaneously to display rules and exhibited more complexity in choosing a facial expression. That is, they more often chose an emotion thought to be felt by the character or intentionally chose a different emotion and in both cases showed inferential-causal reasoning compared to selecting a facial expression at random or similar to what the character last looked like. Harris, Donelly, Guz, and Pitt-Watson (1986) demonstrated that a limited ability to distinguish between real and apparent (displayed) emotions is in place in children as early as four years of age. This early understanding is very fragile however and possibly limited to the case of hiding negative situations (Joshi & MacLean, 1994). In the aforementioned study by Harris et al. (1986), 6-year-olds were more proficient in correctly labelling and differentiating between real and apparent emotions than 4-year-olds but there was no significant difference between 6 and 10. In terms of justifications for different emotions each age group surpassed its younger peers. However even 10-year-olds on average were correct in only half to three-quarter of the stories indicating that understanding the distinction between real and apparent emotions and knowledge of display rules is not at ceiling even in older children. The age differences in understanding hiding emotions between 4 and 6 years have also been shown in Japanese, Indian and English children (Gardner, Harris, Ohmoto, & Hamazaki, 1988; Joshi & MacLean, 1994).

3.1.3.7. Emotion regulation.

Emotion regulation is a special case in that some aspects of it (e.g. the knowledge about emotion regulation strategies) can be subsumed under emotion understanding while other aspects go beyond it. There is a growing literature about automatic (i.e. implicit) emotion regulation of the organism, that takes place beyond deliberate or effortful strategies (Gyurak, Gross, & Etkin, 2011; Mauss, Bunge, & Gross, 2007; DeWall et al., 2011). Additionally, knowing about strategies alone does not suffice to effectively utilize them. Executive functions and effortful control (e.g. orienting attention, inhibition of behaviour or delaying gratification) have been implicated in the regulation of emotions to the extent that they are used interchangeably by some authors (e.g. Eisenberg, Sinrad, & Eggum, 2010). Inhibitory control is associated with emotion regulation in preschool children (Carlson, & Wang, 2007) and cognitive control might mediate the effect of mindfulness training on emotion regulation (Teper, Segal, & Inzlicht, 2013). As this work focuses on emotion understanding, the discussion of emotion regulation in this context is largely limited to knowledge about emotion regulation strategies.

Understanding of emotion regulation strategies becomes more varied and complex throughout childhood (Harris, Olthof, & Meerum Terwogt, 1981). Importantly, cognitive emotion regulation (e.g. distraction) develops during middle childhood. While 6-year-olds only rarely mention *cognitive distraction* as a suitable strategy in distressing situations, 11-year-olds do so comparable to *behavioural distraction* strategies (Altshuler & Ruble, 1989). Avoidance strategies like *escaping* deemed more maladaptive, at least in some contexts, made up a smaller proportion in older compared to younger children. Similarly in another study (Band & Weisz, 1988) the reported use of secondary coping strategies (e.g. social support, cognitive avoidance, re-appraisal) increased from 6 to 12 years. Although the most dramatic changes in the understanding of emotion regulation take place in middle childhood development continues through adolescence and adulthood (Rawana, 2014). Interestingly, there is evidence that use of adaptive emotion regulation strategies and social support (Zimmermann & Iwanski, 2014) and cognitive reappraisal (Gullone, Hughes, King, & Tonge, 2010) declines from middle childhood to early adolescence. Subsequently use of adaptive emotion regulation strategies appears to increase again through emerging adulthood to middle adulthood (Zimmermann & Iwanski, 2014). Furthermore, older adults have been shown to be more competent in regulating emotions as evidenced by self-report (Gross et al., 1997) and experimentally through mood-repair (Kliegel, Jäger, & Phillips, 2007). However developmental differences in emotion regulation seem to be emotion-specific (Zimmermann & Iwanski, 2014). For example while in early adolescence, dysregulation increased for anger, suppression increased for fear and passivity decreased for sadness. One caveat in emotion regulation research in adolescents and adults however is the prevalence of self-report measures which often ask about the use of emotion regulation strategies which does not necessarily coincide with competency in these strategies.

Emotion regulation, probably even more than other aspects of emotion understanding, has been implicated in mental health. Mental health problems such as borderline-personality-disorder, depression, anxiety-disorders, substance-related-disorders, eating disorders and somatoform disorders are all intimately tied to deficits in emotion regulation (Berking & Wupperman, 2012). Emotion regulation viewed as effortful control or self-regulation is related (mostly negatively) to externalizing and internalizing problems from a very young age but the association is more consistent for externalizing (Eisenberg et al., 2010). The link with externalizing and internalizing problems has also been found when looking at the ability to regulate emotions more specifically (Hill, Degnan, Calkins, Keane, 2006; Rydell, Berlin, & Bohlin, 2003). For example, cognitive reappraisal was significantly related to positive and negative indicators of mental health in a recent large meta-analysis (Hu et al., 2014).

3.1.3.8. Mixed Emotions.

Related to other forms of interpretive diversity is understanding of emotional ambiguity or mixed emotions, the recognition that the same event can provoke different emotions in different people or even the same person. Understanding of mixed emotions does not develop at once, it follows a protracted development:

Donaldson & Westerman (1986) differentiated between different levels of understanding ambivalence in emotions. There was a significant effect of age for children between 4 and 11 years. Children between 7 and 8 began to show understanding that the same event can elicit contradictory feelings in the same person and that they are not exclusively linked to external factors but the influence of enduring memories or traits was not acknowledged yet and the different explanations were still separated temporally (e.g. first she feels X, later she feels Y). Children between 10 and 11 appreciated that feelings can coexist, modulate each other, are linked to enduring traits and that these mixed feelings can be confusing. Harter & Budding (1987) found a 5-step developmental sequence concerning recognition of simultaneity of emotions. Subjects first had to select from two stacks of photos showing positive or negative facial expressions, place them on a board and find an emotion label for each picture. They then had to think of a thing that would make them feel both emotions at the same time. Depending on the level to be tested face-pictures were of the same or a different valence. They then had to think either of one thing (i.e. same target) that would make them feel both feelings at the same time or how they could have one feeling about one thing and the second feeling about another thing (i.e. different target). A developmental sequence was found. The youngest children (at about 5) did not find examples even for same valence emotions to be felt simultaneously. Subsequent progress of understanding improved in intervals of about 1.3 years in the following sequence: same-valence-same-target, same-valence-different-target, different-valence-different-target, and different-valence-same-target (for mean ages 7.3, 8.7, 10.1 and 11.3 respectively). There was considerable inter-individual variance of 1 to 2 years in the acquisition of each level which is a general finding in theory of mind literature (e.g. see Wellman, 2014 for false-belief understanding and Pons & Harris, 2005 for emotion understanding). Gnepp, McKee and Domanic (1987) found a linear age trend in children from 5 to 8 to attribute more than one emotion (by choosing between three drawn happy, sad and afraid facial expressions) to a character in short equivocal situational vignettes although they were explicitly asked if they are thinking of one or two emotions. However, even at age 8 half of the time the children reported only one emotion in response to equivocal vignettes. When prompted later whether almost all children like/don't like/are scared of XY (according to the child's previous response of happy, sad, scared) or whether some children like and some don't like XY, even 5-year-olds endorsed diversity of emotional experiences in about 70% of responses while 8-year-olds responded with mixed emotions nine out of 10 times. A later study replicated the general findings of the previous study and further demonstrated that development progresses until 12 years (Gnepp & Klayman, 1992). Intriguingly young adults in this study, like the 12-year-olds, acknowledged diverse emotions in equivocal situations only 70% of the time. They also found that including a cue about inter-individual differences in emotional experience in a specific situation (e.g. some kids feel X, some kids feel Y) immediately before asking about the feelings of a character in this situation did not significantly increase the likelihood for attributing mixed emotions to this character. This sheds light on epistemological aspects of emotion understanding. Children acknowledge individual differences in emotional reactivity relatively earl, but do not use this information to infer mixed emotions in the same individual. In contrast cueing intra-individual diversity in emotional experience (e.g. the character sometimes feels X and sometimes feel Y) did increase attribution of mixed emotions, as did leaving one of two possible endings of the scenario open (e.g. either she gets chocolate milk or carrot juice).

Other studies used snippets of animated films to investigate attribution and experience of mixed emotions. Larsen, To, & Fireman (2007) confronted children with the scene of a bitter-sweet farewell from *The Little Mermaid* and discriminated several interesting aspects of mixed emotion understanding. They asked children aged 5 to 12 about emotions the cartoon character (Triton) as well as they themselves had experienced. They also compared whether mixed emotions were reported only after a prompt, spontaneously and if even the simultaneity of these mixed emotions was acknowledged. Several interesting patterns emerged highlighting the protracted development of mixed emotions understanding. There was a significant age effect to attribute mixed emotions (prompted, spontaneously and simultaneously) to the cartoon character. About half of 5- to 6-year-olds reported mixed emotions at all, while only a quarter did so spontaneously and hardly any viewed them as simultaneous. For 8- to 9-year-olds, four in five reported mixed emotions, half of the children spontaneously and about a third as simultaneously occurring. For 11- to 12-year-olds, nine in ten reported mixed emotions, three in four spontaneously and two third of the children as simultaneous. With respect to their own experience, less children reported mixed emotional for themselves than for the cartoon character. Furthermore almost all children who reported experiencing mixed emotions also attributed mixed emotions to the cartoon character while the reverse was true for just under half of the children. Another study using an animated cartoon found evidence that half of a sample of children between 3 and 5 years recognized and reported two opposite-valence emotions (Smith, Glass, & Fireman, 2015). Subjects were first asked how the character felt. If the child didn't report mixed emotions these questions were followed up by asking if he felt anything else and finally by directly asking about him feeling happy / sad. Unfortunately the authors did not report separate results for spontaneous vs. prompted two-emotion responses. If, based on previous literature, one assumes that most were prompted the gap in regard to the findings of the previous study narrows. Still, this constitutes evidence for a fairly early competence of mixed emotions understanding that warrants further attention. In the study by Smith et al. (2015) the final prompt asked for the two emotions separately. Children might have remembered the two emotional states of the character as two different instances which are endorsed independently without acknowledging the existence of both emotions at the same time. Also, the stimuli in the studies by Larsen et al. (2007) and Smith et al. (2015) were not verbal descriptions of situations but snippets of cartoons with (anthropomorphic) characters displaying emotions through facial expressions, body, posture and voice which might have conveyed both emotions expressively and facilitated the acknowledgement of mixed emotions. However there is also evidence that, even without the aid of facial emotion recognition, a partial understanding of mixed emotions already develops between 4 and 5 years. Kestenbaum and Gelman (1995) presented short three-sentence stories to children. The first sentence designated the context (e.g. playground), the second and third sentences each described an event associated which could both elicit the same emotion, different emotions of the same-valence (e.g. sad-mad) or different-valence (e.g. happy-sad). Subjects were asked whether the character feels just emotion A, both emotion A and B, or just emotion B. Five-year-olds significantly provided more mixed emotion responses for the multiple-emotions stories than 4-year-olds but still only for half of the stories.

Another source for interpretive diversity in emotional experience is consideration of inter-, or intra-individual differences in knowledge, traits, experiences, social categories, or thinking styles (Lagattuta, Kramer, Kennedy, Hjortsvanc, Goldfarb, & Tashjian, 2015). Understanding of how different mental strategies and thinking styles influence emotional experiences (e.g. thinking optimistically vs. pessimistically, rumination, distraction, etc.) forms an important part of cognitive emotion regulation which I discussed in the previous section.

From the evidence above it seems that considering mixed emotional responses in others when prompted starts at about age 5. However only in second or third grade do children start to regularly attribute mixed emotions. The acknowledgment of the simultaneity of mixed emotions is an even later achievement of older childhood. Why is (growing) understanding of mixed emotions such a relatively late achievement? This might be due to several factors. Acknowledging two or more emotions at once puts a higher burden on executive functions. One has to hold the ambivalent emotions in working memory, inhibit the more salient emotion and shift flexibly to the other emotion concept. On the other hand the way children learn about emotions may bias their thinking and recognition of them. Adults might talk to their younger children about emotions in a simplified way in an effort not to confuse them. This might promote the development of an overly discrete and one-dimensional concept of feelings which they later have to learn to overcome.

3.1.3.9. Moral Emotions.

During the primary school years around the age of 8, there is a shift in attributing emotions in the context of moral transgression from more immediate gain- or desire-based reasoning to considering other agents feelings and internal moral standards. Four- to 6-year-old children consistently attribute positive emotions to a wrong-doer or victimizer (Nunner-Winkler & Sodian, 1988; Arsenio & Kramer, 1992). This applies even if the transgression is severe and he does not profit materially from it but not if he unintentionally harmed the other person (Nunner-Winkler & Sodian, 1988). In contrast 8-year-olds attribute negative emotions, primarily or after additional prompts to wrong-doers (Nunner-Winkler & Sodian, 1988; Arsenio & Kramer, 1992). Conversely 8-year-olds more often attribute positive emotions to someone resisting temptation to morally transgress despite forgoing a material gain (Nunner-Winkler & Sodian, 1988). Adding ecological validity to these findings, attributing morally oriented emotions to a story character submitting to temptation was related to not peeking in an experimental cheating task (Lake, Lane, & Harris, 1995). A more recent study looked at emotion predictions and explanations for rule violations and willpower decisions (i.e. abiding to the rule) in vignettes that made explicit the character's desire and the prohibitive rule and where in transgression endings there was no negative outcome (Lagattuta, 2005). Four-, 5-, and 7-year-old children and even adults mostly predicted rule-breakers to feel good and rule-abiders to feel bad according to their desires, but not as good or bad as those whose desires are fulfilled or blocked not by their decisions but by external circumstances. However as in Arsenio and Kramer (1992) prompting for a possible second emotion led to 7-year-olds and adults to acknowledge the possibility of desire-emotion mismatches (e.g. feeling bad after breaking the rules or feeling good after abiding to the rules) in the majority of vignettes. Similarly, starting at age 7, subjects also attributed mixed emotions to the majority of story characters. Concerning emotion justifications, 7-year-olds and adults explained emotions more often in terms of rules than 4- or 5-year-olds. Interestingly 7-year-olds also explained emotions more often in terms of future consequences than younger children but also adults. Taken together, there is firm evidence that around 7 or 8, when attributing emotions, children start to take moral or rule considerations into account beyond simply applying a desire framework.

3.1.4. Individual differences in emotion understanding.

Age and language ability are two factors contributing strongly to individual differences in emotion understanding. Together they explained 72% of variance of emotion understanding, as measured with the Test of Emotion Comprehension, in 4- to 11-year-old children (Pons et al., 2003). Of this variance, 27% was explained by language ability and 20% by age alone. Gender did not predict emotion understanding. Despite this considerable influence of age, some 7-year-old children can have a higher level of emotion understanding than some 12-year-olds. Emotion understanding improves over time but also shows remarkable stability (Pons & Harris, 2005). In a linear regression analysis, level of emotion understanding, gender and age together explained 49% of variance of emotion understanding one year later but neither gender nor age were significant predictors for themselves. The missing influence of gender is in line with the former study. The smaller sample size (42 vs. 80) and overall older age (7 to 11 vs. 4-11 years) in this study might have contributed to the non-significant effect of age in the regression.

3.1.5. Cross-cultural differences in emotion understanding.

Only few studies to date have investigated cultural differences in emotion understanding. In these studies differences have emerged although not unequivocally. Two studies have examined differences in appearance-reality understanding of emotions in collectivistic vs. individualistic societies. A study with Indian and English school entry children found a cultural difference but only very specifically in younger Indian girls in child-adult vignettes about negative emotions (Joshi & MacLean, 1994. In contrast a study with 4- and 6-year-old Japanese children did not find an accelerated understanding of appearance-reality emotions compared to a preceding study with English children (Gardner, Harris, Ohmoto, & Hamazaki, 1988). Two studies comparing Chinese and English children found stronger evidence for cultural differences. In the first study, European-American children at age 3 expressed a higher level of emotion knowledge, as measured by generating situational examples for basic emotions, than first generation immigrant Chinese or native Chinese children (Wang, 2008). This difference might be due to different foci of childrearing beliefs between western countries where individuality and raising an emotionally intelligent child is fostered and China where conforming to social norms is stressed (Chao, 1995). At age 4 only the difference between European-American and native Chinese children remained emphasizing the probable influence of early education on emotion understanding. The second study found that English children were better in assessing the impact of an external reminder on emotions while Chinese children had an advantage in understanding the influence of moral reasoning on emotions (Chen, 2009). A study utilizing the Test of Emotion Comprehension found some differences in the rank order of emotion understanding components. While more English children acknowledged that a situation can elicit mixed emotions, more Quechua children affirmed that a memory cue can reactivate an emotion (Tenenbaum, Visscher, Pons, & Harris, 2004). Looking at cultural differences within Europe, Italian children showed a better understanding of appearance-reality (hiding) of emotions in the Test of Emotion Comprehension than German children but not of other components (Molina, Bulgarelli, Henning, & Aschersleben, 2014). A possible explanation for this result are differences of the importance that is placed on autonomy vs interdependence in Germany vs Italy respectively.

3.1.6. Antecedents of emotion understanding.

I am aware of the fact that many studies reported here only used correlational designs that don’t allow for strong inferences about causality. However, based on theoretical considerations and analogous to related constructs (e.g. theory of mind; see Miller, 2012) I will cautiously divide the reporting of related constructs in two sections, one with factors more commonly discussed in the context of influencing the development of emotion understanding and one with factors presumably being influenced by emotion understanding. Whenever studies have applied longitudinal or experimental designs, thus supporting causal interpretations, it will be explicitly stated. This distinction however should not imply that emotion understanding and these other factors cannot influence each other reciprocally. Emotion understanding for example might improve peer likability. Better peer relationships in return might provide experiences to further develop understanding of emotions thus leading to a reinforcing loop.

That being said, a substantial part of the literature on the antecedents of theory of mind (see chapter 2.1.1.2) could have been merged with this chapter under the umbrella of mental understanding. Indeed, a large number of those studies have included tasks of emotion understanding (although mostly emotion recognition and situational emotion knowledge). A distinguishing feature between the two strains of literature however is that the former has almost exclusively utilized false-belief tasks as their primary outcome variable while in the latter the focus is genuinely on emotion understanding. Still, the separation is more a reflection of research programs than of fundamental differences.

3.1.6.1. Parental influences.

Parent's modelling of expressive behaviour and emotional responsiveness and quality of marital relationship all influence children's emotion understanding (Denham, Mitchell-Copeland, Strandberg, Auerbach, & Blair, 1997). Children, whose parents do not discourage their expression of feelings by ignoring or negative reinforcement and who maintain more positive affectivity during difficult child rearing situations, demonstrate a better understanding of emotions. Conversely, children who more positively attend to parent's emotions and whose parents are affectively more balanced also show more positive and less negative emotions towards peers. Children of mothers who report higher levels of marital conflict, show less knowledge of emotional display rules (i.e. hiding emotions) (Nixon & Watson, 2001). Infant-parent attachment has also been implicated to play a role in the development of emotion understanding. In a longitudinal study secure vs. insecure infant-mother attachment at 1 year predicted understanding of mixed emotions at age 6 (Steele, Steele, Croft, & Fonagy, 1999). A recent meta-analysis confirmed the link between attachment and emotion understanding and found that it was of medium strength (Cooke, Stuart-Parrigon, Movahed-Abtahi, Koehn, & Kerns, 2016). The mechanisms of transmission from attachment to emotion understanding are open to speculation however. It may be that securely-attached children are better able to signal emotions, particularly distress, to their mother and in turn receive more sensitive and responsive feedback which provides a safe learning environment and makes emotional expressions worth paying attention to. It may be as well that parents with secure-attachment styles are better in understanding emotions themselves which passes on to the child by genetic transmission and/or learning mechanisms.

3.1.6.2. Linguistic abilities, conversation and mental state talk.

General verbal intelligence and linguistic abilities have been repeatedly linked to emotional competence (Pons et al., 2003; de Rosnay, & Hughes, 2006). In particular receptive vocabulary and literacy seem to be associated with emotion recognition, knowledge about external causes of emotions and mixed emotions (Beck, Kumschick, Eid, & Klann-Delius, 2012). Harris (2005) reviewed the extensive evidence of the influence family conversations about emotions have on childrens' understanding of emotions. In his view, language not only helps children to talk about emotions but also to better remember and imagine future emotion eliciting events. More generally, family mental state talk has been found to predict emotion understanding in longitudinal studies. The frequency and quality of emotional state talk of 3-

year-olds with their mothers and siblings was associated with their ability to recognize emotions in an affective perspective taking task three years later (Dunn, Brown, & Beardsall, 1991). Mother's mental state talk about desires with 15-month-olds predicted their emotion labelling performance on a task about external causes of emotions at 24 months (Taumoepeau & Ruffman, 2006). In another study with preschool children, father's use of causal explanatory mental state language predicted children's emotion understanding (emotion recognition and hiding emotions) two years later (LaBounty, Wellman, Olson, Lagattuta, & David, 2008). After controlling for intelligence, emotion understanding at time 1, intelligence and mother's mental state language, however, it didn't continue to significantly predict children's emotion understanding.

Intervention-studies are another possible source for support of a causal role of mental state language in the development of emotion understanding. In a randomized controlled trial (Gavazzi, Ilaria Grazzani, & Ornaghi, Veronica, 2011) preschool children were read stories rich with emotional state terms two times a week for 2 months but only the intervention group played conversational language games afterwards to foster their use of mental state language. Emotion comprehension (assessed with the TEC) improved significantly in the experimental group compared to the control group. Another study suggested that the understanding of emotional-state language, rather than its use, is associated with children's emotion understanding (Ornaghi & Grazzani, 2013). Giménez-Dasí, Quintanilla, and Daniel (2013) implemented and evaluated the *Thinking Emotions* program with preschool children, an intervention to promote emotion understanding and social skills. This program is based on the *philosophy for children* (Lipman, Sharp & Oscanyan 1980) which teaches children critical thinking skills in a three-step approach: reading a story, collecting questions and peer dialogue initiated by activites like drawing, role-play or miming. In addition to 16 sessions of philosophy for children, 14 sessions were filled with different activities. Sessions targeted basic, complex and mixed emotions, real- vs. apparent emotions, empathy and social competence. As an ancillary to this curriculum, the parents were also given home activities and school teachers were instructed to reintroduce emotion regulation strategies when appropriate situations arose. Four- and 5-year-olds in the program improved in measures of social competence while the latter also exhibited increased emotion understanding compated to a control group.

3.1.6.3. Executive function & nonverbal intelligence.

The link between executive function and theory of mind, in particular false belief understanding, is well established (e.g. Perner & Lang, 1999; Devine & Hughes, 2014). The link between executive function and emotion understanding has not been investigated as thoroughly. At least for the mental level of emotion understanding (desire-based & belief-based emotions, hiding emotions) similar influences are to be expected conceptually. If the associations found for theory of mind and executive functions are similar in regard to emotion understanding remains to be seen. Here I only present studies that included at least one emotion-related outcome variable. For studies investigating the link between executive functions and (cognitive) theory of mind, see chapter 2.1.3.

For external emotion understanding an influence of executive functions is likely. In a study utilizing a procedure similar to the TEC's external emotion component, set shifting (operationalized with a version of the dimensional change card sort task) predicted emotion recognition questionnaire performance beyond theory of mind, language ability, IQ and mothers' age in preschool children (Martins et al., 2016). Furthermore, it has been shown that behavioural control (a construct similar to executive functions) in preschool predicts emotion knowledge 2 years later in primary school (Schultz, Izard, Ackerman, & Youngstrom 2001). For mental emotion understanding the evidence is less clear. In a prospective study, executive functions at 2 and 3 years predicted theory of mind performance at age 3 and 4 respectively, beyond prior theory of mind task performance, age and verbal ability (Muller et al., 2012). At age 4, the theory of mind assessment included an unexpected content belief emotion task where the target's emotion instead of belief was inquired from the child but otherwise contained non emotional tasks (diverse desires, diverse beliefs, level 2 visual perspective taking and unexpected content false belief). Separate results for the belief emotion task were not reported unfortunately. The task however loaded highly on a single factor with the other theory of mind tasks, suggesting that they share a latent construct. In a study with 3- to 6-year-old children (Henning, Spinath, & Aschersleben, 2011), performance in the dimensional change card sort, a task of executive function, was significantly related to an aggregated theory of mind score (consisting of diverse beliefs, diverse desires, knowledge access, unexpected location and unexpected content false beliefs and real-apparent emotions) after controlling for age, sentence comprehension, parental education and child temperament. Looked at separately, executive function and understanding of appearance-reality emotions as well as diverse beliefs were not associated significantly, despite a rather large sample size of 169. As with verbal intelligence, nonverbal intelligence (a sub-division of executive functions, Schellig, Drechsler, Heinemann, & Sturm, 2009) has been implicated in better emotion understanding (Albanese, De Stasio, Di Chiacchio, Fiorilli, & Pons 2010; von Salisch, Haenel, & Freund, 2013), particularly of the more mental and reflective levels.

3.1.7. Consequences of emotion understanding.

3.1.7.1. Mental health.

Deficits in emotion understanding can be found in a wide range of populations with mental health problems or groups at risk for developing a mental disorder (Southam-Gerow & Kendall, 2002). Self-reported difficulty identifying emotions is associated with more mental health problems (Ciarrochi, Scott, Deane, & Heaven, 2003). There is evidence though that in externalization disorders these deficits only apply when understanding the emotional states of oneself as opposed to those of another person (Casey & Schlosser, 1994). In terms of aggression and hard-to-manage behaviours, emotion understanding seems to be more related to the kind of conflict resolution than to the occurrence of it (De Rosnay, Harris, & Pons, 2008). Not least, attribution biases seem to play an important role in emotion understanding deficits in clinical populations (Collin et al., 2013). Emotion knowledge seems to be negatively related to behaviour problems, particularly internalizing problems. Emotion knowledge at age five predicted caregiver-rated behaviour problems and more specifically internalizing but not externalizing problems at age nine in children from economically disadvantaged families after controlling for verbal ability and temperament (Izard et al., 2001). In another longitudinal study with children from economically disadvantaged families, emotion knowledge at age seven predicted self-reported internalizing symptoms at age 11 beyond family income, verbal abilities and teacher reported externalizing as well as internalizing problems at age seven (Fine, Izard, Mostow, Trentacosta, & Ackerman, 2003). The fact that this association has been shown with self-report as well as other-report data further strengthens the evidence. But also physical health and health behaviours are influenced by emotional competence, supposedly through physiological pathways (e.g. autonomous nervous system, neuroendocrine system, immune system), which has been shown in a study using objective proxies like medical expenses, drug consumption or days spent at the hospital (Mikolajczak et al., 2015).

3.1.7.2. Prosocial behaviour and social skills.

Emotion recognition/labelling and situational emotion knowledge in preschool children was associated with prosocial behaviour in a structured behavioural observation (Denham, 1986). In a longitudinal study employing the same tasks, emotion understanding at age 3.5 predicted parent reported prosocial behaviour concurrently and at age 4.5. (Eggum et al., 2011). Emotion recognition/labelling at age five predicted children’s parent rated social skills even 4 years later at age nine (Izard et al., 2001).

3.1.7.3. Peer acceptance / likability.

Several studies have examined the link between socio-metric status (i.e. peer rated likability) and emotion understanding. An early study found an association between emotion recognition and higher peer likability in third to fifth grade children (Edwards, Manstead, & MacDonald, 1984). In a study with preschool children, emotion recognition/labelling and emotional situation knowledge (external causes) predicted peer rated likability and also mediated the relationship between age and likability (Denham, McKinley, Couchoud, & Holt, 1990). In contrast, another study found that emotional situation knowledge but not emotion recognition/labelling predicted peer acceptance in preschool children (Garner, Jones, & Miner, 1994). The conflicting findings of these two studies might be due to the second study controlling for age and maternal emotion socialization practice first before looking at the influence of emotion recognition on peer acceptance. The link between peer likability and emotion understanding is further supported by a study using a different method for the assessment of emotion understanding. Answers in an interview including questions about emotion recognition, causality of and reactions to emotions, predicted children's peer acceptance of kindergarten and first-grade children. Additionally, it influenced the association between maternal and paternal expressiveness and peer acceptance (Cassidy, Parke, Butkovsky, & Braungart, 1992). Results from a more recent study suggest that the effect of emotion knowledge on peer acceptance is mediated by social skills (Mostow, Izard, Fine, & Trentacosta, 2002). That makes sense if one acknowledges that knowledge not necessarily translates into behaviour. Understanding a person’s emotions alone may not be a sufficient condition to act pro-socially or get accepted by peers. The knowledge has to be translated into action. If I don’t know how to functionally respond to the emotions of my peer, my understanding will most likely not help me much in fostering the relationship. Likewise, the motivation to apply them and act pro-socially is crucial and may be influenced by factors such as moral motivation, sympathy (Malti, Gummerum, Keller, & Buchmann, 2009) and empathy (Roberts & Strayer, 1996).

3.1.7.4. Academic competence.

Emotion knowledge at age five predicted academic competence at age nine and also mediated the effect of verbal ability on academic competence (Izard et al., 2001). Another aspect of emotion understanding assumed to influence academic competence is emotion regulation. In a longitudinal study, parents rated children’s emotion regulation competencies at age 4.5. Approximately one year later, academic achievement was tested. Emotion regulation predicted literacy, math and listening achievement even after controlling for IQ and maternal education. It is important to note however, that the emotion regulation subscale of the Emotion Regulation Checklist (Shields & Cicchetti, 1998) used in this study does not directly ask for emotion regulation strategies and also measures aspects of emotion understanding, empathy and equanimity, thus being broader in definition than emotion regulation as assessed with the Test of Emotion Comprehension.

Not many studies have looked at the link between emotion understanding and academic competence in detail but there is evidence that it is mediated by several factors. It is assumed for example that better understanding of emotions leads to higher academic competence through enhanced social skills and better relationships with peers and teachers which help in attaining academic proficiency. One possible mediator is teacher-student relationship. Variables like closeness or quality of relationship between teacher and student have been shown to be associated with both, academic competence and emotion understanding and -regulation (Birch & Ladd, 1997; Graziano, Reavis, Keane, & Calkins, 2007; Garner & Waajid 2008). Another kind of possible mediator is peer acceptance (i.e. likability), measured through socio-metric ratings and peer relationships. The link between emotion understanding and peer acceptance has been shown in numerous studies (see above). In respect to academic achievement studies have also found an association with peer acceptance / relations (Muma, 1965; Austin & Draper, 1984; Ladd, 1990) but the findings are not unequivocal (DeRosier, Kupersmidt, & Patterson 1994). In a longitudinal study, Lecce, Caputi, Pagnin, and Banerjee (2017) found that social competence (which included peer nominations) mediated the relationship between social understanding (comprised of theory of mind and emotion understanding measures) in preschool and later academic achievement in early primary school. Another longitudinal study found that low peer acceptance in fourth grade indirectly influenced low academic achievement in sixth grade through low academic self-concept and more internalizing symptoms in fifth grade (Flook, Repetti, & Ullman, 2005). These findings emphasize how constructs like peer acceptance, that are suspected of mediating the influence of emotion understanding on academic achievement, can themselves be subject to mediation by other factors. When considering the known association of emotion understanding and internalizing problems (see chapter 3.1.7.1) it becomes evident that we are presumably dealing with a very complex mode of bidirectional influences and feedback loops, which we have only began to investigate in parts. Finally, attention to academic tasks has recently been investigated as a potential mediator between emotion understanding and academic achievement (Trentacosta, Izard, Mostow, & Fine, 2006). In a recent study (Trentacosta & Izard, 2007) the mediating effects of these three factors, teacher-student closeness, peer likability and attention on the link between emotion knowledge, emotion regulation and academic competence have been examined applying path-analyses. Emotion knowledge in the last year of kindergarten predicted academic achievement in first grade of school, while an indirect effect of teacher-rated emotion regulation, mediated through teacher attention ratings, was found as well. More specifically, emotion knowledge predicted reading but not mathematical or spelling achievement. It should be noted that this effect was independent of verbal ability which itself directly predicted academic achievement.

3.1.8. Goal of study 2 – developing and evaluating the computerized Test of Recognizing and Understanding Emotions (cTRUE)

The Test of Emotion Understanding has proven to be a useful and comprehensive measure to assess emotion understanding in children. Despite its usefulness, the procedure can be improved for older children. Several studies have shown that most children aged 9 or above score correctly on all but one or two components of the TEC as expected according to the theory (Pons et al., 2004; Albanese, De Stasio, Di Chiacchio, Fiorilli & Pons, 2010; Ornaghi & Grazzani, 2013). There is growing evidence however that understanding of mental states, even if present on a basic level, continues to develop and that performance on many continuous tasks of advanced theory of mind is not at ceiling even for adults (Miller, 2012). Two basic components (recognition and external causes), although consisting of several items, are routinely scored dichotomously. The problem of dichotomizing continuous variables is an associated loss of power (Cohen, 1983). Furthermore, without denying early conceptual milestones that might be sufficiently captured by categorical assessment, a continuous view on mental understanding across the lifespan is more compatible with recent findings (Apperly, 2011; Apperly, Warren, Andrews, Grant, & Todd, 2011; Miller, 2012; Lagattuta et al., 2015) and might provide additional and meaningful information.

We (Haslinger & Leyrer, 2013) developed a new experimental task of emotion understanding, based on the Test of Emotion Comprehension (Pons, Harris & de Rosnay, 2000). There were three main goals in the development of the computerized Task of Understanding and Recognizing Emotions (cTRUE). The first goal was to avoid ceiling effects in the higher echelons of competence and to measure the level of performance in the different components of emotion understanding non-dichotomously to be able to investigate individual differences and relationships with related variables more thoroughly. This goal should be reached by including more items (see Table 14 for an overview), thus leading to less correct answers per chance and more variance compared to a dichotomous measurement. Particularly more difficult components like beliefs, hiding, regulation or mixed emotions were represented by five items instead of a single item. Secondly, cartoons should include different emotions to account for the findings of varying difficulty in the labelling of different emotions (Denham & Couchoud, 1990; Wintre & Vallance, 1994). Thirdly, the inclusion of response times was thought to provide additional useful information beyond accuracy scores. Response times are routinely included in the scoring schema of most performance tests of cognitive functions (e.g. attention, executive functions, etc.). In the scientific and practical assessment of abilities like theory of mind or emotion understanding, response times are usually neglected. Recently however there has been a growing interest in the speed aspect of performance in theory of mind (e.g. Apperly, Samson, & Humphreys, 2009; Dumontheil, Apperly, & Blakemore, 2010; McCleery, Surtees, Graham, Richards, & Apperly, 2011). In respect to emotion understanding, the investigation of response times has mainly been limited to the field of emotion recognition (e.g. Kestenbaum & Nelson, 1992; Fink, de Rosnay, Wierda, Koot, & Begeer, 2014). One might ask why the measuring of response times should be of any interest when dealing with theory of mind or emotion understanding. On the one hand it is of a purely theoretical interest. No performance based account of a skill or ability can be complete without considering speed differences between populations or within a person between different contexts. The consideration of response times is even more important for studies where the time course of neurophysiological processes is compared to the behavioural time course. On the other hand imagine a social interaction where understanding the causes of a person's emotions is warranted. As the understanding of the mental states of the other person grows, the verbal and behavioural responses are adjusted accordingly, taking into account that knowledge. Considering the fast pace of communication where a response may be appropriate in one moment and obsolete a second later, it makes a difference if that knowledge is available sooner or later. Yet to my knowledge, there is no other comprehensive measure of emotion understanding to this date, utilizing response times.

In summary, goal of study two was to develop an adapted version of the Test of Emotion Comprehension that fulfilled following criteria: (1) increased variance and sensitivity in older children, (2) coverage of a larger range of social situations and emotion outcomes, (3) inclusion of response time measurement. Subsequently the reliability and validity of the task and its relation to other measures of emotion understanding and socio-cognitive competence was to be evaluated in a sample of primary school children.

3.2. Methods

3.2.1. Participants.

Sixty-one children were recruited through two primary schools and two day care centres in Linz and its environs (Austria). School and centre authorities were contacted through phone and personally. Informed consent and assent was acquired from the parents and children respectively. On the first four children, the procedure was tested and it became apparent that the cTRUE was too long for the children's attention span. Thus items were eliminated to make the procedure more comfortable for the children and data from the first four children was not used in subsequent analyses. Two other children were excluded, one because it scored lower than two standard deviations below the mean on the vocabulary test and another child because it had only been in Austria and learning German for a year. The final subject pool consisted of 55 children (26 female) aged 6;8 to 11;11 (M=8.88, SD=1.35). Age and gender distributions are shown in Table 11.

Table 11
Age and gender distribution of the sample of study 2

age (years)	total	female	male
6	1	1	0
7	16	9	7
8	18	10	8
9	8	1	7
10	6	3	3
11	6	2	4
Total	55	26	29

Complete caregiver ratings could only be collected by 41 participants (19 female) with a mean age of 8.64 years (*SD*=1.28) because one school declined to fill out the forms due to time restraints of the teachers and one questionnaire which was not completed by the teacher. Age and gender distributions of the sample for which caregiver ratings are available are shown in Table 12.

Table 12
Age and gender distribution of the sample for social skills, social-roles and social competence

age (years)	total	female	male
6	1	1	0
7	9	4	5
8	16	9	7
9	7	1	6
10	6	3	3
11	2	1	1
Total	41	19	22

For the Participants Roles Questionnaire the sample was identical except that the questionnaire from a 9-year-old boy, for which the SSIS was missing was completed (*M*=8.86, *SD*=1.21). For the FASC, data from nine subjects was lost due to audio recording failure resulting in a sample of 46 participants (21 female) with a mean age of 8.47 (*SD*=1.31, see Table 13).

Table 13
Age and gender distribution of the sample for FASC

age (years)	total	female	male
6	0	0	0
7	16	9	7
8	12	7	5
9	8	1	7
10	4	2	2
11	6	2	4
Total	46	21	25

3.2.2. Procedure.

Children were individually assessed in their respective school or after school centre facilities in a quiet room between March and July 2014 by a master student (Stefanie Reichetseder) under supervision of the author. Children from primary schools were assessed during their normal school schedule at noon while children from school care centres were assessed in the afternoon. Before starting the standardized assessment the experimenter explained the procedure and established a friendly atmosphere. To counter possible training effects from TEC and cTRUE, presentation order was randomized with the software Mix (Van Casteren & Davis, 2006). The session-duration did not exceed 90 minutes and breaks were implemented on an individual basis as required. Additionally teachers or caretakers filled out two questionnaires rating social skills and social role behaviour of the child.

3.2.3. Measures.

3.2.3.1. Test of Emotion Comprehension.

The Test of Emotion Comprehension is a task to assess nine components of emotion understanding: recognition of facial emotions, external causes of emotions, memorial cues as triggers for emotions, desire- and belief-based emotions, hiding emotions, emotion regulation, mixed emotions and moral emotions. The experimenter reads out scenarios assisted by simple cartoons. The child is then asked how the character in the story feels and responds by pointing to one of four cartoon facial expression. These four response options may include the emotions happy, sad, scared, angry and alright. The response format for three components differs: For emotion recognition the experimenter reads one of the five emotion words and the child has to point to the corresponding drawn face on the response sheet. Answers for emotion regulation target behaviours (e.g. cognitive regulation) and mixed emotion include answers with two emotions (e.g. happy and scared). All components consist of a single item, except recognition and external causes which consist of five items each. Components are scored dichotomously (with a cut-off of three for recognition and external causes) and summed up for a total score of emotion comprehension ranging from 0 to 9. For a detailed description of each component see Pons, et al. (2004).

3.2.3.2. Computerized Task of Emotion Understanding.

The Computerized Task of Emotion Understanding was very similar to the Test of Emotion Comprehension in procedure and content. The sequence of components was identical to the TEC, the item sequence within components was randomized. Components were scored as quotients and the total emotion understanding score was computed by adding all nine component scores yielding a continuous score of 0 to 9 to make it comparable to the discrete score of the TEC.

There were also a few important distinctions however. First, the Computerized Task of Emotion Understanding was administered on a computer via E-Prime® 2.0.8.90 (Psychology Software Tools, Inc., 2010). All verbal instructions and pictures were previously recorded and presented through the software but the experimenter was present in case of technical difficulties or problems of understanding. Responses were made by the child with a computer mouse. No child had difficulties controlling the mouse. The initial instruction for the task (originally in german) was as follows:

> Hello! You will now be shown a couple of pictures and asked some questions about them. Your goal is to find out how the children on the pictures might feel. You can respond by selecting the face that matches the feeling. You can also respond before the story has ended if you believe to know the answer. After the story has ended, select the face that shows how the child in the story feels. Select a face even if you have already selected a face during the story. Find out now how the children on the pictures feel and select next to start. (Haslinger & Leyrer, 2013)

Figure 11 shows examples of the five facial expressions represented by the three male and three female characters which were used as the response faces in the male and female version of the cTRUE.

Figure 11. Sample emotion faces used as response options in the cTRUE

Second, each component contained more items than the TEC (see Table D 1 in Appendix D for a list of covered emotions). Particularly more difficult components like beliefs, hiding, regulation or mixed emotion*s* were represented by five items. Components recognition and external causes contained 15 instead of five items, desires two and reminder three vignettes. All cartoons were conceived and created by a master student (Haslinger, 2014) under supervision of this author (who also contributed ideas for the social situations depicted in the vignettes) with the web-tool Bitstrips (www.bitstrips.com) under the following rules: New items should be as similar (e.g. in terms of complexity, length) to the original items of the TEC as possible, include situations primary school age children might be familiar with or at least can relate to, and encompass a variety of four basic emotions (happy, sad, angry, scared). Original items from the TEC were also re-created to asses, if the newly devised cartoons were comparable to the established cartoons in terms of item characteristics. Additionally it was of interest whether the items present in both tasks differed in terms of difficulty, suggesting an influence of task-format (paper vs. computer). Third, response times were collected after the prompt that followed the end of the cartoon (e.g. "How does [name of character] feel. Does he/she feel sad, angry, scared or alright?"). Only response times for correct responses were used. Data were corrected for outliers (see chapter 3.2.4). Response times were averaged for each person per component and a grand mean over all components was calculated for the average response time of the total emotion understanding score. See Figure 12 for an example of a cartoon for *external causes* and Appendix D for example vignettes of the other components.

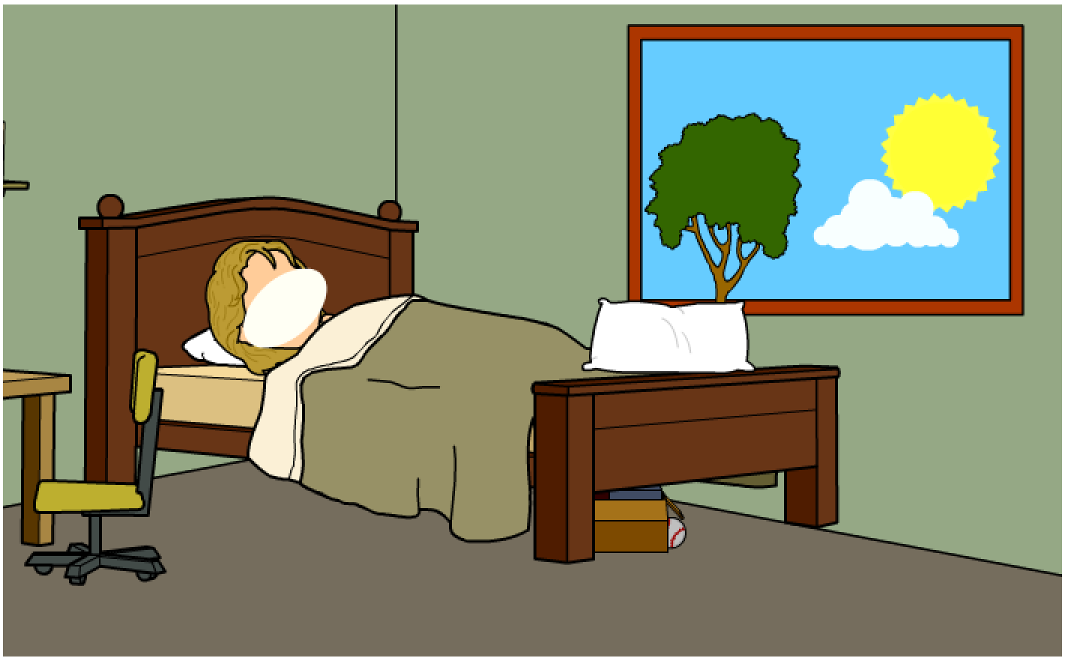

Figure 12. CTRUE example item of the external-causes component.

The description for this vignette is as follows: This girl is sick in bed and outside the weather is really nice. How does the girl feel? Does she feel happy, sad, just ok or scared?

3.2.3.3. Emotion knowledge.

Emotion knowledge was assessed similarly to previous studies (e.g. Schultz, Izard, & Bear, 2004) through three measures. In the emotion labelling task children had to label 24 pictures of facial expressions for four simple emotions (happy, sad, angry and scared) and for eight ambiguous expressions. Because the author of the original task did not respond to a request for the stimulus material, pictures were taken from the NimStim set of facial expressions (http://www.macbrain.org/resources.htm) for which ratings by 81 adult subjects are available (Tottenham et al., 2009). Facial expressions were used for the simple emotions category if the reported recognition rate for the intended emotion was above 80% and for the ambiguous category if it was below 50%. The social situations and social behaviour tasks each consisted of 15 one to three sentence scenarios, three for the emotions happy, sad, angry and scared and three for ambiguous scenarios. Scenarios were taken from the appendix of Schultz et al. (2004). Ambiguous items were used in the aforementioned study to investigate emotion attribution biases but were not analysed in the current study. They were however included to keep the procedure comparable. An overall emotion attribution accuracy score was computed by summing the means of the three separate scales (recognition, social situations, and social behaviours). Internal consistency of the emotion attribution accuracy score in the present study was Cronbach $\alpha = .82$ compared to $\alpha = .68$ and $\alpha = .75$ in previous studies (Schultz et al., 2004; Trentacosta & Izard, 2007) suggesting the reliability of the modified task used in this study. To use a comparable scoring scheme to the cTRUE and not have facial emotion recognition (which consists of more items) overrepresented, accuracy-ratios were computed for each sub-task and the three scores summed for a total emotion attribution score used in all analyses.

3.2.3.4. Multiple emotions.

Children were given four scenarios from the multiple emotions task (Meerum Terwogt, Koops, Oosterhoff, & Olthof, 1986) which was taken from Rieffe, Meerum Terwogt and Kotronopoulou (2007). They were asked to imagine how they would feel in the given situation: sad, angry, happy or scared? In a prior example story they were explicitly instructed about the possibility of people experiencing multiple emotions simultaneously. In a variation to the original task the four drawings of facial expressions to respond to were taken from the TEC and the story vignettes were not accompanied by drawings since the material was not obtainable by request. Children were awarded a point for each scenario in which they responded with both correct emotions in accordance with the coding scheme in Beck, Kumschick, Eid and Klann-Delius (2012).

3.2.3.5. Flexibility and Automaticity of Social Cognition.

For a detailed task description see study 1. Because of the lengthy and extensive test battery, only two vignettes were used. To ensure maximum comprehension in this diverse age range, only non-verbal vignettes were considered and one unambiguous (*restaurant*) and one ambiguous (*computer class*) item from study one selected.

3.2.3.6. Social Skills Improvement System.

Teachers or school caregivers completed the teacher version of the Social Skills Improvement System (Gresham & Elliot 2008) which is the successor of the Social Skills Rating System. It consists of three aggregated scales: social skills, problem behaviours and academic competence. The social skills and problem behaviour items are rated on 4-point scale: *never*, *seldom*, *often*, and *almost always*. Items of academic competence are rated in terms of relative rank performance in the classroom: lowest 10%, next lowest 20%, middle 40%, next highest 20% or highest 10%. Because no German version of the questionnaire was available to date, it was translated by a master student and back-translated by an English native speaker who was also fluent in German. For items that are also present in the Social Skills Rating System, an existing German translation was used (Hess et al. 2013).

3.2.3.7. Social competence scale of the Youth Self Report.

To obtain a measure of self-rated social competence items of the social competence scale of the Youth Self Report were presented in an interview. Originally the Youth Self

Report is intended for children aged 11years and older but several studies support the validity of this measure in children younger than 11 years of age (Kolko & Kazdin, 2003; Yeh & Weisz, 2001; Ebesutani, Bernstein, Martinez, Chorpita, & Weisz, 2011), at least for the broad symptom scales.

3.2.3.8. Participant Roles Questionnaire.

The Participant Roles Questionnaire (Belacchi, 2008) assesses pro-social and pro-bully role behaviour through other ratings. Teachers or school caregivers rated how frequently the child shows behaviours indicative of pro-social (consoler, defender, mediator) pro-bully (bully, assistant, reinforcer), victim or outsider roles. The questionnaire was taken from the appendix of Belacchi and Farina (2010) and translated into German. Reported measures of reliability are .94 for prosocial roles, .92 for pro-bully roles, .62 for outsider, and .68 for victim (Belacchi & Farina, 2010).

3.2.3.9. EEG

Children also underwent a short procedure designed to assess EEG reactivity to cognitive load, positive / negative affect, frustration, and empathy (calibration; G. Siegle, personal communication, October 11, 2013). EEG was recorded with the Emotive Epoc headset comprising 14 electrodes during the calibration procedure and completion of the cTRUE. The EEG recording is not part of this study and therefore not further referenced.

3.2.3.10. Verbal intelligence / vocabulary.

To control for verbal intelligence the German version of the vocabulary subscale from the Wechsler Intelligence Scale for Children IV (Petermann & Petermann, 2011) was administered. See chapter 2.2.3.3 for a detailed description of the measure.

3.2.4. Data inspection and processing.

Outcome variables (except for those that were clearly non-normal by visual inspection) were subjected to outlier analysis and extreme outliers (two times the inter-quartile-range) corrected by replacing them with the next non-extreme outlier value + 1 unit. Most cTRUE component scores (accuracy and response times, e.g. regulation), participant roles scores and FASC scores proved to be non-normally distributed (particularly skewed, see Table E 1 in Appendix E), thus Spearman rank correlations were chosen in all analyses including these scores. Response time distributions are regularly skewed to the right. That means, frequency of responses rises quickly and then drops slowly with a long right tail (Ratcliff, 1993). This is because there is a lower boundary for responses given by limits of information processing speed and motor reaction time. On the right end of the distribution however, there is a spread caused by fluctuations in attention, inter-individual range on abilities tapped by the task and different response strategies (e.g. impulsive vs. reflective). A liberal approach of outlier correction was chosen. Threshold for outliers at the right end of the distribution was defined as the upper quartile + 2 * inter-quartile range which pertained to 4% of all correct responses. Outliers were windsorized and not eliminated because of the small number of trials per component compared to traditional response time experiments. Additionally, late responses in complex tasks like this one are more likely to occur in the context of cognitive task demands (i.e. reasoning about the correct answer) than outstanding responses in simple perception tasks. Early responses were not subjected to this criterion since due to the task design the response decision could be formed prior to the end of the prompt thus very fast responses cannot be considered spurious.

3.3. Results

3.3.1. Descriptives.

Table 14

Descriptive statistics of study 2

Variable	*n*	*M*	*SD*	min	max
Age (years)	55	8.88	1.35	6.72	11.80
Vocabulary (stand)	55	10.51	3.27	4.00	17.00
cTRUE	55	7.18	1.09	3.33	8.93
cTRUE: RT (ms)	55	1683.00	457.00	979.00	2826.00
TEC	55	6.91	1.39	4.00	9.00
Emotion Knowledge	55	29.78	4.64	16.00	37.00
MET	55	0.45	0.86	.00	3.00
FASC: CRT (s)	46	30.13	15.72	4.91	81.53
FASC: IRT (s)	46	3.36	1.56	.79	7.72
FASC: ORT (s)	46	22.89	16.19	3.90	75.05
FASC: TR	46	2.83	1.10	1.00	6.00
FASC: MSJ	46	1.74	1.06	.00	5.00
FASC: MJTRr	46	.60	.32	.00	1.00
FASC: word-count	46	45.48	35.51	6.00	147.00
FASC: IST	46	3.22	2.88	.00	11.00
FASC: ISTr	46	.07	.05	.00	.20
YSR: Social Competence	55	5.97	1.65	2.67	9.67
SSIS: Social Skills	41	90.76	18.74	32.00	127.00
SSIS: Problem Behaviors	42	23.81	13.53	2.00	54.00
SSIS: Academic Competence	42	25.02	6.86	7.00	35.00
PRQ: Prosocial	42	7.75	1.94	3.00	11.33
PRQ: Probully	42	5.47	1.71	3.00	10.00
PRQ: Bully	42	4.93	2.29	3.00	11.00
PRQ: Reinforcer	42	6.90	1.74	3.00	10.00
PRQ: Assistant	42	4.57	1.76	3.00	11.00
PRQ: Defender	42	7.81	2.24	3.00	12.00
PRQ: Consoler	42	7.74	2.24	3.00	12.00
PRQ: Mediator	42	7.69	2.11	3.00	11.00
PRQ: Victim	42	4.17	1.51	3.00	9.00
PRQ: Outsider	42	7.36	1.85	3.00	12.00

Note. CRT= cartoon reading time, cTRUE = Computerized Task of Recognizing and Understanding Emotions, FASC = Flexibility and Automaticity of Social Cognition, IRT = initial response time, IST = sum of internal state terms, ISTr = ratio of IST to total number of words, MET = Multiple Emotions Task, MSJ = sum of mental justifications, MSJTRr = ratio of MSJ to sum of total responses, ORT = overall response time, PRQ = Participant Roles Questionnaire, RT = response time (in ms), SSIS = Social Skills Improvement System, TEC = Test of Emotion Comprehension, TR = sum of total responses, stand = standard score

See Table E 2 for descriptive statistics of subscales / components of emotion understanding tasks.

3.3.2. Reliability.

Reliability for the cTRUE total score was Cronbach's Alpha=.81. For the component scores of *mixed emotions* and *emotion regulation* reliability was .72 and .70, for *hiding*, *recognition* and *desires* it was .62, .55, and .54, for *external causes* and *morality* it was .44 and .31 while for *beliefs* and *reminder* it was .16 and .11. After removing the reconstructed TEC-items from the cTRUE-score, Cronbach's Alpha was still .77. Randomly removing an identical number of new items instead lowered the internal consistency to .75. The internal consistency for the means of the nine emotion understanding components was .60. In comparison, Cronbach's Alpha for the nine components of the TEC was .28. To investigate how reliability of each component could be improved, an optimized version was computed by removing items with a negative item-component correlation and subsequently balancing emotion categories. This led to a substantial reduction of items while retaining the same level of overall scale-reliability and improving reliability in several components. Since these changes were made post-hoc on the basis of the current sample it would not be sound to use it in the subsequent analyses. This supposedly improved version has to be validated with a new sample. See Table 15 for a summary of reliability scores.

Table 15
Reliability of cTRUE components

	Original cTRUE		cTRUE excl. TEC items		cTRUE excl. random items		cTRUE optimized	
cTRUE component	items	α	items	α	items	α	items	α
cTRUE: total	54	.81	41	.77	42	.75	40	.80
cTRUE: recognition	11	.55	11	.55	11	.55	7	.65
cTRUE: ext. causes	15	.44	10	.35	10	.58	10	.55
cTRUE: reminder	3	.11	2	.20	2	.06	2	.20
cTRUE: desires	2	.54	1	-	1	-	2	.54
cTRUE: beliefs	5	.15	4	.41	4	.02	3	.42
cTRUE: hiding	5	.62	4	.58	4	.57	5	.62
cTRUE: regulation	5	.70	4	.65	4	.63	5	.70
cTRUE: mixed	5	.72	4	.64	4	.62	4	.77
cTRUE: morality	3	.31	2	-.19	2	.42	2	.42

Note. For some components, the number of items in the table differs from the number of items used in the task as items that had an accuracy of 1 were not included in the calculation of Cronbach's Alpha. cTRUE = Computerized Task of Recognizing and Understanding Emotions

3.3.3. Item analysis.

Mean item difficulty was .84 (*SD*=.14) and ranged from .42 to 1 (Median=.87). Item-scale correlations varied between -.20 and .59 (Median=.24, *M*=.23, *SD*=.19). Three items from the recognition component, and one from belief based emotions and morality had, negative item-scale correlations. One has to keep in mind however, that all but one of these items had a difficulty of .96 or .98, so one spurious response can skew the correlation drastically. The only item with a negative item-scale correlation that did not show near-zero variance was the morality-item recreated from the TEC. See figure Table E 3 in Appendix E for a table with item statistics.

3.3.4. Convergent validity.

Table 16 shows inter-correlations between tasks of emotion understanding as well as age and vocabulary. Correlations between the cTRUE and TEC as well as emotion knowledge were high. There was no significant correlation between the multiple emotions task and any other task of emotion understanding. This might be in part due to 73% of all subjects scoring zero on this task. The cTRUE mean response time was only significantly negatively related to emotion knowledge. Vocabulary was significantly related to all tasks of emotion understanding while only cTRUE and emotion knowledge were significantly associated with age.

Table 16
Inter-correlations of tasks of emotion understanding

Measure	cTRUE	cTRUE: new	cTRUE: RT	TEC	EK	MET[1]
cTRUE: new	.94**	-	-	-	-	-
cTRUE: RT	-.24	-.12	-	-	-	-
TEC	.57**	.54**	-.20	-	-	-
EK	.54**	.57*	-.33*	.53**	-	-
MET[1]	.16	.09	-.18	.09	.24	-
Age	.31**	.38**	-.26	.19	.34*	.03
Voc	.37**	.41**	-.09	.55**	.54**	.31*

Note. N=55. cTRUE = Computerized Task of Recognizing and Understanding Emotions, cTRUE:new = cTRUE excluding recreated TEC items, RT = response times, TEC = Test of Emotion Comprehension, EK = emotion knowledge, MET = Multiple Emotions Task, Voc = Vocabulary (standard score)
[1]rang correlation (spearman)
* p<.05, ** p<.01

Comparing the total score of cTRUE for the original TEC items that were recreated vs. the new items (excluding *recognition* where the new/old distinction is not applicable), both item sets correlated strongly ($r(55)=.63$, $p<.01$) with each other and comparably high with the TEC total score (new vignettes: $r(55)=.54$, $p<.01$; old TEC-vignettes: $r(55)=.49$, $p<.01$) which is similar to the correlation of the TEC and the cTRUE total score including all items ($r(55)=.57$, $p<.01$). This suggests that overall, the new items were comparable in terms of convergent validity.

For individual components correlations between TEC and cTRUE, as well as cTRUE scores computed only from new vignettes can be seen in Table 17 below.

Table 17
Correlations between TEC and cTRUE component scores

TEC	cTRUE	cTRUE new vignettes
1. Recognition	.03	.03
2. Ext. causes	.05	.05
3. Reminder	.24^{+}	.04
4. Desires	.61**	.50**
5. Beliefs	.24^{+}	-.05
6. Hiding	.37*	.30*
7. Regulation	.52*	.56**
8. Mixed emotion	.66**	.58**
9. Morality	.57**	.43**

Note. N=55. cTRUE = Computerized Task of Recognizing and Understanding Emotions, TEC = Test of Emotion Comprehension
$^{+}$ p<.10, * p<.05, ** p<.01

3.3.5. Concurrent validity.

3.3.5.1. Social competence.

To establish concurrent validity with variables thought to be related to emotion understanding, correlations between cTRUE scores and the Social Skills Improvement System (SSIS), a teacher reported questionnaire of social skills, problem behaviours and academic competence, as well as the social competence sub-scale of the Youth Self Report were computed (see Table 18). Neither social skills nor problem behaviours, academic competence or self-reported social activities were significantly correlated with any task of emotion understanding except the multiple emotions task which was positively correlated with social skills and negatively with problem behaviours. Regarding individual emotion understanding components, there was a trend association between emotion regulation and academic competence and mixed emotions and both, social skills and academic competence. The component morality was negatively related to self-reported social activities.

Table 18

Correlations between tasks of emotion understanding and social skills

Measure	SSIS:SS	SSIS:PB	SSIS:AC	YSR
cTRUE	.07	-.03	.21	-.09
TEC	.17	.15	.05	-.14
EK	.14	.06	.15	.06
MET[1]	.41**	-.31*	-.04	.02
cTRUE: recognition[1]	.12	-.11	-.05	.01
cTRUE: external causes[1]	-.10	.07	.25	-.07
cTRUE: reminder[1]	.19	-.20	-.06	.25
cTRUE: desires[1]	-.25	.19	-.06	-.04
cTRUE: beliefs[1]	.16	-.12	-.05	-.03
cTRUE: hiding[1]	.17	-.14	.24	.06
cTRUE: regulation[1]	-.06	-.10	.28^{+}	-.06
cTRUE: mixed emotions[1]	.28^{+}	-.06	.27^{+}	-.13
cTRUE: morality[1]	-.15	.03	.12	-.35**

Note. N=42, except n=41 for SSIS-SS and n=55 for YSR. cTRUE = Computerized Task of Recognizing and Understanding Emotions, TEC = Test of Emotion Comprehension, EK = emotion knowledge, MET = Multiple Emotions Task, SSIS = Social Skills Improvement System, SS = social skills, PB = problem behaviour, AC = academic competence, YSR = Youth Self Report

[1]rank correlation (Spearman)

$^{+}$ p<.10, * p<.05, ** p<.01

For cTRUE response times, there was a negative correlation between external causes response times and academic competence and significant positive and correlations between reminder and social skills and problem behaviours respectively (see Table 19).

Table 19

Correlations between cTRUE response times and social skills

Measure	SSIS:SS	SSIS:PB	SSIS:AC	YSR
cTRUE	-.06	.08	-.23	-.05
cTRUE: recognition[1]	.02	-.14	.01	-.00
cTRUE: external causes[1]	.07	-.04	-.41**	-.09
cTRUE: reminder[1]	.38*	-.41**	-.03	-.02
cTRUE: desires[1]	-.15	.19	-.19	.21
cTRUE: beliefs[1]	.05	-.02	-.10	-.15
cTRUE: hiding[1]	-.13	.14	-.19	-.05
cTRUE: regulation[1]	-.09	.05	-.25	.11
cTRUE: mixed emotions[1]	.04	.03	-.22	-.15
cTRUE: morality[1]	-.11	.27	.10	-.09

Note. N=42, except n=41 for SSIS-SS and n=55 for YSR. cTRUE = Computerized Task of Recognizing and Understanding Emotions, SSIS = Social Skills Improvement System, SS = social skills, PB = problem behaviour, AC = academic competence, YSR = Youth Self Report

[1]rank correlation (Spearman)

$^{+}$ p<.10, * p<.05, ** p<.01

For the TEC, there was a significant correlation between mixed emotions and social skills (see Table E 4).

3.3.5.2. Participant roles.

There were several significant or highly significant correlations between tasks of emotion understanding and participant roles (see Table 20). Teacher-reported prosocial roles were significantly related to the TEC, emotion knowledge and multiple emotions task. There was a correlation at trend level with the cTRUE. Pro-bullying roles were negatively related to the multiple emotions task. The role of outsider, not included in the prosocial or pro-bully scores, was negatively related to the cTRUE total score. The role of victim was positively correlated with the TEC and emotion knowledge scores and at a trend level negatively related with the multiple emotions task. On a component level, emotion regulation and mixed emotions from the cTRUE were positively correlated with prosocial-roles. Mixed emotions understanding was also negatively related to the role of outsider. Surprisingly desire based emotion understanding was positively associated with pro-bully roles.

Table 20
Rank correlations (Spearman) between tasks of emotion understanding and participant roles

Measure	PS	D	C	M	PB	B	R	A	O	V
cTRUE	.26$^+$	.38*	.11	.18	.09	.13	-.01	.14	-.33*	.21
TEC	.38*	.47**	.30$^+$	.16	.00	.17	-.17	-.00	-.20	.38*
EK	.46**	.44**	.36*	.44*	.15	.21	.08	.13	-.15	.36*
MET	.31*	.30$^+$	.20	.24	-.33*	-.25	-.32*	-.30$^+$	-.09	-.29$^+$
cTRUE										
Recogn.	.07	-.05	.11	.04	-.12	-.04	-.25	-.02	.07	-.08
Ext. causes	.29$^+$	.37*	.13	.32*	.22	.26	.10	.28$^+$	-.09	.19
Reminder	.27$^+$	.18	.25	.27$^+$	-.26	-.17	-.23	-.26$^+$	-.11	-.02
Desires	-.24	-.01	-.35*	-.31*	.34*	.28*	.25	.45**	.03	.15
Beliefs	.15	.16	.01	.18	-.05	-.07	-.06	.00	-.09	.00
Hiding	.22	.17	.13	.17	-.01	-.02	.03	-.02	-.23	-.03
Regulation	.31*	.27	.14	.42**	-.04	-.02	.02	-.08	-.25	-.03
Mixed e.	.34*	.39**	.25	.22	.00	.11	-.19	.04	-.37*	.18
Morality	.01	.11	-.09	.01	-.01	.02	-.04	.00	-.04	.13

Note. N=42. cTRUE = Computerized Task of Recognizing and Understanding Emotions, TEC = Test of Emotion Comprehension, EK = emotion knowledge, MET = Multiple Emotions Task, PRQ = Participants Roles Questionnaire, PS = prosocial roles, D = defender, C = consoler, M = mediator, PB = pro-bullying roles, B = bully, R = reinforcer, A = assistant, O = outsider, V = victim
$^+$ p<.10, * p<.05, ** p<.01

For cTRUE response times, reminder was negatively associated with victim, desires based emotion understanding was positively correlated with reinforce and assistant while emotion regulation was negatively related to prosocial roles, in particular mediator (see Table 21)

Table 21
Rank correlations between cTRUE response times and participant roles

Measure	PS	D	C	M	PB	B	R	A	O	V
cTRUE	-.15	-.10	-.11	-.18	.04	.01	.09	.01	.15	.05
Recognition	-.22	-.23	-.12	-.19	-.12	-.23	.05	-.14	.14	-.11
External causes	-.27+	-.23	-.14	-.30+	-.12	-.16	-.06	-.11	.20	.01
Reminder	.14	.10	.11	.07	-.21	-.21	-.10	-.23	-.07	-.40**
Desires	-.09	.00	-.08	-.17	.28+	.13	.31*	.30*	-.03	.28
Beliefs	.02	.06	-.06	-.01	.00	.03	-.01	.04	-.01	.02
Hiding	-.07	.00	-.06	-.11	.13	.13	.03	.16	.01	.22
Regulation	-.31*	-.21	-.24	-.34*	-.10	-.02	-.17	-.10	.21	-.01
Mixed emotions	-.08	.11	-.14	-.12	-.06	-.04	-.06	.02	.10	.09
Morality	.07	-.13	.17	.16	.09	.17	.08	-.04	.06	.13

Note. N=42. cTRUE = Computerized Task of Recognizing and Understanding Emotions, TEC = Test of Emotion Comprehension, EK = emotion knowledge, MET = Multiple Emotions Task, PRQ = Participants Roles Questionnaire, PS = prosocial roles, D = defender, C = consoler, M = mediator, PB = pro-bullying roles, B = bully, R = reinforcer, A = assistant, O = outsider, V = victim
$^{+}$ p<.10, * p<.05, ** p<.01

3.3.5.3. Social cognition.

There were no significant correlations between cTRUE total score and FASC (see Table 22).

Table 22
Rank correlations between tasks of emotion understanding and FASC

Measure	MSJ	MSJTRr	IST	ISTr	IRT
cTRUE	-.19	-.10	.00	.02	-.07
TEC	-.04	.01	.16	.09	.04
EK	-.20	-.11	-.00	.05	-.25+
MET	.26+	.25+	.37*	.22	-.16
cTRUE: recognition	-.29+	-.10	-.14	-.05	-.07
cTRUE: external causes	-.19	-.12	-.21	-.14	-.07
cTRUE: reminder	.14	.20	.19	.23	-.05
cTRUE: desires	-.09	.01	.05	.05	-.04
cTRUE: beliefs	-.10	.04	.07	.12	.26+
cTRUE: hiding	-.36*	-.20	-.18	.02	-.09
cTRUE: regulation	-.07	-.12	.03	.14	-.22
cTRUE: mixed emotions	-.07	-.00	.15	-.11	-.22
cTRUE: morality	.09	.00	.07	-.00	.06

Note. N=46. cTRUE = Computerized Task of Recognizing and Understanding Emotions, TEC = Test of Emotion Comprehension, EK = emotion knowledge, MET = Multiple Emotions Task, IRT = initial response time, MSJ = sum of mental state justifications, MSJTRr = ratio of MSJ to sum of total responses, IST = sum of internal state terms, ISTr = ratio of IST to total number of words
$^{+}$ p<.10, * p<.05, ** p<.01

Neither TEC nor emotion knowledge was significantly related to any FASC variables although there was a negative trend for significance between emotion knowledge and the initial response time. The multiple emotions task was significantly related to the total number of internal state terms used and there was a trend for an association with the number of mental justifications and the ratio of mental to total justifications. Looking at the individual cTRUE-components, hiding emotions was negatively related to the number of mental state justifications. There was a trend for recognition to be negatively related to the number of mental state justifications and belief-based emotions to be positively related to the mean initial response time.

There was a significant negative correlation between the cTRUE mean response time and the ratio of internal state terms to the total number of words contained in the answers, mainly carried by negative associations with mixed emotions and morality (see Table 23). Morality was negatively related to all mental state justification and mental state term based FASC variables and recognition was negatively linked to mental state justifications.

Table 23
Rank correlations between cTRUE response times and FASC

Measure	MSJ	MSJTRr	IST	ISTr	IRT
cTRUE	-.15	-.09	-.22	-.32*	.19
cTRUE: recognition	-.32*	-.20	-.24	-.08	.14
cTRUE: external causes	.00	-.09	-.13	-.08	.16
cTRUE: reminder	.04	.14	-.02	-.18	.04
cTRUE: desires	-.04	.07	.01	-.02	.29+
cTRUE: beliefs	-.08	.01	.00	-.12	.14
cTRUE: hiding	.25	.25	.14	-.07	.21
cTRUE: regulation	-.15	-.11	-.21	-.15	.11
cTRUE: mixed emotions	-.15	-.12	-.27	-.37*	-.01
cTRUE: morality	-.33*	-.37*	-.36*	-.36*	-.11

Note. n=46. cTRUE = Computerized Task of Recognizing and Understanding Emotions, IRT = initial response time, MSJ = sum of mental state justifications, MSJTRr = ratio of MSJ to sum of total responses, IST = sum of internal state terms, ISTr = ratio of IST to total number of words

+ p<.10, * p<.05, ** p<.01

3.3.6. Incremental validity.

On the TEC eight subjects (14.5%) reached ceiling (a score of 9) while on the cTRUE none of them achieved the maximum score of 9. From Table 24, it is obvious that statistically, cTRUE scores are less constrained by ceiling effects. This additional variance however is not necessarily useful or valid. To investigate incremental validity hierarchical regressions were conducted. Only associations of cTRUE components with social competence or participant roles outcomes for which a significant correlation had been found in the previous analyses were subjected to this analysis.

Table 24
Number of subjects reaching ceiling performance in TEC and cTRUE

	TEC	cTRUE
Total score	8 (14.5)	0 (00.0)
recognition	49 (89.1)	27 (49.1)
external causes	46 (83.6)	9 (16.4)
reminder	47 (85.5)	33 (60.0)
desires	45 (81.8)	33 (60.0)
beliefs	47 (58.2)	28 (50.9)
Hiding	40 (72.7)	15 (27.3)
regulation	46 (83.6)	30 (54.5)
mixed	33 (60.0)	17 (30.9)
morality	32 (58.2)	28 (50.9)

Note. n=55. Percentage in parentheses. Ceiling numbers of TEC recognition and external causes scores are based on means, not cut-off scores for better comparability with cTRUE

In the first step the control variables age and vocabulary were entered, in a second step the TEC and finally the cTRUE. To assess incremental validity for cTRUE response time scores they were included in a third step after control variables and cTRUE and TEC scores. Although significantly correlated, cTRUE regulation and mixed emotion accuracy scores did not predict prosocial roles once age, vocabulary and TEC component score was controlled for (see Table 26 and Table 27). CTRUE external causes response times however predicted academic competence beyond age, vocabulary, TEC external causes and cTRUE external causes accuracy (see Table 25). For cTRUE regulation response times a trend emerged to predict prosocial roles beyond age, vocabulary, TEC regulation and cTRUE regulation accuracy (see Table 26).

Table 25
Summary of hierarchical regression analysis for external causes emotion understanding variables predicting academic competence

	Model 1			Model 2			Model 3		
Variable	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
Age	-.10	.07	-.21	-.14	.07	-.29	-.16	.07	$-.34^{*}$
Vocabulary	.50	.35	.22	.34	.35	.15	.00	.36	.00
TEC ext. causes				-1.74	3.94	-.07	-3.36	3.71	-.13
cTRUE ext. causes				19.78	10.37	.30	14.40	9.86	.22
cTRUE ext. causes RT							.00	.00	$-.42^{*}$
R^2		.30			.42			.55	
F for change in R^2		1.95			1.89			6.79^{*}	

Note. N=42. Reg = emotion regulation
* p<.05, ** p<.01

Table 26
Summary of hierarchical regression analysis for emotion regulation emotion understanding variables predicting prosocial roles

	Model 1			Model 2			Model 3			Model 4		
Variable	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
Age	.02	.02	.14	.02	.02	.13	.02	.02	.11	.01	.02	.08
Vocabulary	.20	.10	.31*	.17	.11	.26	.16	.11	.25	.16	.11	.25
TEC regulation				.75	.91	.14	.49	1.12	.09	.60	1.09	.11
cTRUE reg							.66	1.62	.08	.68	1.58	.09
cTRUE reg RT										-.00	.00	-.26+
R^2		.12			.13			.14			.21	
F for ΔR^2		2.60+			0.68			.16			3.09+	

Note. n=42. Reg = emotion regulation
+ p<.10, * p<.05, ** p<.01

Table 27
Summary of hierarchical regression analysis for mixed emotions emotion understanding variables predicting prosocial roles

	Model 1			Model 2			Model 3		
Variable	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
Age	.02	.02	.14	.02	.02	.15	.02	.02	.15
Vocabulary	.20	.10	.31	.13	.10	.21	.14	.10	.22
TEC mixed emotions				1.35	.57	.35	1.50	.76	.39
cTRUE mixed emot.							-.34	1.16	-.06
R^2		.34			.48			.48	
F for change in R^2		2.60+			5.56*			0.09	

Note. N=42
+ p<.10, * p<.05, ** p<.01

3.4. Discussion

Goal of this study was to develop a computerized task of emotion understanding based on the TEC and evaluate its psychometric properties in a primary school age sample. More precisely the goals were to (1) investigate reliability, establish (2) concurrent validity through comparison with the TEC and (3) convergent validity by relating it with other measures of social competence and social cognition, and finally to (4) investigate its value in terms of incremental validity. First the results of these four goals are discussed in turn. Thereafter I highlight findings of interest which are limiting the task's psychometric quality but also emphasize the complexity of emotion understanding not yet captured in its entirety by existing measures.

3.4.1. Reliability.

Reliability for the cTRUE total accuracy score was good. Hayward & Homer (2017) recently investigated the reliability and validity of established theory of mind tasks in a very similar age group (7 to 12 years) of normally developing children. Only the Faux Pas task and the interpretive ambiguous figures task displayed a level of reliability similar to the cTRUE with second-order false belief tasks, interpretive restricted-view tasks, Strange Stories and in particular the Eyes task ranging below. As Cronbach's alpha increases with number of items this comparison might be misleading though. If we look at the internal consistency computed from cTRUE's 9 component scores alone, reliability was still similar to the Strange Stories' alpha with 24 items and far exceeding the Eyes task with 28 items. Compared to these tasks the cTRUE and the TEC assess a wider variety of competencies so a lower internal consistency is to be expected. More troublesome is the heterogeneous internal consistency of individual component scores. In particular morality, beliefs and reminder displayed very low internal consistency. Likely influences pertaining to item characteristics are discussed in chapter 3.4.5.

3.4.2. Convergent validity.

As expected, cTRUE and TEC were strongly correlated. There was also a large association with emotion knowledge. This provides first evidence that the cTRUE measures the construct of emotion understanding similarly to the established Test of Emotion Comprehension. Concerning individual components, correlations were mostly medium or large except for external level components and belief based emotions. For emotion recognition, external causes and reminder, the missing associations are most probably owed to the performance being at ceiling, especially for the TEC. Regarding belief-based emotions see chapter 3.4.5 for a discussion of possible explanations.

3.4.3. Concurrent validity.

There were few associations between cTRUE and the social skills, problem behaviour or social cognition scores. These results were largely backed by missing correlations between the TEC and these measures in this study. There were stronger correlations with prosocial roles. These findings are discussed in turn.

3.4.3.1. Social competence and problem behaviours.

There were small to medium strength associations between understanding of mixed emotions and social skills, as well as understanding of mixed emotions and emotion regulation and academic competence but they only reached trend level. For problem behaviours there was no significant correlation with any cTRUE component. This is in line with a study that found TEC performance to be largely unrelated to internalizing and externalizing problems in primary school age children (Gobel, Henning, Moller, & Aschersleben, 2016). Concerning speed of responses, the medium to large negative association between cTRUE external cause's response times and SSIS academic competence score stood out among the otherwise small to medium correlations of cTRUE scores. Emotion knowledge, a construct largely overlapping with understanding of external causes of emotions, in Kindergarten has been shown to predict academic performance in primary school (Izard et al., 2001; Trentacosta & Izard, 2007). More unexpectedly, response times for the reminder component showed a positive correlation with social skills and a negative correlation with problem behaviour. That is, children that took longer to respond correctly to prompts about which emotion a reminder elicits were rated higher on social skills and lower on problem behaviours by their teachers and guardians. This association is fairly cohesive and cannot be attributed to outliers (see Figure E 1 and Figure E 2 in Appendix E). Could this relationship be an artefact of the data fed into the variable? Only response times for correct responses were counted because information about response speed in incorrect responses is uninformative or at best equivocal. Maybe subjects with a lower accuracy score in the reminder component also took longer to give their correct responses? This was not the case since reminder accuracy was not related to the reminder mean response time (r=.09, p=.54). The robust associations with social skills and problem behaviours was mainly with the first item that is a slightly modified version of the original TEC item. This vignette tells the story about a child's cat (the cat from one belief item) that was chased away by a dog and did not return. Later the child looks at a photograph of the pet and the participant is asked how the child feels. Let us look at another reminder item for comparison. A child is frightened by an aggressive looking dog jumping up the kid. Later the child sits at a bus station and a woman holding the dog (looking neutral now) on a leash comes by. The structure of this item is similar to the original item but does not show these correlations. There is a critical difference inherent in the test procedure however. The item with the cat is the only one referring to a previous item in the test procedure from another component (belief based emotions: the cat not knowing that a dog is hiding behind the bush). May this slower response be caused by remembering and processing the earlier vignette and be indicative of a good comprehension of emotions or mind-reading motivation? Looking at correlations between the item response time and cTRUE (r=.02, p=.89) and TEC (r=−.17, p=.25) total scores and the accuracy of the corresponding belief item (r=.24, p=.11) this does not seem to be the case. As it is this finding turns out hard to explain and requires further studies of emotion understanding utilizing response times.

There were no relations with the self-reported social activities and peer contact (the social competence score of the youth self-report) except for a non-significant low to medium correlation with reminder and a significant negative correlation with morality. Interestingly reminder was the only TEC component to be negatively correlated with social withdrawal in the study by Gobel et al. (2016). May the unexpected negative correlation with moral emotion understanding be a consequence of cTRUE item idiosyncrasies? Correlation with the TEC morality score was non-significant but pointed in the same direction (r_s=−.21, p=.13). Looking at individual cTRUE morality items the negative relationship was visible with the recreated TEC-item (r_s=−.31, p=.01) and the third morality item about coming home too late (r_s=−.29, p=.03). Thus this finding is not attributable to item characteristics of the newly created vignettes. Discerning the relationship further in respect to the YSR social competence sub-scores, morality was only significantly negatively related to frequency of meeting friends (r_s=−.29, p<.01) and club activity (r_s=−.27, p=.05). Next, might idiosyncrasies in the sample have caused this puzzling correlation so that a sub-sample exhibited a lower/higher score in these two variables paired with a higher/lower morality emotion understanding score? Looking at group differences between recruiting sites (primary schools, day care centre) with Kruskal-Wallis-Tests there were no significant differences for the variables frequency of friends met ($\chi^2(2)$ = 3.54, p=.17) or club activities ($\chi^2(2)$ = 5.36, p=.07). At the same time the mean cTRUE morality mean ranks between groups were virtually the same, speaking against this hypothesis ($\chi^2(2)$ =0.93, p=.63). One has to bear in mind however that reported frequency of social activities and peer contact was only modestly and non-significantly related to teacher reported social skills (r=.22, p=.16). Insofar the relevance of this result remains unclear and open to further investigation.

3.4.3.2. Flexibility and automaticity of social cognition.

The results pertaining to the relationship between cTRUE and FASC indicate that both tasks capture largely different processes. Children who were more adept at understanding emotions did not produce a larger variety of mental justifications or mental state terms in response to social vignettes. This is in line with findings of a recent study comparing the TEC and a similar interview procedure like the FASC where children were asked about the feelings of characters in social vignettes (Castro, Halberstadt, & Garrett-Peters, 2016). Performance in the TEC was not related to the number of different affective mental state terms used.

The trend level negative correlation between cTRUE response time and FASC's internal state term ratio warrants further attention however. It seems that children whose responses in the FASC were more densely packed with internal state terms also tended to be quicker to respond to the cTRUE prompts. This could be due to a certain response behaviour (impulsive vs. reflective) but the absent correlation with initial response time in the FASC makes this unlikely. Another explanation is that of a genuine difference in mentalizing efficiency. This possibility should be tested in further studies.

3.4.3.3. Social role behaviour.

Medium strength correlations between cTRUE total accuracy score, recognition, external causes, regulation and mixed emotions component scores and prosocial roles emerged in line with previous findings with the TEC in preschool children (Belacchi & Farina, 2010). There was a negative association between mixed emotions and the outsider role not found in the study by Belacchi & Farina (2010). Concerning response times, faster responses to emotion regulation and external-causes vignettes were associated with prosocial roles while slower responses to reminder items were related to lower scores in the victim role.

Unexpectedly, the component of desire based emotion understanding was negatively related to the prosocial roles of consoler and mediator and positively related to pro-bully roles. Looking at correlations with individual cTRUE-desires item scores (see Table E 5 in Appendix E), the remade TEC item (*salad*) was negatively correlated with prosocial roles while the new cTRUE item (football) was positively related to pro-bully roles. Associations with the TEC single item scores and combined desire score were also generally in the same direction albeit smaller and mostly non-significant. This suggests that the link found cannot be sufficiently explained by idiosyncrasies of the cTRUE items. To investigate if age or vocabulary mediated this relationship, partial ordinal correlations were computed. Correlations for consoler and mediator remained significant speaking against cohort effects associated with age or verbal intelligence. Concerning pro-bullying roles, correcting for verbal intelligence did not influence the result but partialing out age did lower the association below the threshold of significant although the effect remained at almost medium size.

Highly developed mentalizing abilities have mainly been linked to prosocial behaviour (see chapters 2.1.1.5 and 3.1.7.2) while anti-social behaviour like bullying has been associated with social cognitive deficits (Dodge, Pettit, McClaskey, & Brown, 1986; Crick & Dodge, 1994). On the other hand the Machiavellian intelligence hypothesis posits, that behaviours such as deception and manipulation are adaptive and have been evolutionary shaped through selective pressure in social groups (Byrne & Whiten, 1988). Similarly it has been argued not to take a deficit but individual difference perspective on bullying with at least ring-leader bullies possessing adaptive social skills irrespective to the ends they are utilized (Sutton, Smith, & Swettenham, 1999b; also see the comment of Crick & Dodge, 1999 and response of Sutton, Smith, & Swettenham, 1999c). Bullies also seem to be a heterogeneous group with some scoring sub-par and others scoring normal on theory of mind tasks (Gasser & Keller, 2009). In a noteworthy prospective longitudinal study about environmental risk factors on development, a large sample at age 5 was assessed with a range of first and second order theory of mind measures and followed up at ages 7, 10 and 12 (Shakoor et al., 2012). Poor theory of mind in childhood increased the risk of being a victim or bully-victim in early adolescence over and above child- and family specific factors. There was a similar association for becoming a bully although it was statistically explained by low socio-economic status and child maltreatment. However, the negative relationship between desire based emotion understanding and pro-social roles was an isolated finding and other correlations with emotion-understanding were in the positive direction so this finding has to be replicated before drawing further conclusions. In summary, associations between cTRUE and teacher reported social role behaviour were largely in line with previous findings.

3.4.4. Incremental validity.

There were interesting medium-strength correlations between cTRUE regulation and mixed emotions accuracy and response time scores with teacher rated academic competence not visible with the TEC. They only reached trend level however, arguably due to the limited sample size, so their relevance is questionable until replicated. Overall, cTRUE accuracy scores do not seem to add incremental validity to the TEC in respect to relations with measures of social competence. This does not preclude however the possibility that they hold meaningful variance in respect to other aspects of socio-emotional competence, not considered here. In contrast, response times of cTRUE external causes component did explain variance in academic competence beyond TEC and cTRUE accuracy scores. Could another common factor better explain this link? Children who are better at controlling a computer mouse might excel at school because their manual dexterity indicates advanced maturation or because their familiarity with computers might arise from belonging to a household with a higher socio-economic status. Or else speed of processing would be arguably beneficial for academic competence, as well as responding swiftly on cTRUE. However, if that was the case, we would expect all cTRUE component response times to show this link with academic competence. A link between external aspects of emotion understanding and academic competence has already been found before albeit for accuracy (Izard, 2001). The question, why the link here has been found only for speed is open to further investigation. For cTRUE emotion regulation response times the variance to predict pro-social roles beyond TEC and cTRUE accuracy scores reached trend level. Remember that only response times of correct responses entered the response time variable. This suggests that there is a difference between children who appraise cognitive emotion regulation strategies as most effective faster or slower in terms of prosocial behaviour. Taken together response time latencies do seem to contain useful information in tasks like the TEC and should be used more routinely.

3.4.5. Heterogeneity in emotion understanding components reminder, beliefs and morality.

The three cTRUE components reminder, beliefs and to some lesser degree morality displayed exceptionally low internal consistency. What causes and properties could this fact be attributed to?

Two of the three reminder-vignettes showed low item-scale correlation, all three did not correlate with each other and only the recreated TEC item correlated with the original TEC item. This is surprising since all items follow the same rationale: (1) character in an emotion eliciting situation with the emotion being explicitly stated, (2) character goes to sleep to indicate passage of time, (3) character in a situation eliciting a different emotion which is quizzed by a control question, (4) description of a reminder for the original (target) emotion eliciting situation (e.g. aggressor, lost pet) followed up by the test question about the current emotion. The three items do however differ regarding the baseline and target emotions (see above). In the recreated original TEC item it is happy-sad, in the two new items it is ok-scared and ok-angry. Contrasting a positive vs. negative emotion might require a different kind of understanding than contrasting a neutral with a negative emotion. Yet the two items featuring a neutral affective stance do not correlate either. Also in terms of difficulty the happy-sad and ok-scared items are more similar than the ok-angry item which is most difficult. So the kind of initial-final emotion contrast seems not to be accountable. Regarding low item-scale-correlation, might the accuracy in response to the control question have something to do with it? For the vignette similar to the item in the TEC (a lost pet) only about half of the subjects answered the control question correctly. Looking only at subjects who answered the control question correctly, item-scale correlation was only slightly higher speaking against this explanation. It is also possible that because of the relatively low item difficulty for two of the three items and the associated low variance, the lack of association might be a spurious result (Goodwin & Leech, 2006). A replication with a younger, larger, and more homogeneous sample could shed light on this issue.

Turning to the component of belief-based emotions, the low item-difficulty for all but the recreated TEC item is notable. A slight change in the wording of control questions might have inadvertently caused this gap. In three of four new items, the initial scene description and/or the control questions explicates the thoughts of the character. Consider following example:

> This is Alex. Today is his birthday but when he comes home nobody is there to congratulate him. Please click on the door to see what's behind. In the other room Alex's mother and his friends are hiding to surprise him. Please click on the door again to get back to Alex. [Control question:] Does Alex know that there is a surprise party for him? [Control question answered correctly:] That's right, Alex does not know that there is a surprise party for him. He thinks everyone forgot about his birthday. (Haslinger & Leyrer, 2013)

On the other hand, the item "empty fridge" does not reference the character's thoughts and yet it has an item difficulty similar to the other three new vignettes. Another possible explanation can be discussed within the context of low item inter-correlations. Only the cat-item of the cTRUE belief component related to its counterpart (featuring a rabbit) in the TEC. Furthermore only the "laugh-down" item (emotion elicited through false belief: anger, emotion that a true belief would elicit: happy) correlated with the other new cTRUE items, but not with the recreated or original TEC item. The item difficulties of four of the five items were so low that even more so than in the case of the results of reminder, a few spurious results might have distorted the correlations. Indeed, in the case of the sadness-vignette (i.e. surprise party), the only subject failing the item showed a relatively high total emotion understanding score resulting in a negative item-scale correlation. However there are also more fundamental possible explanations. For one thing, valence effects may play a role. The cat-item differs conceptually insofar as a negative-valence reality (fear of the dog) needs to be overridden. In the other items, a positive-valence reality (surprise party, mother brings pizza, etc.) needs to be overridden. A salience effect for attributing negative emotions has been found in the case of external causes and real vs. apparent emotions (Fabes et al., 1991; Harris et al., 1986). In the case of emotive false belief reasoning, at least for typically developing children, there does not seem to be a difference between valence of immediate and eventual emotions (Hughes et al., 1998; Parker, MacDonald, & Miller, 2007; Seidenfeld, Ackerman, & Izard; 2017) speaking against valence effects. These studies, in contrast to the present, have used strictly narrative vignettes however, so valence effects based on procedural differences cannot be ruled out. For example, supporting the narrative prompt with a picture might prime external emotion reasoning and thus, if external emotion attribution (e.g. empty room) and false belief based emotion (e.g. everyone forgot about my birthday) coincide, result in a higher accuracy. A design in which the emotion associated with the external context contrasts with the false belief based emotion attribution (e.g. a birthday character standing in a room full of party guests, falsely believing they have to leave early) might better capture pure belief based emotion understanding. Systematically varying these factors could shed light on this possible interaction. A procedural variant between the items may also have impacted performance. In the original TEC item a wolf is hiding behind a bush (reality) behind a rabbit that is eating a carrot unsuspecting of the predator (false belief). In the test-book, this scene is realized on a single page with the bush being a lid that is opened by the child to be informed of presence of the wolf. Similarly in the corresponding cTRUE item, a dog is hiding behind a bush behind a cat, enjoying the sun. In the other cTRUE items, the reality disconfirming the false belief is presented on another page (e.g. surprise party behind the door). The presence of the bush, hiding the predator in the field of view might increase the salience or awareness of the danger to the rabbit (potential cause of fear) in the child's mind, thus increasing the false-belief vs. reality conflict. Indeed this item was markedly more difficult (*M*=.65) than the other belief-based emotion items (*M*=.91). Interestingly this unexpected finding might hint at the role of executive functions in theory of mind. If performance on this task was only dependent on conceptual understanding an increase in difficulty was not to be expected (note that the comprehension of the false belief control question was near perfect, *M*=.97). If however correctly judging the false belief dependent emotion requires inhibition of the child's knowledge of the real state of affairs, greater salience of the conflict should increase difficulty. This might also explain the low item difficulty of the new vignettes. Unfortunately we did not include a measure of inhibitory control to answer this question empirically. Again a replication with a younger sample needs to affirm this lack of association before concluding that these items are conceptually different.

Inter-correlations between morality items were also very low and two vignettes displayed near zero item-scale correlations. There was no procedural difference between the items as in the belief-based emotions component that could explain this finding. Comprehension was very good (*M*=.96) and performance in the test question was not at ceiling. Might the different scenes depicted themselves be the origin of this heterogeneity? Let me first recollect the three scenarios: In the first scenario (cookies) the character eats a cookie without asking and later in the evening tells her mother about her day but not about having taken the cookie. In the second scenario (homework), the character lies to her mother about the homework so she can go out and play. In the evening the character reminisces the day and remembers having lied to her mother. In the third scenario (home late), the character decides to stay at a friend longer than agreed upon with her mother. She then comes home late and her mother is already waiting at the door (importantly scared is not a response option in this scenario). What are the communalities? All three scenarios include a moral transgression: not asking before taking something belonging to someone else, lying to someone dear and not honouring an agreement. There are also important differences that might have had an effect. The first two items are both about being dishonest to the mother although in the first item indirectly and the second openly. Conversely in the first item remembering the dishonesty occurs while being with the mother, in the second while being alone. In the third item there are not really mental states (deception) involved. While the cookie-item unsurprisingly correlates strongly with its TEC-pendant but not with the cTRUE total score, the homework-item correlates with both. The third item eventually isn't associated with either. It seems attributing emotional consequences to moral transgressions is not a unitary process and may depend on contextual factors and probably personality variables. The cause of this heterogeneity could be elucidated in a future study that systematically varies moral transgressions with vs. without dishonesty, direct vs. indirect dishonesty and presence of attachment-figures during remembrance of the moral transgression.

3.4.6. Influence of emotion categories.

Might heterogeneity in emotion understanding components be attributed in part to differences in response characteristics between emotion categories? To answer this question I looked at item difficulty and internal consistency within emotion categories, not emotion understanding components. Because of item characteristics not every component could be included in this comparison. For desire based emotions and morality there was only one correct response for each item (happy/sad and sad, respectively). For mixed emotions there are two possible responses per item and for emotion regulation the responses are based on behaviour not emotions. Emotion recognition was also excluded from this comparison despite containing multiple emotion response categories because of the qualitatively different requirements (no understanding based on narrated story vignettes but only facial visual features). The emotion category *happy* did not appear in reminder, belief-based emotions and hiding emotions so it was omitted in the comparison. For scared, sad and angry item difficulties and Cronbach's alpha was computed for vignettes of components external emotions, reminders, belief based emotions and hiding emotions. Mean item difficulty for scared, sad and angry was *p*=.88, *p*=.88 and *p*=.75 respectively. Mean internal consistency within emotion categories for above mentioned components was α=.59, α=.29 and α=.59 for scared, sad and angry items respectively. Two things can be gleaned from these additional analyses.

First, angry items have a slightly higher item difficulty. This is in line with previous findings of causal emotion attributions in preschool children. In a study with 4- to 5-year-old children, blinded expert-raters were more inaccurate in matching children's causal explanations with the corresponding emotion for anger than for fear but equally well for sadness (Russell, 1990). In a more recent study 3- and 4-year-old children had more difficulties producing justifications for why someone might feel angry that expert-raters could blindly match to the corresponding emotion than for the emotions happiness, sadness or fear (Russell & Widen, 2002). Likewise, Widen and Russel (2013) found that pre-schoolers could more easily identify a story-character's emotions from situational causes (e.g. birthday) than from their behavioural consequences (e.g. jumping up and down and clapping hands) for happiness, sadness and fear while the reverse was true for attributing anger. Thus there is evidence that causal explanations for anger are more difficult to produce for children or seen from another angle, anger attributions might be more dependent on individual response styles than general situational or behavioural attributions. Finally, potentially anger eliciting situations might inherently be more ambiguous and requiring additional information about beliefs, desires and mental states in comparison to - let's say - happiness eliciting situations. An example from the latter study illustrating this ambiguity would be a mom yelling at her son which can plausibly be attributed as causing anger, sadness or fear.

Second, why is internal consistency so much lower for sadness-vignettes? For external-causes items, internal consistency was roughly equal to the other emotion categories. Looking at individual item statistics, the hiding-item for fear displayed a substantially higher item-scale correlation than the sadness-item in respect to the emotion-category as well as total emotion understanding score despite having almost identical item difficulty. For the sadness item (i.e. child that fell off his/her skateboard laughing to not show how he/she is feeling really inside) the response options were happy, scared, angry and sad. As argued above sadness- and anger-attributions are more ambiguous than for other emotions (also see Borke, 1971). For the reminder item (which was the modified TEC reminder item), there was no competing anger response option to choose from. For belief-based emotions the low item-scale correlation for one sadness item is attributable to a spurious result as already mentioned in the previous chapter. The other sadness item however was not at ceiling concerning item difficulty and had an item-scale correlation similar to the other emotions. Taken together this suggests that the lower internal consistency of sadness-vignettes is not attributable to a consistent emotion category effect but rather to individual item idiosyncrasies.

3.4.7. Problems, possible solutions and open questions.

First, problems of the existing cTRUE task are addressed and general solutions discussed before turning to solutions for individual components of emotion understanding.

As far as it concerns psychometric standards of individual assessment are concerned, internal consistency and item inter-correlations at the level of components are insufficient for all but emotion regulation and mixed emotions although in the realm of experimental research tasks with the goal to examine relations or differences on a group level, such numbers are not unusual (e.g. the Eyes task, see Hayward & Homer, 2017). Nevertheless, the goal has to be to increase internal consistency and provide re-test reliability in future studies. To achieve this, item characteristics such as target emotion, distractors, accompanying vignette descriptions or included social norms should be manipulated systematically and compared to explain the heterogeneity for some items.

Most items of the cTRUE are also too easy for the intended age group of primary school children although they may be perfectly suited for children with emotional or developmental disorders. Applicable to most components of emotion understanding, a higher difficulty could be achieved through the inclusion of social emotions (e.g. guilt, embarrassment, envy, jealousy, pride, etc.), either in established components (e.g. reminder, beliefs) or as new components (e.g. Quintanilla & Gimenez-Dasi, 2017). A problem for the normative answer scheme of the cTRUE might be, that social emotions are in general even more difficult to map to specific situations as they are more dependent on the understanding of other mental states and social knowledge such as scripts and norms than so-called basic emotions. A similar problem has been found in respect of the Meyer-Salovey-Caruso Emotional Intelligence Test where for several items the correct answer, as defined by an expert panel in the development of the task, was not the answer associated with the highest ability of emotional intelligence (Fiori et al., 2014). To reduce ambiguity, it might be necessary to convey additional information about characters such as the kind of relationship between characters, expectations or social class. This in turn might make the task more culture-dependent. This assumption of course would have to be tested empirically. It may as well become apparent that in complex situations there is not one correctly associable emotion. Of course a statistical norm could be collected which tells us what emotion a majority of people regard as most fitting for a specific scenario (a similar approach has been chosen in the creation of the training stimuli in the third study). Another fruitful avenue might lie in asking subjects both how they would feel and what they think most other people would feel in a given situation. Apart from an ability stance on emotion understanding, response patterns can be associated with clinical symptoms or social behaviour to investigate emotional biases as was done with the ACES (Schultz et. al, 2004).

To make sure performance in higher components of emotion understanding was maximally independent from demands of facial emotion recognition, the four response vignettes were also acoustically labelled (e.g. „does he feel happy, sad, angry or alright") and subjects were required to response after this prompt. Therefore participants who knew the answer right after the vignette at the start of the prompt could still give the response only after the end of the prompt making it impossible to distinguish them from participants who only figured out the correct response at the end of the prompt. This had a similar effect to windsorizing early responses. A follow-up version of cTRUE could require subjects to respond as fast as possible, even before the vignette or prompt has ended. Even more interesting would be to track eye-movements. An alternative time-variable could be created by considering the first fixation with a given minimum length of the chosen emotion as a possible estimator of speed of emotion understanding. Let us now turn to a discussion of problems and possible solutions for each component of emotion understanding.

Recognition: Facial expressions could be varied in intensity to adjust difficulty. However, most available pictures sets used in research feature mostly exaggerated facial expressions (e.g. Ekman & Friesen, 1976; Lundqvist, Flykt, & Öhman, 1998; Tottenham et al., 2009). Apart from problems such as ecological validity and fundamental questions about natural kinds of emotions (e.g. Barrett, 2006), this leads to very easy instances of emotion recognition. A paradigm similar to the Emotion Recognition Task (Montagne, Kessels, De Haan, & Perrett, 2007) or the Emotion Hexagon Test (Young, Perrett, Calder, Sprengelmeyer, & Ekman, 2002) where graded intensities of facial expressions are presented could be adopted. Additionally emotion recognition through the voice or other nonverbal bodily cues (posture, gestures) could be included.

External Causes: The social situations items from the emotion knowledge task were more difficult (.75) than external causes in the cTRUE (.87) despite also describing situations that are in the range of a child's every-day experiences. Presumably this might be due to two procedural differences. First, children were only read the vignette and did not see a corresponding picture. Second, in the emotion knowledge task, there were five instead of four response options. Excluding visual stimuli from the cTRUE would presumably make the task more dependent on verbal abilities. Indeed, while the external causes component in the cTRUE correlated only moderately with the vocabulary test (r=.24, p=.07) the social situations score of the emotion knowledge task displayed a highly significant association (r=.48, p<.001). Conceptually, including social emotions (e.g. envy, shame) could make the task more difficult for children in middle childhood.

Reminder: This component could be broadened to include how children integrate mental states over time. Changing the response format to include a continuous, or at least ordinal judgement of the strength of emotion could pave the way for a more fine grained analysis of emotion reminding cues and emotion understanding in general (e.g. other components in the TEC/cTRUE). For the specific component of reminder, the perceptive or symbolic similarity of reminding cues to the original emotion eliciting object and its influence on the attributed reinstated emotion could be taken into account. Most 5-year-olds (repeatedly) but only few 3-year-olds (and only sporadically) explain emotions through cognitive cues if they are only symbolically related and not exact parts of past events (Lagattuta, Wellman, & Flavell, 1997). Another variable is how consistent emotional experiences are linked with an object or contextual cue and using that information to predict future emotions, thoughts and decisions. It has been shown that even 4- to 5-year-old children make more positive predictions for unambiguously positive past experiences followed by ambiguously and finally unambiguously negative past experiences (Lagattuta & Sayfan, 2013). Interestingly the most recent past event was weighted more heavily, particularly if negative, which is interesting from a temporal discounting perspective. The differentiation of this sequence became more pronounced with age suggesting the development of judgemental heuristics.

Desires: This component could be expanded to include social desires (e.g. desire to impress/hurt/etc. someone), not only material desires (e.g. desire to obtain a toy). In a more general approach, psychological needs and goal theory could be systematically incorporated (e.g. Heckhausen & Kuhl, 1985; Epstein, 1990; Grawe, 2004). In this vein emotions are responses to (un-)successful approach or attainment of goals with motivational goals being grounded in basic psychological needs (e.g. attachment, control/orientation, self enhancement, pleasure-gain/pain avoidance).

Beliefs: Emotion understanding depending on higher order theory of mind would increase the difficulty which is the case with social emotions which require beliefs about mental states of other beings. To give an example: being afraid of hamsters and (wrongly) believing my new classmate I am visiting has one is a first-order belief-based emotion.

Feeling ashamed (a social emotion) because knowing that my peers know that I am afraid of hamsters (and find it ridiculous) is a second-order belief-based emotion.

Hiding: additional to assessing if children are aware of the possibility to supress the display of emotions at all, the sensitivity to detect such dissimulation tendencies could be explored. In the TEC (and cTRUE), the child is explicitly prompted that the character in the vignette displays a certain emotion because he doesn't want to know the others to know how he is actually feeling. Spontaneous detection of display rules might be investigated. Alternatively the likelihood and difficulty to hide emotions across different external and internal factors could be explored (e.g. personality; the duration to uphold the "forged" emotion; varying difficulty in the face of repeated cues that elicit the real emotion).

Regulation: A greater variety of emotion regulation strategies could be included in the multiple choice options. Following Grob and Smolenski (2005) adaptive strategies could include: problem-focused action, distraction, enhancing mood, acceptance, forgetting, reappraisal and cognitive problem solving. Maladaptive strategies could include: giving up, aggression, withdrawal, self-devaluation and perseveration.

Mixed: In the cTRUE and TEC the different thoughts/attributions leading to the mixed emotions are explicitly stated to probe if conceptual understanding of intrapersonal interpretive diversity is there at all, which establishes a lower boundary of competency. Spontaneous attribution of mixed emotions (given ambiguous past experiences/associations) would likely increase inter-individual variance but might prove to tap more into personality variables or thinking styles than an ability.

Morality: certain social emotions like shame or guilt are responses to social or moral transgressions. To focus on the moral quality of these emotions, the moral transgression should not be followed by repercussions. In some sense the TEC item (and the cTRUE item homework) already incorporates the social emotion guilt but labels it as sad instead.

3.4.8. Summary.

The cTRUE was devised a computerized modification of the TEC for non-dichotomous assessment of emotion understanding components in middle childhood. cTRUE and TEC total scores were strongly correlated, even when excluding non-shared vignettes suggesting that the new computerized task captures emotion understanding similarly to the established pen and paper task. On a component level, reflective components of emotion understanding showed strong associations with respective TEC components but emotion recognition and external causes were unrelated most likely due to low variance and belief-based emotions vignettes showed great heterogeneity. Internal consistency of the cTRUE total score was good but reliability for single components varied greatly. Particularly the components of reminder, beliefs and morality displayed item heterogeneity hinting at complex factors influencing emotion attribution. There were few associations with the external criteria of social skills and mental state reasoning but stronger links to social role behaviour, consistent with the results for the TEC. In terms of incremental validity cTRUE accuracy scores failed to significantly add information in respect to the external criteria but cTRUE external causes response times predicted academic competence beyond control variables, the TEC and cTRUE external causes accuracy scores. The present results suggest that the more complex social vignettes are, the harder one-to-one emotion mappings become and the more individual affective response styles are probably captured beyond emotion understanding. Furthermore the study provides tender evidence for the utility of response time measurement in conventional theory of mind and emotion understanding tasks.

4. Study 3: EmoJump. A computer game designed to promote emotion understanding.

4.1. Introduction

4.1.1. Improving emotion understanding.

Despite the breadth of knowledge about the role of socio-emotional competence in psychopathology and behaviour problems, only recently have studies to improve emotion understanding emerged in greater number. An up-to-date meta-analysis (Sprung, Münch, Harris, Ebesutani, & Hofmann, 2015) looked at 19 training studies and found robust effects for external, mental and reflective aspects of emotion understanding. Aside from integrating effects of different studies, thus increasing the number of subjects and power, meta-analyses provide the opportunity to look at moderators of intervention effects. Interestingly for external aspects the effect was larger in studies using a group protocol while for reflective aspects the effect was larger for individual settings and for longer training sessions. Since reflective aspects of emotion understanding are more complex than external aspects, it is possible that longer training sessions in an individual setting provide more opportunity and focus for improving them. It should be noted, that the effect size of external and reflective aspects was double the size than for mental aspects of emotion understanding. Even more remarkable, there was no significant moderating effect of training composition (combination of external, mental and reflective aspects trained). However the number of studies in the meta-analysis was rather small for moderator analyses so the moderating effects have to be interpreted with caution. Noticeably, most interventions have targeted clinical populations, particularly people with the autism spectrum disorder, which is not surprising, since the condition has been firmly linked with deficits in theory of mind. Still, the meta-analysis has not found a significant moderating effect of population, only a trend, which suggests that emotion understanding can even be enhanced in normally developing children.

4.1.2. Conventional emotion understanding interventions.

Some interventions are realized as school- or kindergarten-curriculums. The *Funny Faces* program for example, is designed to promote recognizing and understanding of simple and complex emotions, understanding of situations as elicitors of emotions and how emotions affect behaviour. In a study with deaf children, improvements in understanding situational causes of emotions and emotion vocabulary were observed (Dyck & Denver, 2003). In another study (Pons, Harris, & Doudin, 2002) 9-year-olds underwent the *School Matters In Lifeskills Education* (SMILE) program in classroom for half an hour a day for three months. The SMILE teaches understanding of emotions in oneself and others, including past, future and present emotions and expressed vs. felt emotions. The intervention group improved in the Test of Emotion Comprehension (TEC) total score compared to the control group. Psychological adjustment and mental health was the target of a quasi-experimental study using an emotional intelligence training program (INTEMO) based on the Mayer-Salovey four-branch model of emotional intelligence (Ruiz-Aranda et al., 2012). In 24 weekly one-hour training sessions in class, divided between two 6-month periods over two years, adolescents worked on emotional intelligence skills (emotion perception, emotional facilitation of thinking, understanding of emotions, regulation of emotions) through a variety of tasks like role-playing, reflective activities, art or group work. Psychological maladjustment decreased compared to the control group at post-test. At follow up, half a year after end of training, the effect on maladjustment had increased and positive effects concerning negative affect and mental health emerged. The effect sizes were small or small to medium. *The Roots of Empathy (ROE)* (Gordon, 2001) is a curriculum to promote socio-emotional competence for children in kindergarten through 8th grade and has been implemented in communities of several countries. The curriculum stretches over 9 months and comprises 26 sessions across nine themes. At the start of the program each class "adopts" an infant. The infant and its parents visit the class each month. This serves as a starting point for discussions and activities that are implemented in the general education plan and fostered during visits of a ROE-coordinator three times a month. Children learn about infant development, parental practices, perspective taking and empathy. They for example read a book and discuss the mixed emotions of a character that is experiencing toothing. In the following infant visit the children ask the parents about their feelings about their toothing baby. During the whole program particular importance is placed on a caring and understanding classroom ecology. In a large evaluation of the program (Schonert-Reichl, Smith, Zaidman-Zait, & Hertzman, 2012) with over 500 children assigned to treatment or control group, improvements in peer-rated prosocial behaviour (approximate mean effect size of Cohen's-d = 0.5) as well as teacher reports of aggression (approximate mean effect size of Cohen's-d = 0.4) were observed. Additionally children produced more emotion focused explanations for causes of infant crying (Cohen's-d = 0.25), they did not however improve in self-reported empathy or perspective taking. An overview of many other school-based curricula for social emotional learning from pre- to high-school can be found on the resourceful website of the Collaborative for Academic, Social, and Emotional Learning (CASEL, casel.org).

Quite a lot of studies have targeted children with autism spectrum disorder. An early training study investigated the possibility of teaching emotion-, belief- or pretence-understanding to children with autism (Hadwin, Baron-Cohen, Howling, & Hill, 1996). Emotion understanding training and assessment included facial emotion recognition, external causes, desire-based and belief-based emotions. The emotion understanding training group improved in the taught material as well in untaught material (near transfer) but not in belief understanding or pretence play. In another study, 8- to 12-year-old boys with autism spectrum disorder underwent a clinic-based group-curriculum consisting of 20 weekly 1.5 hour sessions divided upon two modules separated by a break (Solomon, Goodlin-Jones, & Anders, 2004). The program focused on emotion recognition in oneself and others, theory of mind (e.g. perspective taking), executive functions (especially individual and group problem solving) and conversation skills through activities like modelling, role playing, games, visual templates or planning of a party. Concurrently, parents themselves participated in psycho-educative sessions explaining the curriculum's lessons, how temperament, cognition and disorder interact and produce their child's problem behaviours and discussing weekly problem behaviour logs and developing solutions. The intervention group improved in facial affect recognition compared to the control group. A manualized theory of mind group treatment program (Gevers, Clifford, Mager, & Boer, 2006) that involves training sessions about precursors of theory of mind (e.g. listening, recognizing intentions and desires), first order mental state reasoning (e.g. deception, humour) and second order mental state reasoning was used in a randomized controlled trial with children and adolescents with a diagnosis of autism spectrum disorder. Overall the experimental group did not improve on emotion knowledge but they did more often include mixed and complex emotions in their responses (Begeer et al., 2011). Another approach are multi-modal protocols, based on an ecological systems model (Bronfenbrenner, 1979) teaching and practicing social concepts (e.g. what a group is), affective education (e.g. verbal and nonverbal emotional cues, emotional display rules, etc.) and interpersonal problem solving. Such a protocol has been used in three studies with children with the autism spectrum disorder aged 7 to 11 years (Bauminger, 2007a; Bauminger, 2007b) and mean age 11 years (Bauminger, 2002). These studies suggest that the training protocol leads to improvements in recognizing emotions and emotion knowledge (generating examples for emotions). The results were more robust for complex (e.g. embarrassment, jealousy) than simple emotions.

Schizophrenic spectrum disorders are a second area of focus in psychiatry for interventions in social cognition and emotion understanding. The *Training of Affect*

Recognition (TAR) is a manualized computer-assisted intervention over 12 sessions of about 45-60 minutes and is implemented in a setting of two patients and one therapist. It is described in this section since the computer program is more of a tool used by the therapist than a standalone training. The focus is on recognizing facial affect but emotion understanding in social scenes, including complex and mixed emotions, is also covered. Training techniques are based on restitution (i.e. repeated training) and compensation (alternative strategies, e.g. verbalization, using contextual clues and nonverbal signs). In a randomized controlled study (Wolwer & Frommann, 2011) with patients with schizophrenia and schizoaffective disorder, one group received the TAR twice a week over 6 weeks, while the other group completed a cognitive remediation training, similar in respect to structure and methodology (e.g. also incorporating compensation strategies). The TAR-Group improved in facial affect recognition, prosodic affect recognition and theory of mind compared to the cognitive remediation group.

Some interventions target a broad range of socio-emotional and behavioural competences, including some form of emotion understanding training. The nature of such protocols makes it hard or even impossible to disentangle the relative effect of the emotion understanding module. One such intervention program is *Strengthening Early Emotional Development (SEED)* which specifically targets anxiety. It consists of 10 weekly group sessions separately for parents and children and incorporates procedures from different evidence based group programs (ParentCorps, Preschool PATHS, Cool Kids Program, and Coping Koala). Parent sessions primarily focus on psycho-education (e.g. anxiety, exposure) and fostering parenting-skills (e.g. child play, dealing with feelings, managing child anxiety). Child sessions include emotion recognition, emotion talk, relaxation techniques and social skills. SEED has been shown to improve emotion recognition (Fox et al., 2012). Two studies of the same research group, using such broad range interventions, targeted schizophrenic patients. In the first study (Horan et al., 2009) the social cognition training consisted of two phases of six sessions each. While the second phase primarily focused on teaching and training of social cognitive strategies (e.g. to avoid jumping to conclusions) and integration of social cues (e.g. for evaluating non-literal speech and deception), the first phase concentrated on recognizing emotions through facial expressions, non-verbal cues (e.g. posture, tone, etc.) and social contexts. Pictures, audio recordings and video clips were used as training materials. The social cognition training group showed significant improvements with a large effect size in facial affect perception compared to the control group receiving relapse prevention and illness self-management education. There were no significant improvements in measures of social cognition (e.g. understanding non-verbal cues, hostile attributional style, and higher-order theory of mind or mentalizing) over the control group. In the second study (Horan et al., 2011) with subjects with psychotic disorders, they compared four different treatments: social cognitive skills training, computerized cognitive training, a combination of both and illness self-management. The social cognitive skills training group improved in facial emotion recognition and emotion regulation compared to the other groups.

Other studies emphasize the role of mental state talk in the development and promotion of emotion understanding. They read preschool children stories enriched with mental state language. The experimental groups engaged in language games afterwards while the control groups were allowed to play freely. The experimental groups improved in emotion understanding as measured with the TEC which does not require expressive language and relies on nonverbal responses. Results also suggest that the positive effect of mental state language games might be higher for very young preschool children since one study only found a significant training effect in 3- but not in 4-year-olds (Ornaghi, Brockmeier, & Grazzani, 2011) while the other study found improvements in 3- and 4- but not in 5-year-olds (Grazzani & Ornaghi, 2011). Also focused on language but not necessarily mental state talk are interventions that use explanatory conversation to increase the understanding of emotions. Explanations have been implicated to play an important role in the acquisition and development of theory of mind (Wellman & Lagattuta, 2004). A study with primary school age children suggests that children who explained or listened to explanations of causes for a story-character's mixed and hidden emotional responses increased their understanding of it while children who listened to the same vignettes but answered questions about other aspects of the story did not (Tenenbaum, Alfierei, Brooks, & Dunne, 2008). Another language-intervention through drama-based roleplaying with kindergarten children found no effect on theory of mind compared to control subjects (Smith, 2011).

4.1.3. Computerized emotion understanding interventions – state of the field.

Putting forth the proposal to train emotion understanding with assistance of computer programs the question may arise why not just use the "real thing" (i.e. real human interaction). Given the premise that the optimal learning environment is planned and controlled human interaction with a specialist (therapist, teacher, etc.) there have to be other benefits to computerized training to justify using this kind of intervention. As already outlined in chapter 1.3, the potential benefits are mainly ease of accessibility of training, economy of resources (e.g. cost per training time), easy compatibility with other forms of intervention and possibly a higher entertainment value, leading to better adherence (Jak, Seelye, & Jurick, 2013; Burdea, 2003). Computerized interventions may also hold benefits for specific target populations. The medium especially lends itself to people with autism spectrum disorder because of its structured, consistent and untiring working mode which can help reduce frustration and anxiety (Panyan, 1984; Silver & Oakes, 2001).

How widespread is computerized training of emotion understanding? A broadly defined keyword search on PsycINFO on 22.05.2015 ([Emotion Understanding OR Emotion Knowledge OR Emotional Competence OR Emotional Intelligence OR Theory of Mind] AND [Intervention OR Training OR Treatment OR Therapy OR Program OR Teaching] AND [Computer Games OR Video Games OR Computer Assisted Therapy OR Computer Assisted OR Computerized OR Computer Applications OR Computers]) returned 69 results. Only eight of these related to computer-assisted training studies which targeted at least some aspect of emotional competence and none targeted the reflective level of emotion understanding (i.e. emotion regulation, mixed emotions, moral emotions). A literature search on PubMed on 22.05.2015 for key terms in Title/Abstract ([emotion understanding OR emotion knowledge OR emotional competence OR emotional intelligence OR theory of mind OR mentalizing] AND [training OR intervention OR treatment OR therapy OR program OR teaching] AND [computer game OR video game OR computer application OR computer assisted OR computerized OR computer] yielded 14 studies of which seven remained after removing double findings from the PsycINFO search and non-computerized training studies. Thus a total of 15 studies were identified (see Table 28). It is remarkable that a majority of the studies (8 of 15) only assessed the component of emotion recognition as an outcome variable even if incorporating other aspects of emotion understanding in the training. Even then, most studies training several aspects of emotional understanding (e.g. the programs Mind Reading or FaceSay^{TM}) are heavily focused on emotion recognition. No intervention was focused on the reflective level of emotion understanding. Most studies targeted adults (10 vs. five children samples) and clinical samples were prevalent (six with autism, six with schizophrenia, three healthy samples, one with PTSD, one with Down-syndrome). Study designs varied from multiple baseline assessments or quasi-experimental pre-post studies lacking controls to randomised controlled trials. A communality found was the overall small sample size used which is typical for clinical samples. Of the programs/tasks used in the identified studies, only FaceSay^{TM} qualifies as a game if you refuse to count multiple-choice quizzes with feedback as a game. MRIGE includes games but only in one of three applications. Taken together there is a clear dearth of research in the field of computer game based interventions to train emotional competence.

Table 28
Literature review of computerized interventions of emotion understanding

Study	age	group	programs	training target	outcome	Study type & sample size
Bölte et al., 2002	Y, A	ASD	Facial expressions	REC	REC	RCT (TG=5, passive CG=5)
* Cherkasova, 2012	A	SCZ	METT-SETT, MRIGE	REC	REC	RCT (CT+SCT=19, CT=21)
Hooker et al., 2013	A	SCZ	METT-SETT, MRIGE	REC, EXT, DES, BEL	REC	RCT (TG=11, active CG=11)
Huelle, Sack, Broer, Komlewa, Anders, 2014	A	H	Facial expression video clips	REC	REC	Pre-post (N=38)
Kandalaft, Didehbany, Krawczyk, Allen, Chapman, 2012	A	ASD	VR - Second Life	REC, DES, BEL	REC, DES, BEL	Pre-post (N=8)
Lindenmayer et al., 2013	A	SCZ	MRIGE	REC, EXT	REG,	RCT (CT+MRIGE =32, CR=27)
* Myszak, 2011	C	ASD	MRIGE	REC, EXT	REC, BEL	MBA (N=3)
* Perez, 2013	C	H	FaceSay™	REC	REC	
Popova et al., 2014	A	SCZ	Facial expression pictures	REC	REC	RCT (CT, facial affect training, TAU 19/19/19
Rice, Wall, Fogel, Shic, 2015	C	ASD	FaceSay™	REC, DES, BEL	REC, DES, BEL	RCT (TG=16, active CG=15)
Sacks et al., 2013	A	SCZ	METT-SETT, MRIGE	REC, EXT, DES, BEL	REC, REG	Pre-post (N=19)
Saunders et al., 2015	A	PTSD	Cogmed, tDCS	WM, tDCS	REC	Pre-post (N=4)
Silver & Oakes, 2001	C, Y	ASD	Emotion Trainer	REC, EXT, DES, BEL	REC, EXT, DES, BEL	Rand of matched pairs (TG=11, CG = 11)
Silver, Goodman, Knoll, Isakov, 2004	A	SCZ	Emotion Trainer	REC, EXT, DES, BEL	REC	Pre-post (T=20)
Swettenham, 1996	C	ASD, DS, H	False Belief Task	BEL	BEL	ASD=8, DS=8, H=8

Note. A=adults, ASD=autism spectrum disorder, BEL=beliefs, C=children, CG=control group, CT=cognitive training, DES=desires, EXT=external causes, H='healthy', MBA=multiple baseline assessment, PTSD=post-traumatic stress disorders, RCT=randomized controlled trial, REC=emotion recognition, REG=emotion regulation, SCT=social cognition training, SCZ=schizophrenic spectrum disorders, TAU=treatment as usual, TG=training group, WM=working memory, Y=adolescents
* Dissertational theses

As apparent from the literature search, computerized interventions to improve social cognition, theory of mind and emotion understanding, have mainly been developed for and evaluated with clinical populations, particularly autism and schizophrenia (see Ramdoss et al., 2012 for a review in autism). This probably has several reasons. First of all, deficits in theory of mind have been claimed to be hallmarks of autism for 30 years (Baron-Cohen, Leslie, & Frith, 1985; Baron-Cohen, 2001) which is also reflected in current diagnostic systems (World Health Organization, 1992; American Psychiatric Association, 2013). Research on social cognition deficits in schizophrenia has expanded significantly over the last 10 to 15 years. This has also been reflected in and stimulated by the Research to Improve Cognition in Schizophrenia (MATRICS) initiative of the National Institute of Mental Health (NIMH) (Marder & Fenton, 2004). The question, why the use of computer technology in social cognition training has been mainly limited to autism and schizophrenia while promotion of these competences in healthy children has focused on traditional, school-based programs is harder to answer. I can only speculate here but I assume that there are more implicit, attitudinal and normative than practical reasons for this phenomenon. Scientist-practitioners in clinical institutions may have always been more open toward assistive computer technologies than the average school teacher or pedagogue. With increasing availability and use of personal computers – IBM PCs and clones saw a tenfold rise in sales figures from 1984 to 1989 (Reimer, 2005) – they also started to find their way into clinics as devices aiding assessment and rehabilitation (Lynch, 2002). Already in 1985, Bracy, Lynch, Sbordone, and Berrol reported a sudden adoption of computers by clinics in the context of cognitive rehabilitation, mostly for neurological patients. A meta-analysis of controlled studies of computer-assisted cognitive remediation in schizophrenia shows that research started to prosper in the mid-90s of the 20th century (Grynszpan et al., 2011). My own literature search suggests, that programs for computerized training of social cognition and research evaluating them started to accumulate at the start of the new millennium. In contrast, recent reviews of intervention studies to promote emotional-competence did not report a single study utilizing a computerized training with healthy subjects (Sprung et al., 2015; Sklad, Diekstra, Ritter, Ben, & Gravesteijn, 2012; Schutte, Malouff, & Thorsteinsson, 2013). In comparison, my own broad literature search including social cognition, theory of mind and emotion understanding related concepts identified only three studies. In contrast, there is a large body of research on cognitive training in healthy subjects (Lampit, Hallock, & Valenzuela, 2014). In conclusion there is not much literature on computerized training studies of social cognition or emotional competence in healthy people.

4.1.4. Computerized emotion understanding interventions – description of existing training programs/protocols.

The first study to use some form of computerized intervention to improve theory of mind found through the literature search was that of Swettenham (1996). The study only focused on false-belief understanding, not emotion understanding but it is still mentioned here because it is exceptional for its early publication date. After it, no other study utilizing a computer based intervention was published for some years and studies only began to grow at the beginning of the second decade of the new millennium. The training/teaching procedure consisted of eight sessions of completing computerized Sally-Ann false-belief tasks spread over four days with a total of 48 trials. The program included feedback and helpful prompts. All text was read out loud by the experimenter. Twenty-four 3-year-old children with autism, down-syndrome and healthy controls, eight in each group, took part in the teaching. Performance in the computer-task increased for all children over the training, but less so for those with down-syndrome. Groups did not differ in a computerized Sally-Ann task - identical to the training program but without the helpful prompts and explanations - at post-test. In other false belief tasks (Sally-Ann doll task, unexpected content tasks, false belief task about the weather and time) healthy children did not differ from those with down-syndrome but children with autism did not solve a single of these tasks. In a follow-up 3 months later, performance for children with down-syndrome or autism was unchanged but healthy children had further improved as to expect in normal development. Interestingly the Sally-Ann vignettes presented on computer screen are referred to as *computer games* in the paper. No uniformly accepted definition of computer games (or games in general for that matter) exist but without having conducted a survey, one can still argue that few people familiar with computer games (that are commercially available for entertainment) would have classified this task, lacking interactivity, as a game.

Several computer programs exist that exclusively train facial emotion recognition. The facial affect training (Popova et al., 2014) consists of four tasks where subjects have to (1) decide whether the faces of two different persons show the same emotion, (2) identify two morphed expressions from seven possible “primary” emotions, (3) reproduce the sequence of two different emotions and (4) recall the location of a picture depicting a specific person/emotion combination. Patients with schizophrenia performed worse at the start of the training than healthy controls that only received one training session. After 20 daily one-hour sessions the schizophrenic treatment group showed a significant performance gain in all four tasks and performed similar to healthy controls in all but one task. The performance changes were also accompanied with changes in alpha power in the magneto-encephalogram (MEG). In another study utilizing a computerize training of facial affect recognition (Bölte et al., 2002), a small sample of five adolescent and adult males with the autism spectrum disorder (ASD) received training for five weeks, two hours a week while five control subjects with

ASD received none. The adaptive training consisted of 500 photographs of faces depicting one of seven “primary” emotions accompanied by visual and acoustic feedback. If an emotion was rated differently than intended, an explanation for the intended emotion as well as a cartoon with an example for the emotion was given. The experimental group improved in a task of reading emotions in the face and the eyes, compared to the control group, but not in other tasks like ratings of pictures of the International Affective Pictures System (IAPS). But facial emotion recognition can also be improved through unsupervised learning (i.e. without an external signal), at least in healthy people as another study shows (Huelle, Sack, Broer, Komlewa, & Anders, 2014). Thirty-eight young adult females received two sessions, roughly half of them with a 2-day interval and the other half with a 2-month interval in between. One training session consisted of watching 100 video clips of different length (2 to 10 seconds) and selecting one of four emotions (sadness, anger, fear, disgust) immediately after the clip adding up to a total duration of 40 minutes. The same clips were used in the second session in a different order. Subject’s performance improved within and between sessions regardless of inter-session interval.

An early program to train emotion understanding beyond emotion recognition is the *Emotion Trainer* (Silver & Oakes, 2001). It consists of four sections (facial emotion recognition, external causes of emotions, desire-based and belief-based emotions) where subjects are confronted with a picture and/or description and have to choose the right answer between four basic emotions. After each response, feedback and a hint, for first-time incorrect responses, or the correct choice for subsequent erroneous responses, is presented. In a randomized controlled study (Silver & Oakes, 2001), adolescents with ASD increased on a measure of external causes, desire- and belief-based emotions and the Strange Stories but not on a facial emotion recognition task after approximately 10 training sessions, distributed over two to three weeks. The training effect was correlated with the number of times the program was used. The program was also evaluated in a study with 20 male patients with schizophrenia (Silver, Goodman, Knoll, & Isakov, 2004). They used the program three times with an interval of two to three days in-between. Performance improved in two tasks of facial emotion recognition (distinguishing between four basic emotions; rating intensity of emotional valence of sad, happy and neutral faces) but not in a task where they had to rate whether or not a pair of faces differed in emotional intensity.

Another program to train social cognition, including emotion understanding, is *Mind Reading: An Interactive Guide to Emotions* (MRIGE) originally designed to use with autism spectrum disorder (Baron-Cohen, Golan, Wheelwright, & Hill, 2004). The program consists of three different parts. In the Emotions Library, 412 emotions, grouped into 24 categories, are conveyed through short videos, story lines and voice expressions with additional explanatory information provided. In the Learning Centre, the user can complete quizzes where she has to find facial affect expressions, or match emotional statements to pictures. Rewards for completing parts of the program consist of collecting various objects, some of them animated or playing music. A particular kind of reward is play time for the Games Zone. In the Games Zone, the user can play games incorporating facial affect recognition (e.g. card matching game, quizzes). In a first evaluation of the program (Golan & Baron-Cohen, 2006) with subjects with Asperger's Syndrome (AS) and High Functioning Autism (HFA), the experimental group used the program for 2 hours a week over 10 weeks at home. They improved in close transfer tasks of complex emotion recognition in faces and voices as well as emotion concepts recognized, using materials from the training. There was no far transfer to other emotion- or mental state recognition tasks (Reading the Mind in the Eyes / Voice) however. In another experiment, reported in the same study, subjects also with AS/HFA used the program again for 2 hours a week over 10 weeks but additionally received weekly group training in emotion identification by a tutor while the control group received a social skills group training. Subjects in the experimental group again improved in closed transfer tasks, of voice recognition, face recognition and number of emotion concepts recognized but not in far transfer tasks (Reading the Mind in the Eyes / Voice). Three recent studies, published almost concomitantly, used MRIGE, exclusively or in combination with other social cognition training programs, with schizophrenic patients. In one experimental study (Lindenmayer et al., 2013) with people suffering from schizophrenia or schizoaffective disorder one group received combined weekly treatment with MRIGE (1 hour) and a cognitive remediation program (2 hours) for 12 weeks while the other group only received the cognitive training (3 hours). The experimental group improved in a facial emotion recognition and discrimination, as well as social cognition task, compared to the control group. Interestingly the outcome in cognitive measures was also better in the combined group even though the cognitive-training-only group used the cognitive remediation program for more hours. In the second study (Sacks et al., 2013), patients with schizophrenia, who had participated as computer game control subjects in another study completed weekly computerized auditory cognitive training and computerized social cognition training over 10 weeks for a total of 50 and 12 hours respectively. The social cognition training was a combination of tasks taken from MRIGE and the *Micro Expressions Training Tool* and the *Subtle Expressions Training Tool* (METT and SETT) (Eckman, 2003). They improved in emotion perception and emotion regulation. The third study (Hooker et al, 2013) used the same combination of computerized cognitive and social cognition training programs with the same schedule except a slightly lower amount of social cognition training (5-15 minutes per day). A control group played non-specific computer games. The treatment group improved in emotion perception compared to the control group but unexpectedly not in a facial emotion recognition task, very similar to the training program, completed during fMRI acquisition. Change of activity in the right amygdala during recognition of happy, surprised and fearful faces predicted improvement in the emotion perception task only in the treatment group.

SocialVille is an online training program, developed by company Posit Science, incorporating principles of neuroplasticity-based learning, trying to target the underlying brain systems instead of the impaired higher-level behaviours (Nahum et al. 2014). One block of tasks focuses on speeded facial affect processing (e.g. recognizing emotions, matching emotions, executive aspects like delaying responses), another block is centred on working memory (largely of facial affect) while a few additional tasks address prosody (vocal affect) and social situations. Subjects with schizophrenia used the program for 24 hourly sessions on average over 8 weeks, three times a week. Subjects' performance significantly increased in seven of 10 of speeded facial affect training tasks but only in one of six working memory tasks, however both composite scores demonstrated significant change. Reaction times for a facial memory task and a prosody identification task decreased. There was no significant change in the perceiving and managing emotions measure of the MSCEIT. Participants also showed a significant increase in self-reported social functioning and a decrease in behavioural inhibition.

There is also some evidence that gains in emotion understanding can be achieved through computerized cognitive training alone, without specifically targeting emotion understanding or social cognition, at least in clinical populations (Saunders et al., 2015). Four patients with post-traumatic stress disorder received a working memory training combined with transcranial direct-current stimulation (tDCS). First, subjects underwent five weekly 20 minutes tDCS treatment sessions. Subsequently they completed 30-45 minute sessions of working memory training five times per week for 5 weeks. Two of those patients also completed a questionnaire of emotional intelligence and reported improvement in the ability to relate to and understand other people.

An interesting novel approach is the use of virtual reality software to practice social cognition, social skills and social functioning. In a training study of young adults with high-functioning autism received 10 training sessions twice a week (Kandalaft et al., 2012). In a

3D virtual world participants created avatars to match their appearance and engage in dynamic situations commonly experienced by young adults (e.g. meeting someone new, job interview, managing conflict, etc.) with avatars controlled by a confederate clinician. A clinician acting as coach also participated through an avatar and asked the subjects about their insights afterwards and provided detailed feedback. Subjects improved in verbal and nonverbal emotion recognition as well as theory of mind.

The focus of computerized social cognition training programs on facial affect recognition somewhat resembles the predominance of false-belief-focus in research and assessment of theory of mind. The latter has surely been furthered by the early success of the studies by Perner and colleagues. The former has probably also science-historical reasons. Facial expressions of emotions is one of the first aspects of emotion understanding and social cognition that has been subjected to scientific inquiry. It has its roots in Darwin's seminal work *The expression of the emotions in man and animals* (1872) and has also been heavily influenced by the work of Ekman (e.g. Ekman & Friesen, 1971). But there may also be practical reasons at work. While every development of a training or treatment program constitutes an extraordinary amount of effort in itself, facial affect recognition trainings are relatively straight forward to implement in comparison to other branches of social cognition. They usually consist of a database of photographs, presented in the form of quizzes (e.g. identifying a specific emotion, matching two faces, etc.) and the occasional instruction, which facial cues to look out for. The computerized training of understanding of external causes of emotions or belief-based emotions calls for more complex procedures and stimulus creations, more so if the program should be interactive and not merely consist of explanatory text. For example a single training vignette for the understanding of belief-based emotions is arguably hard to realize on a single slide because representing the beliefs of different agents and their ensuing emotions on screen will require several different depictions and – again if done in an interactive fashion – more program logic to realize.

While some programs noted above (e.g. MRIGE, FaceSay™) include game-like elements or small games (e.g. quizzes, matching emotions, puzzles) in their training protocol, none tries to deliver its treatment solely through game-like mechanisms nor are they designed as a coherent computer game. This may or may not be beneficial to the effect of the training, and variables like target competence (e.g. facial emotion recognition, emotion regulation) or level of processing (e.g. explicit and elaborate understanding vs. more implicit, automatic processing) are likely to interact with the degree of interactivity and gamification in a training. Studies comparing treatment components delivered in a more traditional Q&A- or quiz-format vs. game-format would be helpful in answering these questions.

4.1.5. Goal of study 3 – developing and evaluating EmoJump, a computer game to promote emotion understanding.

Let me reiterate three important assumptions explored in previous chapters:

(1) Emotion understanding is linked to mental health and related social outcomes (see chapter 3.1.7). (2) Promotion of mental health and prevention of mental illness plays an important role in reducing the global burden of disease (WHO, 2002, 2004a, 2004b). (3) Playing computer games is a common leisure activity among children and adolescents (see chapter 1.1). Following these three assumptions, goal of study three was to develop and evaluate a computer game promoting emotion understanding as a tool to increase emotional resilience.

4.2. Methods.

4.2.1. Development of EmoJump.

Development of EmoJump started in October 2012. The goal was to develop a computer game to improve emotion understanding in children (see Appendix F for credits of the people involved in various aspects of the game development). We finally decided to implement the game as a crossover between two genres: "side-scroller-platformers" (also known as "jump and run" games) is a type of game where the player controls a character, and has to avoid obstacles while navigating a world in 2D from the left of the screen to the right. In "endless-runner" style games, the character can only be controlled to avoid obstacles and collect items but not stopped. This had several reasons: First, platformer games are among the most played games in children between 6 and 13 years (Medienpädagogischer Forschungsverbund Südwest, 2011). Second, mini-games (including games of dexterity and skill, like platformer games) show an even distribution in preference among gender while genres like "shooters" or strategy games are preferred by boys, at least in children and adolescents between 11 and 18 years (Institut für Jugendkulturforschung, 2008). Lastly, compared to other genres (e.g. 3D games, point-and-click adventure games) the amount and complexity of visual assets that have to be created for such a game is kept within reasonable limits which made the development possible with the resources at hand. The game was realized in HTML5 by two master students of the entertainment computing group under Univ.-Prof. Dipl.-Ing. Dr. Helmut Hlavacs. For details of the technical realization see Natascha Schweiger (2014) and Katharina Meusburger (2014).

4.2.1.1. Description of game-play.

The player can select between two avatars, a girl and a boy, and choose a unique username. Next, the player selects a level to play (external causes, belief-based emotions, mixed emotions). Levels are further divided into sub-levels that differ in how many alternative emotion response options are presented (i.e. distractor emotions) and in speed (i.e. time to react). A video-tutorial explains gameplay and controls for each level. The core gameplay (explained using the example of the first level – external emotions) is as follows (also see Figure 13 below).

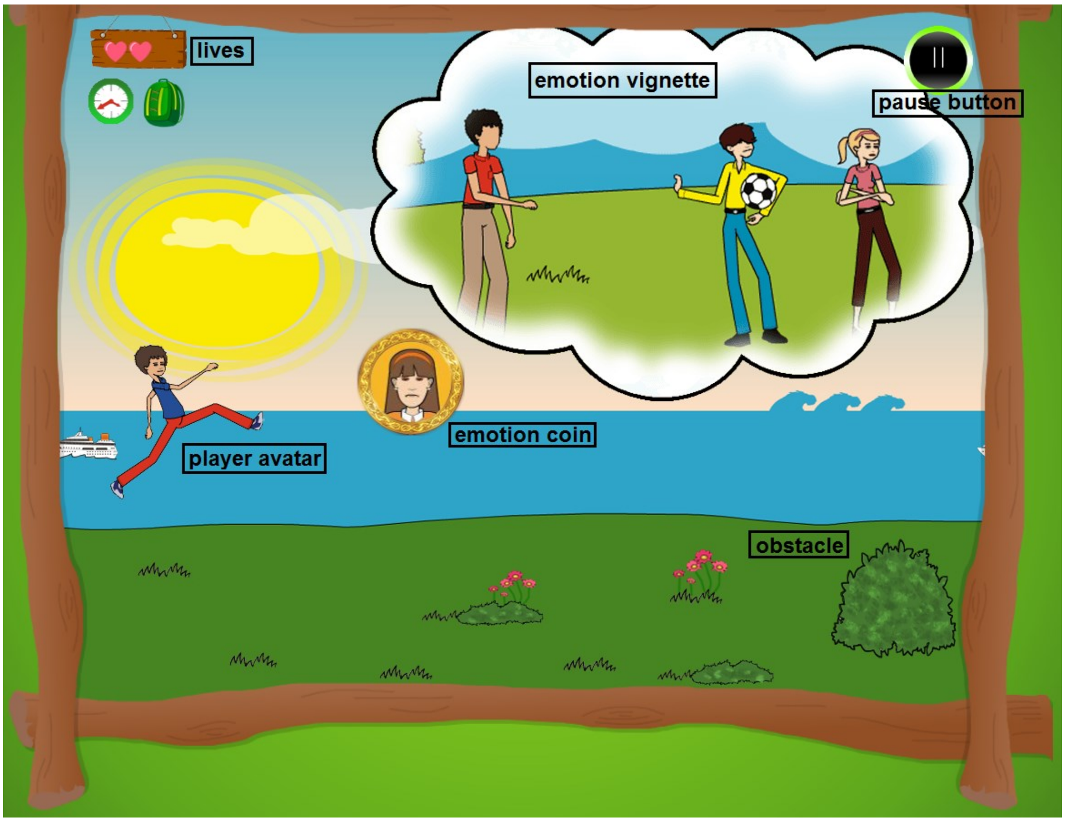

Figure 13. Description of EmoJump interface key elements and example vignette for external-causes visible during the collection phase

The player sees a cartoon in the background depicting a social scene (e.g. a child being excluded from play) and has to find out how the main-character with the red shirt (hereafter referred to as *character*) feels. Importantly the face of the character in the cartoon is empty. This prevents deducing the correct emotion simply by recognizing facial affect which would override the need to interpret the social context and take the perspective of the character. On pressing the space bar, the avatar, controlled by the player, starts running from left to right (actually the background moves from right to left creating the illusion of a moving figure). Additional parallax scrolling of background elements like animals or people creates the illusion of movement. Faces on coins expressing different emotions (happy, sad, angry, scared), as well as obstacles start to appear randomly on the right side of the screen. At one time there cannot be more than one coin and one obstacle on screen however to prevent unmanageable constellations. The player has to collect the right emotion-coins with the help of well-timed jumping (using the space bar) while avoiding the incorrect ones as well as obstacles. After the correct emotion coin to be collected has appeared three times, regardless whether they have been collected or not, the scene ends and a scoring screen for the scene is shown (see chapter 4.2.1.3). Each time the player-avatar crashes into an obstacle, she loses one "life" (i.e. available retries represented through heart symbols). If she loses all three lives the current scene starts again. This lenient limit was introduced to increase engagement and deter from crashing into the obstacles for fun while at the same time being not so hard as to become frustrating and forcing players to frequently repeat vignettes. To further increase emotional salience and address different modalities, coins are accompanied by a corresponding vocal affect expression (laughing for happy, crying for sad, a gasp for scared and a snarl for angry). There is a special power-button (a lightbulb) the player may use a limited number of times (not every cartoon). When pressing the button, a voice, in a neutral tone, explains the scene in the form of a thought monolog of the character in the scene. This serves to help understand the cartoon's narration while not giving away the correct answer.

Level 2 (belief-based emotions) is based on unexpected content false-belief tasks like the smarties-task (Gopnik & Astington, 1988). In this task a child is presented a box of sweets (or holding some other supposedly desired content) and is asked what she thinks is in the box. Afterwards the box is opened to unexpectantly reveal some other, undesired content. The child is then asked what she thought was in the box before it was opened. In a variation of this task, the child is asked how she feels seeing the actual (disappointing) content and how she felt before the box was opened. The child has to acknowledge that she held a different belief and subsequently a different emotion before opening the box. Accordingly this level is built around not a single cartoon but a short comic strip of two or three pictures (see Figure 14 for an example). The player first watches the strip sequentially (i.e. the picture occurring first in the story is highlighted first). The first picture displays a certain situation (e.g. a boy thinking that people might throw tomatoes at him after a failed performance), while the last picture displays some turn in the story for better or worse (e.g. people are applauding him). The player now has to collect the emotion-coins that correspond to the affective state in the last picture as in the external-causes level (i.e. post belief change). After three correct coins have appeared, the first picture of the comic-strip is highlighted and coins have to be collected for this situation despite knowing the outcome of the story (i.e. pre belief change). In this level the player's avatar runs from the right of the screen facing left, symbolizing the retrospective nature of asking about the emotion that the character held before.

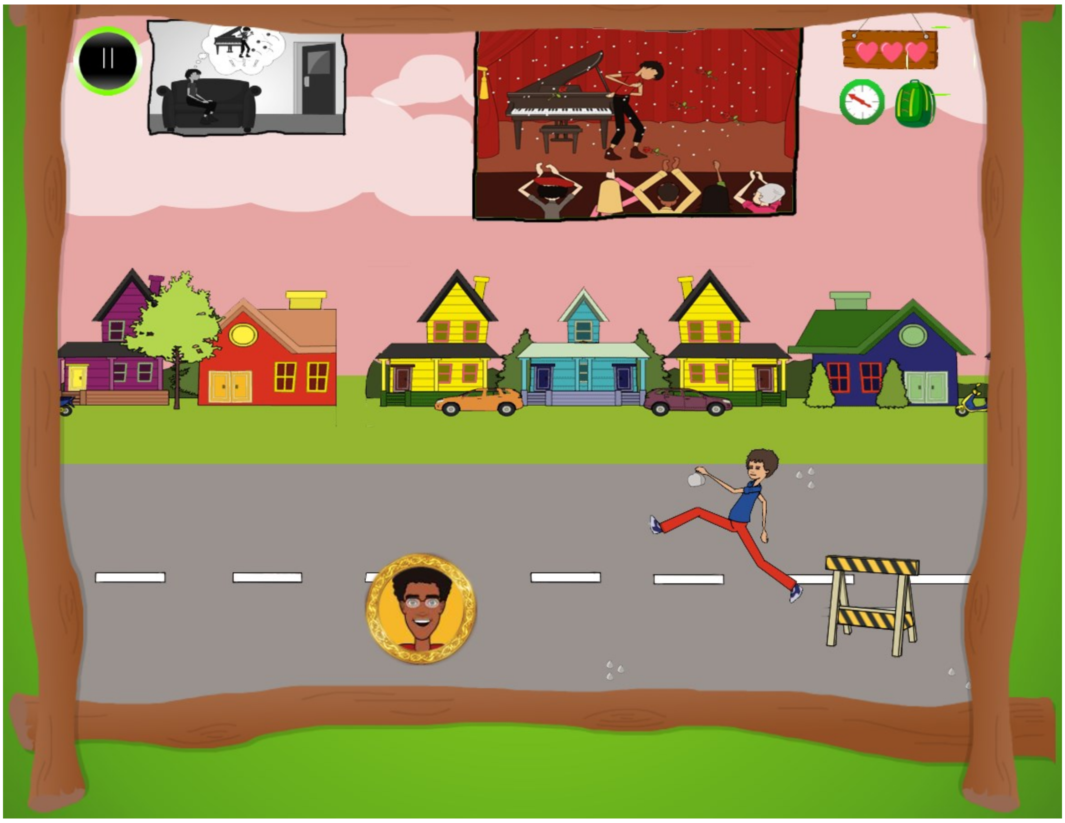

Figure 14. Example vignette for belief-based emotions showing the post belief change phase.

In level 3 (mixed emotions) the player is first shown a cartoon similar to level 1. Before the avatar starts running two thought bubbles are shown that display thoughts the character in the cartoon may associate with the situation (see Figure 15 for an example). The first picture for example might show a boy lying in his bed being sick. In the first associated thought bubble he is looking out of the window from his bed at his friends playing outside. In the second thought bubble his mother is bringing him a toy to his bed. The two thought bubbles are again replaced by the initial cartoon and collecting of the coins ensues. After three correct coins of each emotion associated with the cartoon have been presented, scoring occurs (see chapter 4.2.1.3). To add variety each scene starts with a mini-game where the player can steer the avatar left and right and has to collect an object flying around above her which hints at the cartoon to come (e.g. the toy in the previous example).

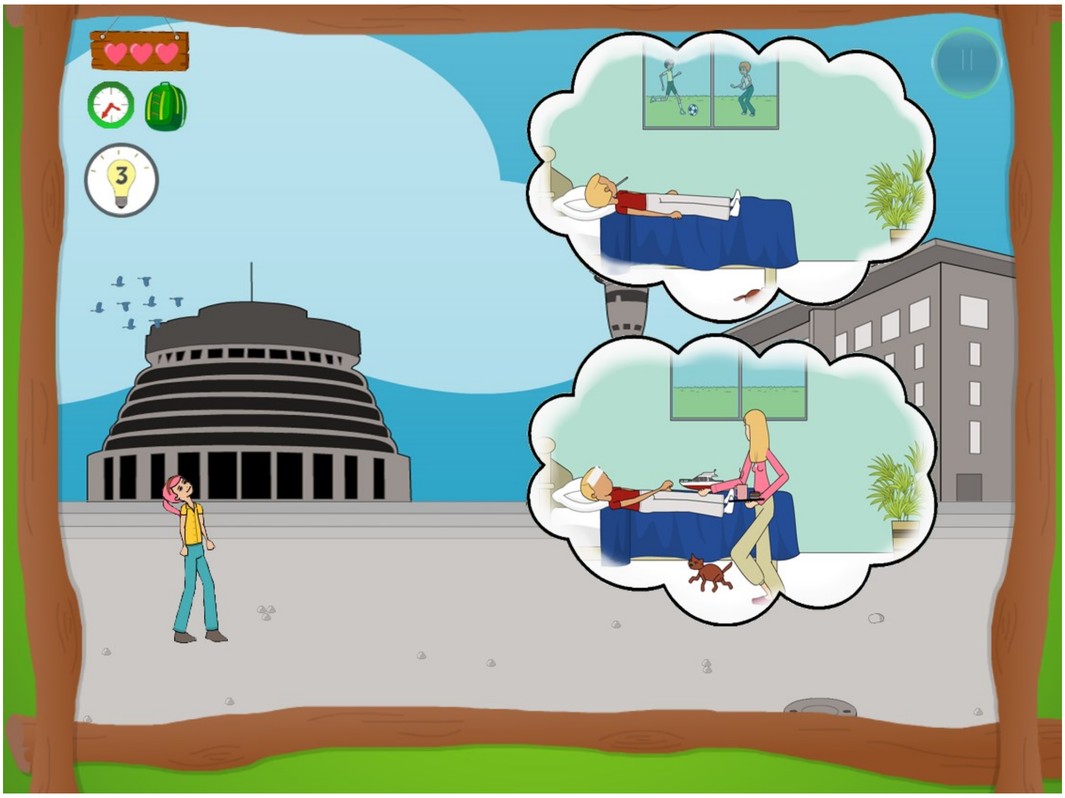

Figure 15. Example vignette for mixed emotions showing the two associated thought bubbles not visible during collection phase.

4.2.1.2. Differences between sublevels.

Each level (i.e. external causes, beliefs, mixed emotions) consists of 12 sublevels that vary in difficulty through increased speed and number of distractors (i.e. incorrect emotion-coins appearing; see Table 29). Each sublevel consists of several scenes (i.e. different cartoons for which emotions have to be collected). For level 1 (external causes) and 2 (belief-based emotions), only one correct emotion has to be collected while for level 3 (mixed emotions) two emotions are correct in each scene. Regarding distractors (i.e. incorrect emotions appearing that have to be avoided) levels 1 and 2 include one, two or three distractors in sublevels 1-4, 5-8 and 9-12, respectively. In level 3, where 2 emotions are allocated as correct emotions already, only one (sublevels 1-6) or two (7-12) distractors are present. Each instance of number of distractors is available in four different speeds. For each scene, three correct emotion-coins appear with distractor emotions randomly interspersed. After they have passed (regardless of whether they are collected or not), the scene ends and a feedback screen appears.

Table 29
Conceptual description of EmoJump sublevels

Level	Number of correct emotions			Number of distractor emotions			Speed (higher numbers indicate higher speed)		
	E	B	M	E	B	M	E	B	M
1	1	1	2	1	1	1	1	1	1
2	1	1	2	1	1	1	2	2	1
3	1	1	2	1	1	1	3	3	2
4	1	1	2	1	1	1	4	4	2
5	1	1	2	2	2	1	1	1	3
6	1	1	2	2	2	1	2	2	4
7	1	1	2	2	2	2	3	3	1
8	1	1	2	2	2	2	4	4	1
9	1	1	2	3	3	2	1	1	2
10	1	1	2	3	3	2	2	2	2
11	1	1	2	3	3	2	3	3	3
12	1	1	2	3	3	2	4	4	4

Note. E = external causes, B = belief based emotions, M = mixed emotions

4.2.1.3. Scoring and success criteria.

On the scene-scoring screen, the player sees how many correct and incorrect coins she has collected, but not which emotions were considered correct or incorrect by the game. This should enhance replayability and keep the player guessing which emotion could also be attributed to this situation, possibly promoting a more elaborate and flexible understanding of emotions. While training programs providing direct feedback seem to be the norm, training effects of emotion understanding have also been shown for unsupervised learning (Huelle et al., 2014). For a scene to be considered completed successfully at least two of three correct emotions have to be collected, while not collecting too many wrong emotions, thus preventing success by merely collecting all possible coins. If the number of presented incorrect emotion coins (C_{pi}) is smaller than the number of presented correct emotion coins (C_{pc}), the success criterion ($X \geq 2$) is the number of collected correct emotion coins (C_{cc}) minus the number of collected incorrect emotion coins (C_{ci}). Elsewise, C_{ci} is weighted by the number of presented incorrect emotion coins exceeding the presented correct emotions coins according to the following formula: $X = \sum C_{cc} - [(\sum C_{pc} \cdot \sum C_{ci}) / \sum C_{pi}]$. This was made to account for occasional erroneous collections occurring through timing or controlling problems. Thus if incorrect emotion coins occur much more often than correct emotion coins, a scene is still counted as completed successfully if three correct coins are collected and incorrect coins are not collected consistently. A level is passed at minimum level of success (i.e. bronze cup) if more scenes have been completed successfully than not. To enhance motivation, higher percentages of successful scenes have to be reached to achieve the silver (80%) or gold cup (90%).

4.2.1.4. Stimulus creation and evaluation.

Cartoons for the game were created by several master students of the department of clinical child and adolescent psychology working on the project (see Appendix F) with the online cartoon creation program *Pixton* (Pixton Comics Inc., 2014) and *Adobe Photoshop* (Adobe Systems, 2014). The goal was to cover emotion eliciting situations typically experienced by children (e.g. fear of animals, separation anxiety, sadness after social rejection or loss of a loved object, anger over someone teasing or impeding the fulfilment of one's desires). This of course means, that not all emotion eliciting situations may be typical for each age. This was a trade-off to achieve a broader applicability. For older children some situations may not be interpreted as eliciting an emotion anymore (e.g. separation anxiety) while for younger children, the range of situations experienced is arguably smaller. Moreover some situations might evoke an emotional response in one person but not in another, across ages. Even though a general consensus about which emotional response is to expect in a given situation often exists, there is variance in normal reactions, even in healthy people. When seeing someone break a beloved toy, the person might show a response of sadness about the loss of the object. He or she might also primarily be angry about the person breaking the toy. The emotional response can also be heavily influenced by beliefs. If the person suspects that the one responsible deliberately destroyed the toy the response is more likely to be one of anger than if it seems to have been an accident. In real social situations this is in part owed due to incomplete information available for interpretation and individual characteristics in responding. Illustrations of such situations provide even less information and social cues which leads to greater ambiguity. We took a pragmatic approach and accepted that a certain amount of ambiguity cannot be eliminated in a game that essentially uses a forced choice format and where response alternatives are scored as either correct or incorrect. Yet to gauge the most likely emotional response we validated the stimulus material. Because of the elongated process of development and different people involved in the project this was done in two steps. First, cartoons of the first level, which had been created earliest, were categorized by a sample of primary school children. Secondly, in a later stage of development, cartoons of level two and three were independently categorized by three master students working in the project. Cartoons for which there was no complete agreement, or for which one or more raters anticipated potential problems in comprehension, were subsequently given a sample of psychology students for categorization to assess which emotion (or emotions in the case of mixed emotions) was most frequently attributed with a certain cartoon (see chapter 4.2.3). Particularly cartoons for which there was not a clear preference of one emotion over the others were placed in such sublevels that the competing emotion(s) did not appear as coins in the scene. In the end we had to accept a certain amount of ambiguity in the cartoon-emotion mappings and that the emotions that had been defined as correct were not always the only conceivable emotions in a situation. Our goal was not to promote emotion understanding by having the subjects memorize a catalogue of situation-emotion mappings but by repeatedly engaging in thinking about emotions in social situations. This approach is similar to interventions where children are read stories by an experimenter and engage in conversations about emotional aspects (e.g. Ornaghi, Brockmeier, & Grazzani, 2014). Therefore, the game-score was expected to represent only a fuzzy measure of success. To take this into account and avoid frustrating the player we decided to set a very liberal success criterion for sublevels at this stage of game development (see chapter 4.2.1.3). Thus, success in a sublevel can even be achieved if some scenes are interpreted differently in terms of emotional responses than by us or the pilot sample.

4.2.2. Pilot study 1.

The aim of the first pilot study, early in game development was primarily to gauge if the cartoons developed for level 1 (external causes) of EmoJump are appropriate for primary school age children and the emotion labels for each cartoon, as intended by the creator, are congruent with the labels assigned by a majority of the children. The understanding of external causes of emotions develops approximately between 3 and 4 years (Pons et al., 2004). In the Test of Emotion Understanding external causes is measured via cartoons that are accompanied by a story read by the experimenter. This is easier than trying to figure out the embedded emotion in a cartoon alone, without additional contextual information, like it is the case in EmoJump. Thus we expected that even children in an age at which mastery on standard measures of understanding of external causes of emotions is achieved would perform below-ceiling when labelling the cartoons. To make sure that the sample to validate the stimulus material is in fact a normal developing one in terms of emotion understanding they also completed the Test of Emotion Understanding.

A master student recruited and assessed 41 children (25 male) from three first classes in a primary school in Vienna. The sample was split to reduce the burden of administration. Group one (21 children, 13 male) rated one half of the stimuli (vignettes 1-46) and group 2 (20 children, 13 male) the other half (vignettes 47-91). Children between groups were matched in terms of age and gender. Mean age was 7.38 years (*SD* = .45, range: 6.28-8.39) for group one and 7.63 years (*SD* = .35, range: 6.94-8.39) for group 2. Average total score of emotion understanding was 5.76 (*SD* =1.55) and 5.9 (*SD* = 1.41) for group one and two respectively. This is comparable to mean scores found in other studies, using the TEC with similar age groups of normally developing children: 6.13 (*SD*= 1.24) in children aged 7.46 years (*SD* = .25) (Pons, Lawson, Harris, & de Rosnay, 2003), 5.80 (*SD* = 1.73) in children aged 7.17 years (*SD*=.16) (Pons et al., 2004), 5.93 (*SD* = 1.37) in children aged 7.22 years (*SD* = 0.25) (Pons & Harris, 2005). This suggests that the sample was adequate to evaluate understanding of the stimuli developed for level 1 of EmoJump. External emotions score in the TEC was correlated with the accuracy ratings of the vignettes, further supporting the validity of the stimulus set (Spearman *r* over both groups: $r(39)=.38$, $p<.05$; for group 1: $r(19)=.51$, $p<.05$; for group 2: $r(18)=.38$, $p<.10$).

A total of 91 cartoons (25 happy, 21 sad, 24 scared, 21 angry) were rated. Percentages of agreement between intended emotion labels and emotion attributions varied greatly for the four emotions over both groups. For happy the agreement was highest (*M* = .89, *SD* = .10), followed by scared (*M* = .71, *SD* = .22) and sad (*M* = .68, *SD* = .16). For angry there was by far the lowest agreement (*M* = .26, *SD* = .21). For the vast majority of all anger-cartoons (90.5%) agreement was 50% or less (mean agreement of 18.75%, *SD* = 9.83). All but one of these cartoons instead received the label *sad* by the majority with a strong consensus among children (*M* = 66.07, *SD* = 13.86). These findings are in line with previous research that found anger to be more difficult to identify in faces and descriptions of social situations than other emotions (Denham & Couchoud, 1990; Wintre & Vallance, 1994). It has also been shown that for young children, the emotions angry and sad are hard to differentiate in social situations (Farber & Moely, 1979; Stein & Jewett, 1986; Stein & Levine, 1987). Apart from developmental issues, this might be in part due to inherent conceptual similarities between these two emotions. For example, the same situation might often either elicit anger or sadness, depending on the agent focussing on the loss of an object or on another person or situation leading to the loss. Another possibility is that some children can understand that they or other people might perceive anger in a situation but refrain from reporting so because of social desirability (Felleman, Carlson, Barden, Rosenberg, & Masters, 1983). These findings in addition to the ambiguous nature of static, nonverbal illustrations of social situations might explain why so few children attributed anger to vignettes designed to fall into this category.

Cartoons for which more than 50% of the children selected the intended emotion, cartoons were included in the stimulus pool for EmoJump. Cartoons with an agreement of 50% or less or for which comprehension problems had become evident during assessment were again validated in a second pilot study with adults before deciding how to proceed.

4.2.3. Pilot study 2.

Goal of pilot study 2 was to collect consensus ratings from a sample of adults for cartoons of level 2 and 3 for which no complete agreement between three independent raters had been achieved or for which one or more of the raters had expressed concerns regarding comprehension. Consensus ratings have been used in development of measures of emotion understanding because of lack of veridical criteria (Mayer, Salovey, & Caruso, 2012). Additionally cartoons from level 1 were included if 50% or less children in pilot study one had agreed with the intended emotion label. If criteria for inclusion in the pilot study applied to a cartoon, all pictures for which subjects had to choose an emotion (e.g. level 2: before – after, level 3: mixed emotions) belonging to that scene were included for validation. The final count was 227 pictures (level 1: 59, level 2: 100, level 3: 68). Participants first completed a few demographic questions like gender and age. Pictures were then presented with the online survey tool SoSci Survey (Leiner, 2014). Presentation was fixed, in level order and the same for all subjects. For each picture, participants had to label which of four emotive states the character was most likely to experience in the situation: happy, sad, scared or angry. They were also asked if they understood the cartoon. If they selected no, they were prompted to briefly describe why or which aspects of the cartoon were ambiguous.

For pilot study 2, 50 psychology students who received bonus course credit for participation were recruited. Two subjects reported to have a dyschromatopsia and were excluded from the study to account for the fact that the main character in the cartoons to be evaluated is designated by a red t-shirt. Average age of the remaining 48 participants (41 female) was 26.61 years (SD = 5.32). Subjects completed the task individually in a room, attended by an experimenter. Mean time of completion was 51.67 minutes (SD = 8.58).

Agreement of the majority of subjects regarding the the intended emotion was found for 181 of 227 pictures. Following rules were then applied: Pictures which turned out to be highly ambiguous were excluded from the item pool. This was the case if either three emotions were similarly frequently nominated (<3% difference) or if the majority of participants had marked the picture as unintelligible. Scenes for level 2 were also discarded, if both pictures of the comic strip (before – after) were labelled with the same emotion. For level 3, scenes were not discarded if receiving the same emotion label since in the game the player is explicitly instructed to collect two different emotions. The second most frequent emotion was locked in the game if the margin was 20% or less. That is, the emotion for the competing interpretation did not appear as an emotion-coin to collect. Additionally, following rules applied to the 46 pictures for which no agreement with the intended emotion was achieved: If the most frequently chosen emotion-label for a cartoon exceeded the frequency for the intended emotion-label by 10% or more, the new emotion-label was assumed for that cartoon. If the top emotion-label for the cartoon exceeded the frequency for the intended emotion-label by less than 10%, the original emotion-label was retained but the competing emotion label was removed as a response option. This forced the player to consider a different attribution. The main reason we did not recode or exclude these pictures was, that the majority of them belonged to the intended anger-category which would have considerably voided this category. As mentioned above, attributing emotional responses in social situations is more ambiguous for anger than for other emotions. Furthermore, the scoring schema (described in chapter 4.2.1.3) was designed to be lenient enough not to require cartoons to be indisputable. In total 19 pictures were discarded, 21 were recoded and for 31 pictures the competing emotion was excluded.

4.2.4. Playtesting.

After completion of the EmoJump prototype to be used in this study, a playtesting session was conducted to evaluate if primary school-age children can manage gameplay and controls and assess the difficulty of parameters like speed and obstacles. Thirty-five children (18 female) from the *A1 Internet für Alle Campus*, aged eight to 12 (M = 10.77, SD = 0.84) played EmoJump for a short time and completed a short survey afterwards in which they were asked about controls, gameplay elements and goals and if they have had fun playing the game. *A1 Internet für Alle Campus* is a cooperation between the communication provider *A1* and the *Vienna University Children's Office* and provides attended internet access for disadvantaged children who might otherwise have no access, as well as computer and media workshops. To limit time-strain and cover more sublevels, subjects were split in two groups. All children played sublevel 1 of level 1 to become acquainted with the general gameplay and controls. Afterwards, one group played two sublevels of level 2 and the other played two sublevels of level 3. All children could explain what they had to do in the game and how the controls worked, after the short play session. A vast majority (83%) of the children reported they have had fun and had wanted to beat the game. Faster sublevels turned out to be more fun than the slower ones. Overall, the majority (57%) of children reported that the game was too easy for them while 9% said it was too difficult and 34% were satisfied with the level of difficulty. We refrained from increasing the difficulty for several reasons. First, the playtesting sample did not extend to the lower age-border intended for the game. An increase in difficulty might have made the game too difficult for 6- and 7-year-olds. Secondly, the children that frequent the *A1 Internet für Alle Campus* are not naïve to computer games since between workshops they can play a selection of games freely. Additionally, during the playtesting, the need for further optimization became apparent as the browser-game did not run smoothly on all devices.

4.2.5. Participants.

An at-risk sample of children between 5 and 12 years was recruited through contacting clinical psychologists, counselling centres for parenting, familial-conflicts and children with parents who suffer from a psychological disorder, children's homes and special educational facilities and asking for permission to put information brochures and flyers on display. Contact persons were also asked to provide the brochures to parents. Additionally, flyers were posted on internet forums about parenting, foster-parents and social media sites. The information brochures and flyers invited parents to participate in a study to train emotional competence through a novel computer game. To lower the burden of participating in the study, potential participants were told that the child can play the game at home if it adheres to the training schedule. Two children's homes, one special education facilities and six families responded to the information material. One children's home and four families decided to not participate due to time constraints. The special education facility provided 24 children, the children's home three, and the two families, three children. Children were randomly assigned to the experimental group (EG) or the control group (CG). Two children from the children's home dropped out because they did not adhere to the training regime when playing alone and it was not possible to accompany them at the children's home. One child from the special education facility dropped out because the pretest could not be completed on three different occasions due to hyperactivity. One 4-year-old child from a family which provided two children was allowed to participate in the study but excluded from the analysis because it did not meet the age requirement. After consultation with the special education facility it was decided that the children would be accompanied by master students while playing the game because most of them would have been unable to do so alone because of hyperactive and impulsive behaviour. The final sample consisted of 26 children (13 in each group). The recruited sample was dominantly male with only one female in each group. Groups did not differ with respect to age in months (EG: $M = 128.31$, $SD = 22.08$, Range: 95 to 153; CG: $M = 124.85$, $SD = 22.63$, Range: 86 to 152; $t(24)=.41$, $p=.68$) or emotion comprehension at pretest (EG: $M = 7.38$, $SD = 1.05$, Range: 5.35 to 9; CG: $M = 7.53$, $SD = 1.04$, Range: 5.42 to 9; $t(24)=-.36$, $p=.72$). Performance for the *Ishihara plates* indicated a red-green colour-blindness for two children. Because this is only a potential issue for the game (as the target character for mental state attributions is designated by a red shirt) and both subjects had been randomly assigned to the control group, they were retained in the sample.

Groups did not differ significantly in vocabulary (EG: *M*=8.31, *SD*=2.90; CG: *M*=10.08, *SD*=2.18; $t(24)=-1.76$, $p=.09$, $r=.34$) but there was a trend with a medium effect size. Additionally, the distribution showed that in the experimental group five subjects scored one standard deviation below the norm compared to none in the control group. If experimental groups differ in respect to a variable it is not correct to include it as a covariate because variance shared with the covariate and the independent variable cannot be disentangled (Lord, 1967; Miller & Chapman, 2001). Thus unfortunately, verbal competence could not be included as a covariate to account for differences within groups.

4.2.6. Procedure.

Children in the experimental group played the training game for 12 sessions, 20 to 30 minutes each. Three master students, working on the project (SA, TR, VZ) accompanied the sessions to ensure adherence to the training protocol but did not provide help. Each level (external causes, belief-based emotions, mixed emotions) was played for four sessions irrespective whether all sublevels per level were completed or not. Children in the control group were told that they could play the game after the end of the study and remained in classes while children in the experimental group were taken out for the training sessions. Before and after the training, children's emotion understanding and related variables were assessed by the master students. There was no significant difference of average time between pretest and posttest between the groups (EG: *M*=35.23 days, *SD*=5.29, Range: 28 to 49; CG: *M*=36.31, *SD*=8.41, Range: 28 to 58; $t(24)=-.39$, $p=.70$). Most subjects in the experimental group started with the training on the same day as the pretest (*Mdn*=0, *IQR*=2) while the median time between last training session and posttest was 3 days (*IQR*=6).

4.2.7. Measures.

4.2.7.1. Computerized Test of Emotion Understanding.

To measure changes in emotion understanding we used the cTRUE, previously developed. This experimental task has been demonstrated to correlate strongly with established measures of emotion understanding like the Test of Emotion Understanding and the Assessment of Children's Emotion Skills in study 2. Because we were notified about possible problems of a long assessment procedure by caretakers of the recruitment institution and one hallmark of children with ADHD is a shortened span of attention (American Psychiatric Association, 2013) we decided to trim the cTRUE to avoid noncompliance of participants and resulting biased or lost data. The focus in respect to number of items remained on components that were targeted by the game (i.e. external causes, belief-based-emotions, mixed emotions; see Table 30 for a complete list of changes in number of items). The shortened version of the cTRUE task took about 20 minutes to complete.

Table 30
Number of items per component in different cTRUE versions

Component	present study	original cTRUE
Recognition	15	15
External Causes	10	15
Reminder	2	3
Desires	1	2
Belief	4	5
Hiding	4	5
Regulation	3	5
Mixed	4	5
Morality	1	3

4.2.7.2. Flexibility and Automaticity of Social Cognition (FASC).

Two vignettes from Hayward et al. (2016) were used, one nonverbal unambiguous (dog) and one nonverbal ambiguous (playground). See study one for a detailed description of the FASC procedure. Because of inconsistencies in the number of follow up questions posed by the master students collecting the data, only responses for the initial question and the first follow up prompt were analysed.

4.2.7.3. Verbal intelligence / vocabulary.

We used the German version of the Vocabulary subscale from the Wechsler Intelligence Scale for Children IV (Petermann & Petermann, 2011) to achieve an estimate of verbal intelligence. See study one for further details of this measure.

4.2.7.4. Executive Functions.

Subjects also completed the Dimensional Change Card Sort Test as well as the Flanker Inhibitory Control and Attention Test from a pre-release version of the NIH Toolbox (see www.nihtoolbox.org for further information) to assess any potential training effects on executive functions as found in studies with commercial games (Green & Bavelier, 2003), and a game specifically designed to train cognitive control (Anguera et al., 2013). See study one for a detailed description of both tasks.

4.2.7.5. Positive and negative affect.

We used a short version of the Positive and Negative Affect Schedule for Children (Laurent et al., 1999), developed by Ebesutani et al. (2012) which consists of five words for positive affect (joyful, cheerful, happy, lively, proud) and five words for negative affect (miserable, mad, afraid, scared, sad). We used the German word translations from the

PANAS-X (Röcke, & Grühm, 2003) except for *mad* and *miserable* for which no German translation was available and which we translated as *wütend* and *niedergeschlagen*. In addition we included a 5-point scale with smileys and asked the children to mark how they had felt during the last two weeks on average.

4.2.7.6. Colour Blindness.

A subset of the Ishihara plates from the Test for Colour Blindness was presented to screen for potential red-green colour-blindness in participants. This was done to ensure that the protagonist in the vignettes used in EmoJump who was was always wearing red, could be discerned as the target for emotion understanding. To ensure that subjects could express their ability to perceive the number even if they did not know it, they were allowed to trace it with a cotton swab.

4.2.8. Data inspection and processing.

Data was inspected for non-normality visually, with z-transformed skew and kurtosis and Shapiro-Wilk tests (see Table G 1 in Appendix G). Analysis including variables for which residuals were found to be clearly non-normally distributed, parametric analysis was followed up by non-parametric analyses (e.g. U-tests of difference scores) to affirm results.

To account for the problem of inflated familywise type I error rates (false positives) in studies with multiple hypotheses tests, a wide range of approaches have been proposed over time ranging from extremely conservative procedures like Bonferroni correction, that greatly inflate the probability of false negatives (type II error) to no correction of significance at all (e.g. Rothman, 1990). Technical procedures aside, it is also well known that significance relies heavily on sample size and per se does not inform about the size of an observed effect. Likewise the presence of a significant effect alone does not allow for inferring that the effect is population-general, for which an adequate sampling process is imperative. Another point to consider is the goal of the study and the possible risk associated with committing type I and type II errors. If a false positive leads to exposing patients to a potentially hazardous substance, the type I error needs to be controlled more tightly than if a study is more explorative in that it seeks to discover possible associations or differences worth pursuing through further research. EmoJump is not a finished training tool yet and not all goals envisioned have been implemented. The present study aims to investigate whether the training game EmoJump holds potential utility and if further development is warranted, not if the program is ready to be implemented as an effective training tool. Thus, a more liberal approach for multiple-comparison correction is taken. Hochberg's step-up procedure (Hochberg & Benjamini, 1990) was used separately for each group of endpoint variables (e.g.

emotion understanding, social cognition, etc.). This approach has been adopted from Tukey, Ciminera, & Heyse (1985) who proposed a (more liberal) multiple-comparison correction procedure to use separately for each *class* of variables (e.g. organ weight, blood serum biochemistry variables) for pharmacological trials of animal toxicity. In the context of the present study, *group of variables* is defined as construct of interest measured by several tasks and/or variables, which can be listed as follows: (1) emotion understanding, (2) social cognition, (3) executive function, (4) affectivity. The error thus is controlled for each construct of interest but not for the experiment as a whole. In other words, the error to wrongly reject H0 for social cognition or any other class of variables is controlled at 5% while the experiment-wide error to get any spurious significant result is not. The author deems this approach in the context of the nature of the study and weighing the possible consequences of false positives and false negatives as acceptable.

Also note that positive effect sizes indicate an increase in performance in the experimental group.

4.3. Results

4.3.1. Descriptives.

Table 31 shows descriptive statistics of the outcome variables, separately for experimental- and control-group and pre-and post-test.

Table 31
Descriptive statistics of study 3

	EG				CG			
	pre		post		pre		post	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Voc (stand)	8.31	2.90			10.08	2.18		
cTRUE: Total	7.38	1.05	7.58	0.93	7.53	1.04	7.32	1.23
cTRUE: Ext	.90	.12	.92	.09	.94	.08	.88	.14
cTRUE: Bel	.83	.16	.83	.16	.83	.19	.90	.16
cTRUE: Mix	.79	.32	.92	.12	.88	.17	.75	.34
DCCS	6.72	2.37	7.49	1.43	6.78	2.79	6.74	1.86
Flanker	8.36	0.90	8.79	0.87	8.52	1.21	8.07	1.46
FASC: MJ	2.62	1.26	2.23	1.17	2.38	1.50	2.23	1.36
FASC: uIST	4.46	3.41	3.77	2.62	5.08	3.12	4.00	2.45
FASC: ISTratio (%)	11.42	6.54	8.45	4.58	8.84	4.24	9.66	8.30
FASC: IRT (s)	3.48	4.12	2.64	1.98	2.06	2.51	1.50	1.81
PANAS PA	19.46	3.69	17.69	4.23	17.15	3.60	18.69	4.13
PANAS NA	9.31	3.23	9.54	2.99	8.46	3.23	6.92	1.55

Note. CG = control group, cTRUE = Computerized Task of Recognizing and Understanding Emotions, EG = experimental group, IRT = initial response time, NA = negative affectivity, PA = positive affectivity, PANAS = Positive and Negative Affect Schedule, Voc = Vocabulary (standard score)

4.3.2. Emotion understanding.

Two participants, one in each group achieved maximum score on the cTRUE at pretest. Because there is only room for a decline in performance, at best with a stable performance the subjects add zero information to the model while at worst, a degradation biases the analysis. Thus these two subjects were removed from the Group x Time mixed ANOVA for cTRUE total score for which sample size thus was 24. For analyses of individual components of emotion understanding, subjects at ceiling could not be excluded because of the high overall performance rate. Thus the sample size of the Group x Time mixed ANOVAS for cTRUE external causes, beliefs and mixed emotions scores was 26. Since components of cTRUE (e.g. external causes, beliefs, mixed emotions) are non-independent from the total score to which they contribute, multiple-comparisons correction was only applied to the analyses of the individual components but not for the total score. To confirm the results of parametric tests, pretest-posttest difference scores were analysed using Mann-Whitney-U tests.

The mixed ANOVA for the cTRUE total score did not yield a significant main effect (see Table 32). More importantly the interaction effect was also non-significant speaking against an overall training effect. Nonparametric analysis of the difference scores confirmed this result (U=61.00, z=-1.21, p=.238, r=-.24).

Table 32
Group x Time mixed ANOVA of cTRUE total score

Source	*Df*	*F*	*d*	*p*
Group	1	0.12	0.14	.738
Time	1	0.06	0.11	.813
Time x Group	1	1.58	0.54	.222
Error (within)	22			

To investigate whether the training had a significant effect on the specific components targeted, additional ANOVAS were conducted (see Table 33).

Table 33
Group x Time mixed ANOVAs of cTRUE external causes, beliefs and mixed emotions scores

		External Causes			Beliefs			Mixed Emotions		
Source	*Df*	*F*	*d*	*p'*	*F*	*d*	*p'*	*F*	*d*	*p'*
Group	1	0.00	0.00	1.000	0.49	-0.29	1.000	0.20	0.18	1.000
Time	1	0.71	-0.35	.816	1.16	0.44	.717	0.00	0.00	1.000
Time x Group	1	1.97	0.57	.293	1.16	-0.44	.293	7.40	1.11	.036
Error (within)	24									

Notes. p' = exact p-values corrected with Hochberg's step-up procedure

Results for external causes and beliefs were non-significant as confirmed with the nonparametric analysis (external causes: *U*=59.00, *z*=1.36, *p'*=.390, *r*=.27; beliefs: *U*=68.00, *z*=-0.95, *p'*=.421, *r*=-.19). The Analysis of mixed emotions however revealed a large and significant interaction effect of Time x Group. The experimental group increased significantly on the mixed emotions score compared to the control group (see Table 33). Non-parametric analysis of difference scores showed a trend for significance (*U*=45.50, *z*=2.36, *p'*=.054, *r*=.46). As evident in the descriptive statistics (see Table 31), the control group's mixed emotions score deteriorated from pre- to posttest (although non-significantly, pre: *Mdn*=1, post: *Mdn*=.88, *z*=−1.63, *p*=.19, *r*=−.33). There is no reason to assume that emotion understanding naturally decreases over time in this sample. A more probable explanation is a motivational deficit at posttest. Although motivation was not directly assessed, affectivity might hint at motivation during the timespan of the study. However, neither negative affectivity (NA), nor positive affectivity (PA) changed significantly from pre- to posttest in the control group as assessed with Wilcoxon signed-rank test (NA: pre: *Mdn*=18, post: *Mdn*=18, *z*=.95, *p*=.359, *r*=.26; PA: pre: *Mdn*=8, post: *Mdn*=7, *z*=1.39, *p*=.359, *r*=0.38). To assess whether the increase in performance in the experimental group would have still been significant if the control group's performance had stayed constant, another ANOVA with modelled data for the control group was conducted. Because this was not a unique hypothesis test but rather an additional validation of the original hypothesis test, no multiple comparison correction was applied. The interaction effect between time and group remained large and significant at trend level (*df*=1, *F*=3.27, *d*=0.74, *p*=.083), suggesting a genuine effect of the experimental group.

4.3.3. Social cognition.

For the ratio of internal state terms to total word count, sample size for the control group was only 12 because one child did not give any answer on the social vignettes. For initial response time, sample size for both groups was 12 because in the experimental group one child did not want its answers recorded on audio so no time was taken while in the control group one child did not provide any answers. As initial response times were strongly positively skewed (Baayen & Milin, 2010) a reciprocal transformation was conducted. Exact p-values were again corrected according to Hochberg's step-up procedure. Neither regarding flexibility (unique mental state justifications, unique mental state terms) nor automaticity of social cognition, any significant Group x Time effects were found (see Table 34). There was a significant group effect on initial response times however, indicating that the control group overall was faster to respond to the vignettes.

Table 34

Group x Time mixed ANOVAs of FASC-variables

		Group			Time			Group x Time		
Variable	*Df*	*F*	*d*	*p'*	*F*	*d*	*p'*	*F*	*d*	*p'*
MSJ	1	0.07	0.11	0.792	0.86	-0.38	0.364	0.16	-0.17	0.719
uIST	1	0.17	-0.17	0.792	2.81	-0.69	0.214	0.13	0.16	0.719
IST-ratio	1	0.88	0.39	0.792	4.60	-0.90	0.172	0.89	-0.39	0.719
IRT	1	7.49	-1.17	0.048	3.14	0.76	0.214	2.99	-0.74	0.392
Error (within)	24									

MSJ = mental justification, uIST = unique internal state terms, IST-ratio = ratio of internal state terms to total word count, IRT = initial response time

p' = exact p-values corrected with Hochberg's step-up procedure

4.3.4. Executive functions.

Residuals of scores of the DCCS and Flanker again proved to be highly non-normal, thus analyses were conducted with nonparametric tests (Mann-Whitney-U) on difference scores only. For the DCCS there was no significant difference in change between groups (EG: *Mdn*=0.12, CG: *Mdn*=-0.25, *U*=68.00, *z*=0.85, *p'*=.418, *r*=.17). For change scores in Flanker however, there was a significant difference (EG: *Mdn*=0.26, CG: *Mdn*=-.24, *U*=33.00, *z*=2.64, *p'*=.014, *r*=.52) with the experimental group showing an increase in performance compared to the control group. Again, descriptive statistics show that the control group decreased in performance on average in the Flanker task. Thus a follow-up analyses with model data (negative change scores replaced with 0 in the experimental and control group) was conducted. This analysis supported the significant change in the experimental vs. control group (*U*=46.00, *z*=2.11, *p*=.036, *r*=.41).

4.3.5. Affectivity.

There was no significant effect on positive or negative affectivity (see Table 35).

Table 35

Group x Time mixed ANOVA of PANAS scores

		Positive Affectivity			Negative Affectivity		
Source	*Df*	*F*	*d*	*p'*	*F*	*d*	*p'*
Group	1	0.00	0.00	1.000	0.49	-0.29	1.000
Time	1	0.71	-0.35	.816	1.16	0.44	.717
Time x Group	1	1.97	-0.57	.293	1.16	-0.44	.293
Error (within)	24						

p': exact p-values corrected with Hochberg's step-up procedure

4.4. Discussion

4.4.1. Emotion Understanding.

In the present study a prototype of a game designed to train emotion understanding in children, was developed and first evidence on effectivity gathered. A large and significant effect on the understanding of mixed emotions was found. The experimental group increased in their understanding of mixed emotions compared to the control group after only 12 training sessions of which only four were targeting mixed emotions. To the best of my knowledge, this is the first time an effect of a computer-game based training on the understanding of mixed emotions has been shown. Even in non-computerized training studies, mixed emotions have seldom been the explicit target of intervention and even more rarely been measured as an outcome (see chapter 4.1.1). In an early study, Peng et al. (1992) demonstrated that prompting to consider a character's emotional reactions can increase acknowledgment of mixed emotions although this effect was measured closely after the experimental manipulation. In another more recent study, explanatory conversations by the child or the experimenter concerning hidden and ambivalent emotions increased children's understanding of emotions as measured with the TEC (Tenenbaum et al., 2008). EmoJump only includes a short tutorial video for level 3 which explicitly states: "In this level you will see that one can experience multiple emotions concurrently in certain situations. Therefore your objective is to collect two faces at the same time". This prompt is followed by an example and gameplay instructions. Feedback given for collected emotion-coins at the end of a scene is ambiguous (i.e. only total number of correct and incorrect coins) and no explanations are given. Thus, the amount of information and feedback provided lies between that given in studies by Peng et al. (1992) and Tenenbaum et al. (2008). Facial emotion recognition skills have been shown to be improvable through unsupervised learning (Huelle et al. 2014). While learning in EmoJump is not wholly unsupervised, there is no feedback loop for each response, only in a condensed form. Being left to guess which responses to the emotion vignette were right or wrong, might elicit trains of thought why the character in the cartoons is supposed to feel this or that way and intensify the process of elaboration. This might have an effect similar to the deliberate elaboration promoted in the Tenenbaum et al. (2008) study. To inform on this explanation further studies could use the method of verbalizing thoughts. To unravel mechanisms of action, experiments with slightly altered versions of the game for each condition could be carried out. To single out the effect of prompts, one group might play the game with and another group without the prompts in the video tutorial. Alternatively to learn about effectiveness of different forms of feedback, at one extreme, the player could be told immediately after collecting an emotion-coin whether the choice was correct or not while at the other end of the spectrum there might be no feedback at all.

There was no significant effect on overall emotion understanding, understanding of external causes or understanding of belief based emotions. Despite non-significant results, supposedly in part due to the small sample size, following observations and cautious interpretations can be made: Most subjects performed at perfect or near-perfect level regarding the component of external causes which severely limited the room for improvement. Thus at present it cannot be stated whether EmoJump has an effect on the understanding of external causes of emotions. In a follow-up evaluation, a younger sample and more sensitive measures should be used to investigate whether the level for external-causes should remain in the game. Understanding of belief-based emotions did not improve and there was even a non-significant improvement for the control group. This however is owed mainly to the marked improvement of one subject in the control group and thus indicates most likely a spurious result. Interestingly in the recent meta-analysis of training studies of emotion understanding, mental aspects of emotion understanding showed the smallest effects size compared to a moderate effect for external and reflective aspects (Sprung et al. 2015). This might indicate that mental emotion understanding is harder to improve than other aspects. Alternatively it may mean that adequate mechanisms of action have not yet been identified. Another possibility is that the concept for the training level targeting belief-based emotions was ill designed. As explained in the description of EmoJump above, subjects had to attribute diverging emotions to the latter scene in a comic strip and subsequently to the former (having the knowledge of what happens next). This was intended to vaguely emulate a "smarties-task" in which the child is asked how she (or some puppet) felt before opening the container and revealing the unexpected content. Similarly in the TEC the fox is still lurking behind the bush (even if hidden inside this variation of a container). In EmoJump, the presence of the latter emotion eliciting situation is less salient (a smaller, greyed out social vignette) to emphasize the current vignette for comprehensibility. This might have led to subjects ”solving” the scenes in terms of external causes, basically ignoring or forgetting about subsequent events in the narrative.

4.4.2. Social Cognition.

We did not find a treatment effect of EmoJump on social cognition as measured with the FASC, neither for flexibility nor for automaticity. There are several possible explanation for this finding. Firstly, while the training game does not require players to think flexibly on mental states and find justifications, FASC does. It is possible that the procedure of EmoJump is not suitable to improve elaborate thinking about mental states but rather implicit understanding of mixed emotions. There might also have been motivational problems with FASC. Several children stated their disapproval or expressed boredom when confronted with the social vignettes at post-test. This is supported by a marked decrease in response length across groups. Lastly, the aspect of social cognition assessed with FASC might be independent from emotion understanding as measured by cTRUE. This explanation is supported by study two which found FASC variables and cTRUE and TEC to be more or less independent. Another study already mentioned in chapter 3.4.3.2 found performance on the TEC and number of affective mental state terms in a social vignette interview to be unrelated (Castro, Halberstadt, & Garrett-Peters, 2016).

4.4.3. Executive Functions.

In the realm of executive functions there was a significant change of rank for the experimental group compared to the control group for cognitive inhibition /selective attention, indicative of a training effect. Gameplay in EmoJump requires to focus attention on a stimulus (i.e. the correct emotion-coin to collect) and inhibit responding to competing stimuli (i.e. incorrect emotion-coins appearing). Several other studies have found effects of computer game playing on visual attention (Green & Bavelier, 2003; Green & Bavelier, 2006). An fMRI study showed greater recruitment of a fronto-parietal network in response to increasing attentional demands by distractors in non-gamers compared to gamers, suggesting a more automatic or efficient filtering in gamers (Bavelier, Achtman, Mani, & Focker, 2012). The effect of action video games on attention has also been shown to translate into better reading skills in dyslexic children (Franceschini et al., 2013). There was no significant difference for cognitive flexibility. This result is in line with the supposed demands the gameplay puts on the player, since stimulus-response mappings do not change frequently in EmoJump. Interestingly, in a recent meta-analysis on effects of video-game play on information processing, in the domain of executive functions, inhibition was the only sub-skill for which a significant improvement among true experimental studies emerged (Powers, Brooks, Aldrich, Palladino, & Alfieri, 2013).

4.4.4. Study limitations & recommendations.

The large number of children who already scored at maximum in the emotion understanding components at pretest (e.g. mixed emotions) reduced the sensitivity to possible treatment effects. In the treatment group 8 of 13 subjects correctly answered all four mixed emotion vignettes at pretest. Importantly of the five remaining, four increased in their mixed emotions understanding score. Presumably, the treatment effect would have been even more pronounced in a younger sample. Screening for emotion understanding and only including children who have not yet (fully) achieved understanding should reduce ceiling effects and increase sensitivity to detect treatment effects. A larger follow-up study needs to be done, to replicate the effects found. Furthermore, several diverse and established tasks of emotional competence should be included to improve validity of findings. Including measures of social functioning and a follow-up assessment several months after intervention is recommended to inform on the utility of the training game and stability of effects.

We did not collect ratings on motivation but overall it can be stated that in today's saturated media-landscape, it is hard for non-professional games to compete. I do not expect EmoJump to be preferred to commercial games by children. In the context of a psychosocial intervention however it might hold some incentive as an additional offer and maybe help lower barriers between therapist and child. Nevertheless, the game would greatly benefit from a revision by a team of experienced game developers. Today with game development software being available cheaply or even for free, it is easier than ever for small teams or even individuals to develop enjoyable games as the lively independent-games scene demonstrates. Still this requires a level of money and/or time and expertise that was not available in the context of this work. Such a professional realization of a training game might even make it to a point where the game can be handed out to individuals with deficits in emotion understanding and be played for the sake of playing.

4.4.5. Implications for further development of EmoJump.

Several implications for the continued development of EmoJump can be drawn from the results of this pilot study. First, the absence of training effects at level 1 and 2 can be seen as indicative that they are more appropriate for younger children or children with a marked delay in emotion understanding. A follow up study should either include a younger group of normally developing children or a large sample of children with different degrees of impairment in emotion understanding (e.g. different grades of severity in the spectrum of autism) to match the three levels of emotion understanding (i.e. external, mental, reflective) and inform on the optimal target group for the game. Secondly, the number of available vignettes (i.e. scenes per sublevel) should be increased to encompass an even wider array of potentially emotion eliciting situations and increase play-length. This increase in stimulus material could also be achieved by switching perspective between characters in a scene for which emotion states are to be attributed. Having a large pool of vignettes, which will clearly be more appropriate for one age or another, will make it feasible to add a gameplay-option to adjust scene selection to fit the player's age. Thirdly, criteria for success in sub-levels were set very liberally to consider younger participants and avoid demotivating effects. A next version of EmoJump should adapt to the player's performance furthering motivation and potentially optimizing training effects. Finally, as discussed in the context of the study about the Computerized Task of Recognizing and Understanding Emotions, the explicit inclusion of more complex emotions (e.g. jealousy, envy, embarrassment) might be a worthwhile endeavour to increase complexity of this training game and make it more suitable for older target populations.

4.4.6. Conclusions.

The present results can be seen as tender first evidence, supporting the continuation of development of EmoJump. A large training effect for the understanding of mixed emotions was found. Additionally, subjects playing the game improved in a standard task of inhibition / selective attention. Future evaluation studies should include a diverse range of emotion understanding tasks to further establish evidence for EmoJump's efficacy.

5. General Discussion

Social cognition and emotion understanding are critical capacities of the human mind for leading a happy and participative life in the context of social groups. Traditionally the study of these mental faculties has focussed on preschool children or clinical populations. The present work was aimed at normally developing primary school age children. Study one explored socio-cognitive reasoning across the lifespan with a new procedure developed by Hayward et al. (2016). The FASC demonstrated developmental differences between children, adolescents, adults and older adults not as clearly seen in studies utilizing established measures of advanced theory of mind utilizing normative scoring schemata (as reviewed in chapter 2.1.1.8.1). The FASC also allows investigating the effect of language and ambiguity on mental state reasoning. Verbal cues and ambiguity were shown to facilitate flexible and rich mental state attributions. Evaluating interventions targeting multiple aspects of emotion understanding requires sensitive and comprehensive measures. The cTRUE expands upon existing measures and partially fulfils this demand as shown in study two. More specifically cTRUE assesses understanding of mixed emotions and cognitive emotion regulation quite reliably and is related to academic competence and pro-social role behaviour. There is more work to be done concerning the other components of emotion understanding before it can be used in primary school age children however. Study three described the development and evaluation of EmoJump, a new computer game intervention to improve external, mental and reflective aspects of emotion understanding. In a randomized controlled trial it was shown to improve understanding of mixed emotions.

Returning to the overarching theme of global burden of mental health problems and the task-shifting approach of the WHO, how do computerized measures and training games like cTRUE and EmoJump fit in? First children or adolescents that may benefit from emotion understanding promotion have to be identified. To maximise coverage, this would ideally take place in the context of school readiness tests or mother-child-health-pass examinations. In countries lacking these implementations, mental health care workers or teachers might conduct impact workshops in disadvantaged neighbourhoods about emotional competency, similar to sex education initiatives in rural Africa. These might directly target children in schools or youth centres or shift tasks through instructing other teachers or community workers in the area. After the initial workshops each child receives internet access to the training game, either directly or through the corresponding teacher or community worker. Motivation to adhere might be enhanced through (virtual) token systems. Ideally, playing the training game would be accompanied by workshops discussing different attributions of mental states that participants have made to the characters in the social vignettes. This could be the starting point to discuss in which situations the participants themselves have experienced similar feelings, how they can respond to them, etc. At the “internet for everyone” campus Vienna (https://a1internetfueralle.at) by the telecommunication provider A1 for example, children between 7 and 14 can learn safe and competent internet use with the help of media-coaches and take part in media-pedagogic workshops. Many children from disadvantaged socio-economic backgrounds go there because they have no computer- or internet-access at home and children are allowed to play selective games under supervision of the coaches. We, the clinical child- and adolescent-psychology work group, collaborated with this initiative on several occasions for example to play-test serious games. Such a setting would be the ideal setting to implement the combination of computerized emotion comprehension promotion and complementary workshops. One activity in such workshops may be to let children create new vignettes themselves. Constructivist learning (Piaget & Inhelder, 1969; Vygotsky, 1978) is thought to improve the learner’s engagement and deepen his knowledge of the subject, ground it in his experience and transfer it to his every-day context.

Another field of application would be of clinical child- and adolescent-psychologists. Play in this setting has long been used to break the ice, stimulate a discussion about a topic or practice certain skills. Computer game-based interventions have the benefit of delivering a much larger number of stimuli in a (semi-)automated fashion compared to a board game, for instance. Running on a computer does not mean that it has to be a solitary activity though, as therapist and client can jointly discuss the player’s decisions or take turns playing (thus enabling observational learning). More interestingly for the therapeutic setting even would be the development of a cooperative game to promote emotion understanding. Therapist and client might have to solve emotion understanding puzzles with each player only having part of the critical information (e.g. character’s mental states). On a meta-level a therapist might address emotions elicited by a computer game (e.g. frustration, joy, sadness, anxiety, etc.; see chapter 1.4 for a short discussion of emotions in video games), how they originated and how to deal with them. This is a less threatening way for children and adolescents to talk about emotions than directly addressing emotion associated with personal live events.

Another important topic cutting across computer use and emotion understanding is that of pro-social behaviour and empathizing in social media. A multitude of factors like in-group/out-group effects, lack of emotional feedback cues (e.g. facial) or perceived anonymity facilitate phenomena like cyber-bullying, cyber-shaming and ostracising. Arguably this behaviour is not (only) a consequence of emotion understanding deficits but of other factors like the motivation to empathize. If I dehumanize someone by allocating him to an out-group and labelling him as morally inferior, I will not try to "walk in her shoes" even though I may be very empathizing with my friends and family. Still, computerized emotion understanding interventions could incorporate this issue by using social media examples as one type of vignettes raising awareness about feelings caused by online communication. An often cited argument for these anti-social behaviours is, that there is limited acknowledgment that behind the username sits a real human being with feelings. This behaviour resembles a behaviour Antonio Damasio (1994) observed in patients with damage in the ventro-medial frontal cortex he dubbed "to know but not to feel" (p. 205). Watching disturbing images, these patients, although knowing that these situations would elicit fear or sadness, did not produce a skin conductance response – indicative of sympathetic nervous system activity – in contrast to the control subjects and also reported reduced subjective feelings. In further experiments he investigated that these patients fared poorly in a decision making task (the Iowa gambling task). In contrast to the normal controls, they more often chose card decks that frequently yielded higher wins compared to other decks, but occasionally yielded very high losses and in the long run were the worse choice. Additionally, contrary to control subjects, over the course of the game, they did not produce an anticipatory skin conductance response when deliberating to choose from the bad deck. He explained these results in terms of the somatic marker hypothesis. In short, it states that we learn to associate emotions and the accompanying bodily changes (i.e. somatic markers) with situations and their outcomes. When deciding between different courses of action, these somatic markers are activated and favour some decisions and their anticipated outcomes over others. Next to this supposed mechanism, Damasio also acknowledged the possible influence of imagining future or counter-factual scenarios. Returning to the topic of anti-social online commentaries, it is as if, some people act as if their actions have no consequences. And in a very real sense this is exactly the case. If we write something scathing, we do not see and feel the other person being hurt. Or in terms of the somatic marker hypothesis, writing something hurtful to an unseen or even unknown other as opposed to saying it to his face does not activate the same bodily feelings. Receiving insulting comments on the other hand does indeed hurt as a bullying survey shows (Ditch the Label, 2017). The ability (and motivation!) to imagine the other person as a real human with positive and negative traits, social ties and feelings, for instance by imagining someone you know being in the other persons place may help activate somatic markers and facilitate empathy.

Different game genres might lend themselves more readily to the promotion of different aspects of emotion understanding. One of my early concepts ("Emotion Detective", presented at the Future and Reality of Gaming conference in Vienna, 08.10.2012) was that of a "point and click" adventure game in which the child plays the role of a detective solving age-appropriate cases through mental state attribution and -reasoning. Detective stories are rich of mental states such as desires, motives, (false) beliefs, hiding emotions, etc. The player would have to rely on different components of emotion understanding to solve the case. Facial expressions might point out persons worth questioning or hint at underlying desires. Conversely, emotions can be hidden and might be deduced through other means (e.g. getting to know someone's desires and motives). In a traditional video-game custom (e.g. quests) the player might have to identify and satisfy non-player characters needs and desires first (e.g. "wants his toy back", "needs someone to accompany him because he is afraid of something") to drive the story forward. Discerning what a non-player character can know under given circumstances (e.g. false belief) might be paramount to identify suspects or witnesses and avoid being side-tracked. A witness might have to be soothed first (i.e. emotion regulation) before he can be questioned. Such a game necessarily relies on a lot of language so that – barring full voice acting – reading skills would be heavily taxed, restricting the lower limit of appropriate age. On the other hand such a game might have the additional benefit of enhancing the child's reading skills. Furthermore, reading fiction has been suggested to promote mentalizing abilities as well (Mar, Oatley, Hirsh, dela Paz, & Peterson, 2006; Kidd & Castano, 2013) so a text-heavy game similar to an interactive novel might be particularly beneficial.

Computer games are a reality and are not going to go away. Instead of demonizing them, they should be better understood and embraced as a force for potential good. On the other hand, indiscriminate euphoria that sees game based interventions as a panacea for every mental health problem is equally uncalled for. One can look at the literature on computerized cognitive training that, despite it is comparably long history, is still full of controversy (e.g. see Simons et al., 2016, Mewborn, Lindbergh, & Miller, 2017, and Harvey, McGurk, Mahncke, & Wykes, 2018 for recent reviews / meta analyses). On October 20, 2014, the Max Planck Institute for Human Development and Stanford Centre on Longevity released a consensus statement signed by 75 psychologists calling the efficacy of brain training into question. As a response, a group of 133 researchers and therapists signed a letter in response agreeing on several points but rejecting the overall assertion that there was no compelling scientific evidence that brain exercise offers a way to reduce or reverse cognitive decline. This debate seems in part intensified by mutual misunderstanding, generalisations and personal motivations underlying the discussion as it is often the case when making broad statements addressing the general public. Looking at the same evidence from two different viewpoints the one group seems apprehensive of raising false hopes and choosing the wrong focus point in regard to the prevention of cognitive decline while the other group worries about the stifling of promising research. For example, while focusing on brain training for age-related cognitive decline, the critical statement about the lack of scientific evidence does not exclude cognitive remediation in acquired brain injury for which the evidence seems more conclusive than for prevention of age-related cognitive decline. As is typical for a research area outgrowing its infancy, global binary questions give way to more differentiated research agendas concerning which cognitive function in which population under which training regime can be improved and at which point a change can be considered clinically useful. Research on computerized trainings to improve emotion understanding will probably take a similar road.

References:

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA School-Age Forms & Profiles.* Burlington: University of Vermont. Research Center for Children, Youth, and Families.

Achenbach, T. M., & Rescorla, L. A. (2003). *Manual for the ASEBA Adult Forms & Profiles.* Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Achenbach, T. M., Newhouse, P. A., & Rescorla, L. A. (2004). *Manual for the ASEBA Older Adults Forms & Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Ahmed, F. S., & Miller, L. (2011). Executive function mechanisms of theory of mind. *Journal of Autism and Developmental Disorders, 41*(5), 667-678.

Albanese, O., De Stasio, S., Di Chiacchio, C., Fiorilli, C., & Pons, F. (2010). Emotion comprehension: the impact of nonverbal intelligence. *Journal of Genetic Psychology, 171*(2), 101-115. doi: 10.1080/00221320903548084

Alexander, W. H., & Brown, J. W. (2011). Medial prefrontal cortex as an action-outcome predictor. *Nature Neuroscience, 14*(10), 1338-1344.

Altshuler, J. L., & Ruble, D. N. (1989). Developmental changes in children's awareness of strategies for coping with uncontrollable stress. *Child Development, 60*(6), 1337-1349. doi:http://dx.doi.org/10.2307/1130925

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Anderson, C. A., Shibuya, A., Ihori, N., Swing, E. L., Bushman, B. J., Sakamoto, A., . . . Saleem, M. (2010). Violent video game effects on aggression, empathy, and prosocial behavior in Eastern and Western countries: A meta-analytic review. *Psychological Bulletin, 136*(2), 151-173. doi:http://dx.doi.org/10.1037/a0018251

Anderson, V. A., Anderson, P., Northam, E., Jacobs, R., & Catroppa, C. (2001). Development of executive functions through late childhood and adolescence in an Australian sample. *Developmental Neuropsychology, 20*(1), 385-406. doi:10.1207/S15326942DN2001_5

Andrews, G., Cuijpers, P., Craske, M. G., McEvoy, P., & Titov, N. (2010). Computer Therapy for the Anxiety and Depressive Disorders Is Effective, Acceptable and Practical Health Care: A Meta-Analysis. *PLoS ONE, 5*(10), e13196. doi:10.1371/journal.pone.0013196

Anguera, J. A., Boccanfuso, J., Rintoul, J. L., Al-Hashimi, O., Faraji, F., Janowich, J., . . . Gazzaley, A. (2013). Video game training enhances cognitive control in older adults. *Nature, 501*(7465), 97-101. doi: 10.1038/nature12486

Apperly, I. A. (2008). Beyond Simulation-Theory and Theory-Theory: why social cognitive neuroscience should use its own concepts to study "theory of mind". *Cognition, 107*(1), 266-283. doi:10.1016/j.cognition.2007.07.019

Apperly, I.A. (2011). *Mindreaders. The cognitive basis of „theory of mind“*. New York, NY: Psychology Press.

Apperly, I. A., Samson, D., & Humphreys, G. W. (2009). Studies of adults can inform accounts of theory of mind development. *Developmental Psychology, 45*(1), 190-201. doi: 10.1037/a0014098

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review, 116*(4), 953-970. doi:10.1037/a0016923

Apperly, I. A., Warren, F., Andrews, B. J., Grant, J., & Todd, S. (2011). Developmental continuity in theory of mind: speed and accuracy of belief-desire reasoning in children and adults. *Child Development, 82*(5), 1691-1703. doi:10.1111/j.1467-8624.2011.01635.x

Arbeitsgruppe Deutsche Child Behavior Checklist. (1993). *Lehrerfragebogen über das Verhalten von Kindern und Jugendlichen; deutsche Bearbeitung der Teacher's Report Form (TRF) der Child Behavior Checklist. Einführung und Anleitung zur Handauswertung, bearbeitet von M. Döpfner & P. Melchers.* Köln: Arbeitsgruppe Kinder-, Jugend- und Familiendiagnostik.

Arbeitsgruppe Deutsche Child Behavior Checklist. (1998). *Elternfragebogen über das Verhalten von Kindern und Jugendlichen; deutsche Bearbeitung der Child Behavior Checklist (CBCL/4-18). Einführung und Anleitung zur Handauswertung. 2. Auflage mit deutschen Normen, bearbeitet von M. Döpfner, J. Plück, S. Bölte, K. Lenz, P. Melchers & K. Heim.* Köln: Arbeitsgruppe Kinder-, Jugend- und Familiendiagnostik.

Arbeitsgruppe Deutsche Child Behavior Checklist. (2009). *Fragebogen über das Verhalten Erwachsener (ABCL/18-59): Deutschsprachige Fassung des US-Originals von 2003.* Arbeitsgruppe Kinder-,Jugendlichen- und Familiendiagnostik (KJFD). Köln.

Arbeitsgruppe Deutsche Child Behavior Checklist. (2009). *Fragebogen für Erwachsene (ASR/18-59): Deutschsprachige Fassung des US-Originals von 2003.* Köln: Arbeitsgruppe Kinder-,Jugendlichen- und Familiendiagnostik.

Armitage, S. G. (1946). An analysis of certain psychological tests used for the evaluation of brain injury. *Psychological Monographs, 60*(1), i-48.

Army Individual Test Battery Manual of Directions and Scoring (1944). Washington DC: U.S. War Department, Adjutant General's Office.

Arsenio, W. F., & Kramer, R. (1992). Victimizers and their victims: Children's conceptions of the mixed emotional consequences of moral transgressions. *Child Development, 63*(4), 915-927. doi:http://dx.doi.org/10.2307/1131243

Aschersleben, G., Hofer, T., & Jovanovic, B. (2008). The link between infant attention to goal-directed action and later theory of mind abilities. *Developmental Science, 11*(6), 862-868.

Aster, M. V., Neubauer, A., & Horn, R. (Eds.). (2006). *Wechsler Intelligenztest für Erwachsene*. Frankfurt am Main: Pearson Assessment.

Astington, J. W. (2003). Sometimes necessary, never sufficient: False-belief understanding and social competence. In B. Repacholi & V. Slaughter [Eds.], *Individual differences in theory of mind: Implications for typical and atypical development* (pp 13-38). New York, NY: Psychology Press

Astington, J. W., & Jenkins, J. M. (1995). Theory of mind development and social understanding. *Cognition and Emotion, 9*(2-3), 151-165. doi:http://dx.doi.org/10.1080/02699939508409006

Austin, A. B., & Draper, D. C. (1984). The relationship among peer acceptance, social impact, and academic achievement in middle childhood. *American Educational Research Journal, 21*(3), 597-604.

Baayen, R. H., & Milin, P. (2010). Analyzing Reaction Times. *International Journal of Psychological Research, 3*(2).

Backhaus, A., Agha, Z., Maglione, M. L., Repp, A., Ross, B., Zuest, D., . . . Thorp, S. R. (2012). Videoconferencing psychotherapy: A systematic review. *Psychological Services, 9*(2), 111-131. doi:http://dx.doi.org/10.1037/a0027924

Baddeley, A. D. (1986). *Working Memory*. Oxford: Oxford University Press.

Baillargeon, R., Scott, R. M., & Bian, L. (2016). Psychological reasoning in infancy. *Annual Review of Psychology, 67*, 159-186.

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences, 14*(3), 110-118. doi:http://dx.doi.org/10.1016/j.tics.2009.12.006

Banerjee, R. (2000). The development of an understanding of modesty. British Journal of Developmental Psychology, 18(4), 499-517.

Banerjee, R., Watling, D., & Caputi, M. (2011). Peer relations and the understanding of faux pas: Longitudinal evidence for bidirectional associations. *Child Development, 82*(6), 1887-1905. doi:http://dx.doi.org/10.1111/j.1467-8624.2011.01669.x

Barlassina, L. & Gordon, R. M. (2017). Folk Psychology as Mental Simulation. In *The Stanford Encyclopedia of* Philosophy. Retrieved from https://plato.stanford.edu/archives/sum2017/entries/folkpsych-simulation.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21*(1), 37-46. doi: http://dx.doi.org/10.1016/0010-0277(85)90022-8

Baron-Cohen, S. (1995). Mindblindness: An essay on autism and theory of mind. *(1995) Mindblindness: An essay on autism and theory of mind xxii, 171 pp Cambridge, MA, US: The MIT Press; US.*

Baron-Cohen, S., O'Riodan, M., Stone, V., Jones, R., & Plaisted, K. (1999). Recognition of Faux Pas by Normally Developing Children and Children with Asperger Syndrome or High-Functioning Autism. *Journal of Autism and Developmental Disorders, 29*(5), 407-418.

Baron-Cohen, S. (2001). Theory of mind and autism: A review. In G. D. Masters [Ed], *International review of research in mental retardation: Autism (vol 23)* (pp 169-184). San Diego, CA: Academic Press

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

Baron-Cohen, S., Wheelwright, S., Spong, A., Scahill, V., & Lawson, J. (2001). Are Intuitive Physics and Intuitive Psychology Independent? A Test with Children with Asperger Syndrome. *Journal of Developmental and Learning Disorders, 5*(1), 47-78.

Barrett, L. F. (2006a). Are emotions natural kinds? Perspectives on Psychological Science, 1(1), 28-58.

Barrett, L. F. (2006b). Solving the Emotion Paradox: Categorization and the Experience of Emotion. *Personality and Social Psychology Review, 10*(1), 20-46. doi:10.1207/s15327957pspr1001_2

Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience, 12*(1), 1-23.

Bartsch, K. (2002). The role of experience in children's developing folk epistemology: Review and analysis from the theory-theory perspective. *New Ideas in Psychology, 20*(2-3), 145-161.

Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind*. New York, NY: Oxford University Press; US.

Bavelier, D., Achtman, R. L., Mani, M., & Focker, J. (2012). Neural bases of selective attention in action video game players. *Vision Res, 61*, 132-143. doi: 10.1016/j.visres.2011.08.007

Beck, L., Kumschick, I. R., Eid, M., & Klann-Delius, G. (2012). Relationship between language competence and emotional competence in middle childhood. *Emotion, 12*(3), 503-514. doi:10.1037/a0026320

Begeer, Sander, Gevers, Carolien, Clifford, Pamela, Verhoeve, Manja, Kat, Kirstin, Hoddenbach, Elske, & Boer, Frits. (2011). Theory of mind training in children with autism: A randomized controlled trial. *Journal of Autism and Developmental Disorders, 41*(8), 997-1006.

Belacchi, C., & Farina, E. (2010). Prosocial/hostile roles and emotion comprehension in preschoolers. *Aggressive Behavior, 36*(6), 371-389. doi:10.1002/ab.20361

Bender, P. K., Pons, F., Harris, P. L., & de Rosnay, M. (2011). Do young children misunderstand their own emotions? *European Journal of Developmental Psychology, 8*(3), 331-348. doi:10.1080/17405629.2010.495615

Berking, M., & Wupperman, P. (2012). Emotion regulation and mental health: Recent finding, current challenges, and future directions. *Current Opinion in Psychiatry, 25*(2), 128-134. doi:http://dx.doi.org/10.1097/YCO.0b013e3283503669

Bernstein, D. M., Thornton, W. L., & Sommerville, J. A. (2011). Theory of mind through the ages: older and middle-aged adults exhibit more errors than do younger adults on a continuous false belief task. *Experimental Aging Research, 37*(5), 481-502. doi:10.1080/0361073X.2011.619466

Bernstein, D. M., Atance, C., Loftus, G. R., & Meltzoff, A. (2004). We saw it all along: Visual hindsight bias in children and adults. *Psychological Science, 15*(4), 264-267. doi:http://dx.doi.org/10.1111/j.0963-7214.2004.00663.x

Bernstein, D. M., Erdfelder, E., Meltzoff, A. N., Peria, W., & Loftus, G. R. (2011). Hindsight bias from 3 to 95 years of age. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(2), 378-391. doi:http://dx.doi.org/10.1037/a0021971

Bernstein, J. (2013, February 14). Please Play This Free, Gorgeous 3-Minute Long Game About Being A Fly. *BuzzFeedNews*. Retrieved from https://www.buzzfeednews.com/article/josephbernstein/please-play-this-free-gorgeous-3-minute-long-game-about-bein

Birch, S. A., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science, 18*(5), 382-386.

Birch, S. H., & Ladd, G. W. (1997). The teacher-child relationship and children's early school adjustment. *Journal of School Psychology, 35*(1), 61-79.

Bisoglio, J., Michaels, T. I., Mervis, J. E., & Ashinoff, B. K. (2014). Cognitive Enhancement Through Action Video Game Training: Great Expectations Require Greater Evidence. *Frontiers in Psychology, 5*. doi:10.3389/fpsyg.2014.00136

Blair, C., & Razza, R. P. (2007). Relating Effortful Control, Executive Function, and False Belief Understanding to Emerging Math and Literacy Ability in Kindergarten. *Child Development, 78*(2), 647-663.

Bloom, P., & German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition, 77*(1), B25-B31.

Bock, A. M., Gallaway, K. C., & Hund, A. M. (2015). Specifying Links Between Executive Functioning and Theory of Mind during Middle Childhood: Cognitive Flexibility Predicts Social Understanding. *Journal of Cognition and Development, 16*(3), 509-521. doi:10.1080/15248372.2014.888350

Bolte, S., Feineis-Matthews, S., Leber, S., Dierks, T., Hubl, D., & Poustka, F. (2002). The development and evaluation of a computer-based program to test and to teach the recognition of facial affect. *International Journal of Circumpolar Health, 61 Suppl 2*, 61-68.

Borke, H. (1971). Interpersonal perception of young children: Egocentrism or empathy? *Developmental Psychology, 5*(2), 263-269. doi:http://dx.doi.org/10.1037/h0031267

Bosacki, S., & Astington, J. W. (1999). Theory of mind in preadolescence: Relations between social understanding and social competence. *Social Development, 8*(2), 237-255.

Bosacki, S. L. (2000). Theory of mind and self-concept in preadolescents: Links with gender and language. *Journal of Educational Psychology*, 92(4), 709-717. doi:10.1037/0022-0663.92.4.709

Botting, N., & Conti-Ramsden, G. (2008). The role of language, social cognition, and social skill in the functional social outcomes of young adolescents with and without a history of SLI. *British Journal of Developmental Psychology, 26*(2), 281-300.

Bracy, O., Lynch, W., Sbordone, R., & Berrol, S. (1985). Cognitive retraining through computers: Fact or fad? *Cognitive Rehabilitation, 3*(2), 10-25.

Bradmetz, J., & Schneider, R. (1999). Is Little Red Riding Hood afraid of her granmdmother? Cognitive vs. emotional response to a false belief. *British Journal of Developmental Psychology, 17*(4), 501-514. doi:http://dx.doi.org/10.1348/026151099165438

Brandone, A. C., & Wellman, H. M. (2009). You can't always get what you want: Infants understand failed goal-directed actions. *Psychological Science, 20*(1), 85-91.

Brbsmoking. (2015, October 19). Answer to "Why Game of Thrones is not very good". *Telltale Games Community General Discussion Forum for Game of Thrones.* Retrieved from https://www.telltalegames.com/community/discussion/99766/why-game-of-thrones-is-not-very-good

Brent, E., Rios, P., Happé, F., & Charman, T. (2004). Performance of children with autism spectrum disorder on advanced theory of mind tasks. Autism, 8(3), 283-299.

Bretherton, I., & Beeghly, M. (1982). Talking about internal states: The acquisition of an explicit theory of mind. *Developmental Psychology*, 18(6), 906-921.

Brocki, K. C., & Bohlin, G. (2004). Executive functions in children aged 6 to 13: a dimensional and developmental study. *Developmental Neuropsychology, 26*(2), 571-593. doi:10.1207/s15326942dn2602_3

Brown, M., & Forsythe, A. (1974). Robust Tests for the Equality of Variances. *Journal of the American Statistical Association, 69*(346), 364-367. doi:10.2307/2285659

Burdea, G. C. (2003). Virtual Rehabilitation - Benefits and Challenges. *Methods Archive, 42*(5), 519-523.

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition, 112*(2), 337-342.

Buttelmann, D., Over, H., Carpenter, M., & Tomasello, M. (2014). Eighteen-month-olds understand false beliefs in an unexpected-contents task. *Journal of Experimental Child Psychology, 119*, 120-126.

Byrne, R. W., & Whiten, A. (1988). Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes, and humans. In *Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes, and humans* (pp. xiv, 413). New York, NY: Clarendon Press/Oxford University Press; US.

Caillois, R. (1961). *Man, Play, and Games*. Urbana-Champaign, IL: University of Illinois Press.

Camerer, C., Loewenstein, G., Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *The Journal of Political Economy, 97*, 1232–1254

Campos, J. J., Thein, S., & Owen, D. (2003). A Darwinian Legacy to Understanding Human Infancy: Emotional Expressions as Behavior Regulators. In *Emotions inside out: 130 years after Darwin's: The expression of the emotions in man and animals* (pp. 110-134). New York, NY: New York Academy of Sciences; US.

Caputi, M., Lecce, S., Pagnin, A., & Banerjee, R. (2012). Longitudinal effects of theory of mind on later peer relations: The role of prosocial behavior. *Developmental Psychology, 48*(1), 257-270.

Carlson, S. M., Mandell, D. J., & Williams, L. (2004). Executive function and theory of mind: stability and prediction from ages 2 to 3. *Developmental Psychology, 40*(6), 1105-1122. doi:10.1037/0012-1649.40.6.1105

Carlson, S. M., & Wang, T. S. (2007). Inhibitory control and emotion regulation in preschool children. *Cognitive Development, 22*(4), 489-510.

Caron, A. J. (2009). Comprehension of the representational mind in infancy. *Developmental Review, 29*(2), 69-95.

Carpendale, J. I., & Chandler, M. J. (1996). On the Distinction between False Belief Understanding and Subscribing to an Interpretive Theory of Mind. *Child Development, 67*(4), 1686-1706. doi:10.2307/1131725

Carpendale, J. I., & Lewis, C. (2004). Constructing an understanding of mind: The development of children's social understanding within social interaction. *Behavioral and Brain Sciences, 27*(1), 79-151.

Carpenter, J. M., Green, M. C., & Vacharkulksemsuk, T. (2016). Beyond perspective-taking: Mind-reading motivation. *Motivation and Emotion, 40*(3), 358-374.

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63*(4), 176.

Carrington, S. J., & Bailey, A. J. (2009). Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Hum Brain Mapp, 30*(8), 2313-2335. doi:10.1002/hbm.20671

Carruthers, P. (2016). Two Systems for Mindreading? *Review of Philosophy and Psychology, 7*(1), 141-162. doi:10.1007/s13164-015-0259-y

Casey, R. J., & Schlosser, S. (1994). Emotional responses to peer praise in children with and without a diagnosed externalizing disorder. *Merrill-Palmer Quarterly, 40*(1), 60-81.

Cassidy, J., Parke, R. D., Butkovsky, L., & Braungart, J. M. (1992). Family-peer connections: the roles of emotional expressiveness within the family and children's understanding of emotions. *Child Development, 63*(3), 603-618.

Cassidy, K. W., Werner, R. S., Rourke, M., Zubernis, L. S., & Balaraman, G. (2003). The Relationship Between Psychological Understanding and Positive Social Behaviors. *Social Development, 12*(2), 198-221.

Castro, V. L., Halberstadt, A. G., & Garrett-Peters, P. (2016). A three-factor structure of emotion understanding in third-grade children. *Social Development, 25*(3), 602-622. doi:http://dx.doi.org/10.1111/sode.12162

Cepeda, N. J., Kramer, A. F., & Gonzalez de Sather, J. C. (2001). Changes in executive control across the life span: examination of task-switching performance. *Developmental Psychology, 37*(5), 715-730.

Ceranoglu, T. (2010). Video games in psychotherapy. *Review of General Psychology, 14*(2), 141-146.

Cheng, S. K., & Dizon, J. (2012). Computerised cognitive behavioural therapy for insomnia: A systematic review and meta-analysis. *Psychotherapy and Psychosomatics, 81*(4), 206-216.

Chipps, J., Brysiewicz, P., & Mars, M. (2012). A systematic review of the effectiveness of videoconference-based tele-education for medical and nursing education. *Worldviews on Evidence-Based Nursing, 9*(2), 78-87. doi:http://dx.doi.org/10.1111/j.1741-6787.2012.00241.x

Ciarrochi, J., Scott, G., Deane, F. P., & Heaven, P. C. L. (2003). Relations between social and emotional competence and mental health: a construct validation study. *Personality and Individual Differences, 35*(8), 1947-1963. doi:10.1016/s0191-8869(03)00043-6

Clarke, A. M., Kuosmanen, T., & Barry, M. M. (2015). A systematic review of online youth mental health promotion and prevention interventions. *Journal of Youth and Adolescence, 44*(1), 90-113.

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development, 9*(4), 377-395. doi:http://dx.doi.org/10.1016/0885-2014%2894%2990012-4

Cognitive Training Data Response Letter (n.d.). Retrieved from https://www.cognitivetrainingdata.org/the-controversy-does-brain-training-work/response-letter/

Collette, F., Hogge, M., Salmon, E., & Van der Linden, M. (2006). Exploration of the neural substrates of executive functioning by functional neuroimaging. *Neuroscience, 139*(1), 209-221. doi:10.1016/j.neuroscience.2005.05.035

Collin, L., Bindra, J., Raju, M., Gillberg, C., & Minnis, H. (2013). Facial emotion recognition in child psychiatry: A systematic review. *Research in Developmental Disabilities, 34*(5), 1505-1520.

Comay, J. (2010). Individual differences in narrative perspective-taking and theory of mind. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 70*(10-B), 6583.

Cooke, J. E., Stuart-Parrigon, K. L., Movahed-Abtahi, M., Koehn, A. J., & Kerns, K. A. (2016). Children's emotion understanding and mother-child attachment: A meta-analysis. *Emotion, 16*(8), 1102-1106. doi:http://dx.doi.org/10.1037/emo0000221

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin, 115*(1), 74-101. doi:http://dx.doi.org/10.1037/0033-2909.115.1.74

Crick, N. R., & Dodge, K. A. (1999). 'Superiority' is in the eye of the beholder: A comment of Sutton, Smith and Swettenham. *Social Development, 8*(1), 128-131.

Cuijpers, P., Marks, I. M., van Straten, A., Cavanagh, K., Gega, L., & Andersson, G. (2009). Computer-aided psychotherapy for anxiety disorders: a meta-analytic review. *Cognitive Behaviour Therapy, 38*(2), 66-82. doi:10.1080/16506070802694776

Cutting, A. L., & Dunn, J. (1999). Theory of Mind, Emotion Understanding, Language, and Family Background: Individual Differences and Interrelations. *Child Development, 70*(4), 853-865.

Damasio, A. (1994). *Descartes' Error*. New York: Putnam

Darwin, C. (1872). The expression of the emotions in man and animals. *(1872) The expression of the emotions in man and animals vi, 374 pp London, England: John Murray; England*.

De Las Cuevas, C., Arredondo, M.T., Cabrera, M.F., Sulzenbacher, H. & Meise, U. (2006). Randomized Clinical Trial of Telepsychiatry through Videoconference versus Face-to-Face Conventional Psychiatric Treatment. *Telemedicine and e-Health, 12*, 341-350. doi:10.1089/tmj.2006.12.341

De Rosnay, M., Pons, F., Harris, P. L., & Morrell, J. M. (2004). A lag between understanding false belief and emotion attribution in young children: Relationships with linguistic ability and mothers' mental-state language. *British Journal of Developmental Psychology, 22*(2), 197-218.

De Rosnay, M., & Hughes, C. (2006). Conversation and theory of mind: Do children talk their way to socio-cognitive understanding? *British Journal of Developmental Psychology, 24*(1), 7-37.

De Rosnay, M. D., Harris, P. L., & Pons, F. (2008). Emotional understanding and developmental psychopathology in young children *Social cognition and developmental psychopathology* (pp. 343-385). New York, NY: Oxford University Press; US.

De Villiers, J. G. Can language acquisition give children a point of view? In: J. Astington J, J. Baird J (Eds.), *Why Language Matters for Theory of Mind* (pp. 186–219). New York: Oxford University Press.

Denham, S. A. (1986). Social cognition, prosocial behavior, and emotion in preschoolers: Contextual validation. *Child Development, 57*(1), 194-201.

Denham, S. A., & Couchoud, E. A. (1990). Young preschoolers' understanding of emotions. *Child Study Journal, 20*(3), 171-192.

Denham, S. A., McKinley, M., Couchoud, E. A., & Holt, R. (1990). Emotional and behavioral predictors of preschool peer ratings. *Child Development, 61*(4), 1145-1152.

Denham, S. A., Mitchell-Copeland, J., Strandberg, K., Auerbach, S., & Blair, K. (1997). Parental contributions to preschoolers' emotional competence: Direct and indirect effects. *Motivation and Emotion, 21*(1), 65-86.

Denham, S. A., Zoller, D., & Couchoud, E. A. (1994). Socialization of preschoolers' emotion understanding. *Developmental Psychology, 30*(6), 928-936.

DeRosier, M., Kupersmidt, J. B., & Patterson, C. J. (1994). Children's academic and behavioral adjustment as a function of the chronicity and proximity of peer rejection. *Child Development, 65*(6), 1799-1813.

Devine, R. T., & Hughes, C. (2014). Relations Between False Belief Understanding and Executive Function in Early Childhood: A Meta-Analysis. *Child Development*. doi: 10.1111/cdev.12237

DeWall, C. N., Twenge, J. M., Koole, S. L., Baumeister, R. F., Marquez, A., & Reid, M. W. (2011). Automatic emotion regulation after social exclusion: tuning to positivity. *Emotion, 11*(3), 623-636. doi:10.1037/a0023534

Diamond, A. (2013). Executive functions. *Annual Reviews of Psychology, 64*, 135-168. doi: 10.1146/annurev-psych-113011-143750

Ditch the Labe (2017). The Annual Bullying Survey 2017. Retrieved from https://www.ditchthelabel.org/wp-content/uploads/2017/07/The-Annual-Bullying-Survey-2017-1.pdf

Dodge, K. A., Pettit, G. S., McClaskey, C. L., & Brown, M. M. (1986). Social competence in children. *Monographs of the Society for Research in Child Development, 51*(2), 1-85. doi:http://dx.doi.org/10.2307/1165906

Donaldson, S. K., & Westerman, M. A. (1986). Development of children's understanding of ambivalence and causal theories of emotions. *Developmental Psychology, 22*(5), 655-662.

Dumontheil, I., Apperly, I. A., & Blakemore, S. J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science, 13*(2), 331-338. doi: 10.1111/j.1467-7687.2009.00888.x

Dunn, J. (1988). *The beginnings of social understanding*. Cambride, MA: Harvard University Press.

Dunn, J. (1995). Children as psychologists: The later correlates of individual differences in understanding of emotions and other minds. *Cognition and Emotion, 9*(2-3), 187-201.

Dunn, J., Brown, J., & Beardsall, L. (1991). Family talk about feeling states and children's later understanding of others' emotions. *Developmental Psychology, 27*(3), 448-455.

Dunning, D., & Hayes, A. F. (1996). Evidence for egocentric comparison in social judgment. *Journal of Personality and Social Psychology, 71*(2), 213-229. doi: 10.1037/0022-3514.71.2.213

Duval, C., Piolino, P., Bejanin, A., Eustache, F., & Desgranges, B. (2011). Age effects on different components of theory of mind. *Consciousness and Cognition, 20*(3), 627-642. doi:10.1016/j.concog.2010.10.025

Dyck, Murray J., & Denver, Esther. (2003). Can the Emotion Recognition Ability of Deaf Children be Enhanced? A Pilot Study. *Journal of Deaf Studies and Deaf Education, 8*(3), 348-356.

Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., . . . Convit, A. (2006). Introducing MASC: A Movie for the Assessment of Social Cognition. *Journal of Autism and Developmental Disorders, 36*(5), 623-636.

Ebert, D. D., Zarski, A. C., Christensen, H., Stikkelbroek, Y., Cuijpers, P., Berking, M., & Riper, H. (2015). Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: a meta-analysis of randomized controlled outcome trials. *PLoS One, 10*(3), e0119895. doi:10.1371/journal.pone.0119895

Ebesutani, C., Bernstein, A., Martinez, J. I., Chorpita, B. F., & Weisz, J. R. (2011). The youth self report: applicability and validity across younger and older youths. *Journal of Clinical Child and Adolescent Psychology, 40*(2), 338-346. doi: 10.1080/15374416.2011.546041

Ebesutani, C., Regan, J., Smith, A., Reise, S., Higa-McMillan, C., & Chorpita, B. F. (2012). The 10-Item Positive and Negative Affect Schedule for Children, Child and Parent Shortened Versions: Application of Item Response Theory for More Efficient Assessment. *Journal of Psychopathology and Behavioral Assessment, 34*(2), 191-203. doi: 10.1007/s10862-011-9273-2

Education Group GmbH. (2014). 4. Oberösterreichische Kinder-Medien-Studie. Retrieved from: https://www.edugroup.at/innovation/forschung/kinder-medien-studie/detail/4-ooe-kinder-medien-studie-2014.html

Edwards, R., Manstead, A. S., & MacDonald, C. J. (1984). The relationship between children's sociometric status and ability to recognize facial expressions of emotion. *European Journal of Social Psychology, 14*(2), 235-238.

Egner, T. (2011). Surprise! A unifying model of dorsal anterior cingulate function? *Nature Neuroscience, 14*(10), 1219-1220.

Eggum, N. D., Eisenberg, N., Kao, K., Spinrad, T. L., Bolnick, R., Hofer, C., . . . Fabricius, W. V. (2011). Emotion understanding, theory of mind, and prosocial orientation: Relations over time in early childhood. *Journal of Positive Psychology, 6*(1), 4-16. doi:10.1080/17439760.2010.536776

Eisbach, A. O. D. (2004). Children's Developing Awareness of Diversity in People's Trains of Thought. *Child Development, 75*(6), 1694-1707. doi:http://dx.doi.org/10.1111/j.1467-8624.2004.00810.x

Eisenberg, N., Spinrad, T. L., & Eggum, N. D. (2010). Emotion-related self-regulation and its relation to children's maladjustment. *Annual Review of Clinical Psychology, 6*, 495-525. doi:http://dx.doi.org/10.1146/annurev.clinpsy.121208.131208

Ekman, P. (1992). Are there basic emotions? *Psychological Review, 99*(3), 550-553. doi:http://dx.doi.org/10.1037/0033-295X.99.3.550

Ekman, P. (2003). *Micro Expressions Training Tool and the Subtle Expressions Training Tool (METT AND SETT)*. Venice (CA): MOZGO Media.

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion Review, 3*(4), 364-370.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality & Social Psychology February 1971;17(2):124-129*.

Elfenbein, Hillary Anger. (2006). Learning in emotion judgments: Training and the cross-cultural understanding of facial expressions. *Journal of Nonverbal Behavior, 30*(1), 21-36.

Epstein, S. (1990). Cognitive-exeriential self-theory. In L.A. Pervin (Ed.), *Handbook of personality: Theory and research.* (pp. 165-192). New York: Guilford.

Eslinger, P. J. (1996). Conceptualizing, describing, and measuring components of executive function: A summary. In G. R. Lyon & N. A. Krasnegor [Eds], *Attention, memory, and executive function* (pp 367-395). Baltimore, MD: Paul H Brookes Publishing

Ewoldsen, D. R., Eno, C. A., Okdie, B. M., Velez, J. A., Guadagno, R. E., & DeCoster, J. (2012). Effect of playing violent video games cooperatively or competitively on subsequent cooperative behavior. *Cyberpsychology, Behavior, and Social Networking, 15*(5), 277-280. doi:http://dx.doi.org/10.1089/cyber.2011.0308

Fabes, R. A., Eisenberg, N., Nyman, M., & Michealieu, Q. (1991). Young children's appraisals of others' spontaneous emotional reactions. *Developmental Psychology, 27*(5), 858-866. doi:http://dx.doi.org/10.1037/0012-1649.27.5.858

Farber, E. A., & Moely, B. F. (1979). *Inferring others' affective states: Their use of interpersonal vocal and facial cues by children of three age levels*. Paper presented at the biennial meeting of the Society for Research and Child Development, San Francisco.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*(7), 307-314.

Felleman, E. S., Carlson, C. R., Barden, R., Rosenberg, L., & Masters, J. C. (1983). Children's and adults' recognition of spontaneous and posed emotional expressions in young children. *Developmental Psychology, 19*(3), 405-413.

Fernyhough, C. (2008). Getting Vygotskian about theory of mind: Mediation, dialogue, and the development of social understanding. *Developmental Review, 28*(2), 225-262.

Festinger, L. (1957). A theory of cognitive dissonance. *(1957, reprinted 1962) A theory of cognitive dissonance xi, 291 pp Stanford University Press*.

Filippova, E., & Astington, J. W. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child Development*, 79(1), 126-138.

Fine, S. E., Izard, C. E., Mostow, A. J., Trentacosta, C. J., & Ackerman, B. P. (2003). First grade emotion knowledge as a predictor of fifth grade self-reported internalizing behaviors in children from economically disadvantaged families. *Development and Psychopathology, 15*(02), 331-342.

Fink, E., de Rosnay, M., Wierda, M., Koot, H. M., & Begeer, S. (2014). Brief report: accuracy and response time for the recognition of facial emotions in a large sample of children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 44*(9), 2363-2368. doi: 10.1007/s10803-014-2084-z

Fiori, M., Antonietti, J.-P., Mikolajczak, M., Luminet, O., Hansenne, M., & Rossier, J. (2014). What Is the Ability Emotional Intelligence Test (MSCEIT) Good for? An Evaluation Using Item Response Theory. *PLoS ONE*, *9*(6), e98827. http://doi.org/10.1371/journal.pone.0098827

Flavell, J. H., Green, F. L., & Flavell, E. R. (1993). Children's understanding of the stream of consciousness. *Child Development, 64*(2), 387-398.

Flavell, J. H., Green, F. L., & Flavell, E. R. (1995). Young children's knowledge about thinking. *Monographs of the Society for Research in Child Development, 60*(1; 243), v-96.

Flavell, J. H., Green, F. L., & Flavell, E. R. (1998). The mind has a mind of its own: Developing knowledge about mental uncontrollability. *Cognitive Development, 13*(1), 127-138.

Flavell, J. H., Mumme, D. L., Green, F. L., & Flavell, E. R. (1992). Young children's understanding of different types of beliefs. *Child Development, 63*(4), 960-977.

Flook, L., Repetti, R. L., & Ullman, J. B. (2005). Classroom Social Experiences as Predictors of Academic Performance. *Developmental Psychology, 41*(2), 319-327.

Flynn, E., O'Malley, C., & Wood, D. (2004). A longitudinal, microgenetic study of the emergence of false belief understanding and inhibition skills. *Developmental Science, 7*(1), 103-115. doi:http://dx.doi.org/10.1111/j.1467-7687.2004.00326.x

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatry Research, 12*(3), 189-198.

Foroushani, P. S., Schneider, J., & Assareh, N. (2011). Meta-review of the effectiveness of computerised CBT in treating depression. *BMC Psychiatry, 11*, 131. doi:10.1186/1471-244x-11-131

Fox, Jeremy K., Warner, Carrie Masia, Lerner, Amy B., Ludwig, Kristy, Ryan, Julie L., Colognori, Daniela, . . . Brotman, Laurie Miller. (2012). Preventive intervention for anxious preschoolers and their parents: Strengthening early emotional development. *Child Psychiatry Hum Dev, 43*(4), 544-559.

Franceschini, S., Gori, S., Ruffino, M., Viola, S., Molteni, M., & Facoetti, A. (2013). Action video games make dyslexic children read better. *Current Biology, 23*(6), 462-466. doi: 10.1016/j.cub.2013.01.044

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of Mind and Rule-Based Reasoning. *Cognitive Development, 10*, 483-527.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences, 2*(12), 493-501.

Gardner, D., Harris, P., Ohmoto, M., & Hamazaki, T. (1988). Japanese children's understanding of the distinction between real and apparent emotion. *International Journal of Behavioral Development, 11*(2), 203-218.

Gardner, J. E. (1991). Can the Mario Bros. help? Nintendo games as an adjunct in psychotherapy with children. *Psychotherapy: Theory, Research, Practice, Training, 28*(4), 667-670.

Garner, P. W., Jones, D. C., & Miner, J. L. (1994). Social competence among low-income preschoolers: Emotion socialization practices and social cognitive correlates. *Child Development, 65*(2), 622-637.

Garner, P. W., & Waajid, B. (2008). The associations of emotion knowledge and teacher-child relationships to preschool children's school-related developmental competence. *Journal of Applied Developmental Psychology, 29*(2), 89-100.

Garnham, W. A., & Perner, J. (2001). Actions really do speak louder than words-But only implicitly: Young children's understanding of false belief in action. *British Journal of Developmental Psychology, 19*(3), 413-432.

Garnham, W. A., & Ruffman, T. (2001). Doesn’t see, doesn’t know: is anticipatory looking really related to understanding or belief? *Developmental Science, 4*(1), 94-100. doi:10.1111/1467-7687.00153

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology, 18*(1), 1-16. doi:10.1002/acp.934

Gentile, D. A., Anderson, C. A., Yukawa, S., Ihori, N., Saleem, M., Ming, L. K., . . . Sakamoto, A. (2009). The effects of prosocial video games on prosocial behaviors: International evidence from correlational, longitudinal, and experimental studies.

Personality and Social Psychology Bulletin, 35(6), 752-763. doi:http://dx.doi.org/10.1177/0146167209333045

Gergely, G., Nadasdy, Z., Csibra, G., & Biro, S. (1995). Taking the intentional stance at 12 months of age. *Cognition, 56*(2), 165-193.

Glass G. V., Peckham P. D., Sanders J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. Review of Educational Research 42, 237–288

Gnepp, J., & Klayman, J. (1992). Recognition of uncertainty in emotional inferences: Reasoning about emotionally equivocal situations. *Developmental Psychology, 28*(1), 145-158. doi:http://dx.doi.org/10.1037/0012-1649.28.1.145

Gnepp, J., McKee, E., & Domanic, J. A. (1987). Children's use of situational information to infer emotion: Understanding emotionally equivocal situations. *Developmental Psychology, 23*(1), 114-123. doi:http://dx.doi.org/10.1037/0012-1649.23.1.114

Gobel, A., Henning, A., Moller, C., & Aschersleben, G. (2016). The Relationship between Emotion Comprehension and Internalizing and Externalizing Behavior in 7- to 10-Year-Old Children. *Frontiers in Psychology, 7*, 1917. doi:10.3389/fpsyg.2016.01917

Golan, O., & Baron-Cohen, S. (2006). Systemizing empathy: teaching adults with Asperger syndrome or high-functioning autism to recognize complex emotions using interactive multimedia. *Developmental Psychopathology, 18*(2), 591-617. doi: 10.1017/s0954579406060305

Golan, O., Baron-Cohen, S., Hill, J. J., & Golan, Y. (2006). The "Reading the Mind in Films" task: Complex emotion recognition in adults with and without autism spectrum conditions. *Social Neuroscience, 1*(2), 111-123.

Golan, O., Baron-Cohen, S., Hill, J. J., & Rutherford, M. (2007). The 'Reading the Mind in the Voice' Test-Revised: A study of complex emotion recognition in adults with and without autism spectrum conditions. *Journal of Autism and Developmental Disorders, 37*(6), 1096-1106.

Goldman, A. I. (1989). Interpretation Psychologized*. *Mind & Language, 4*(3), 161-185. doi:doi:10.1111/j.1468-0017.1989.tb00249.x

Goldman, A. I. (2005). Simulation Theory. In *Encyclopedia of Philosophy.* Retrieved from https://www.encyclopedia.com/humanities/encyclopedias-almanacs-transcripts-and-maps/simulation-theory.

Goodwin, L. D., & Leech, N. L. (2006). Understanding correlation: Factors that affect the size of r. *Journal of Experimental Education, 74*(3), 251-266. doi:http://dx.doi.org/10.3200/JEXE.74.3.249-266

Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development, 59*(1), 26-37.

Gopnik, A., & Wellman, H. M. (1992). Why the child's theory of mind really is a theory. *Mind & Language, 7*(1-2), 145-171.

Gopnik, A., & Wellman, H. M. (2012). Reconstructing Constructivism: Causal Models, Bayesian Learning Mechanisms, and the Theory Theory. *Psychological Bulletin*, No Pagination Specified.

Gordon, M. (2001). *Roots of Empathy: Training manual*. Available from the Roots of Empathy, www.Rootsofempathy.org

Gordon, R. M. (1986). Folk psychology as simulation. *Mind and Language, 1*(2), 158-171.

Granic, I., Lobel, A., & Engels, R. C. (2014). The benefits of playing video games. *American Psychologist, 69*(1), 66-78. doi:10.1037/a0034857

Grawe, K. (2004). *Neuropsychotherapie*. Göttingen: Hogrefe.

Graziano, P. A., Reavis, R. D., Keane, S. P., & Calkins, S. D. (2007). The role of emotion regulation in children's early academic success. *Journal of School Psychology, 45*(1), 3-19.

Grazzani, Ilaria, & Ornaghi, Veronica. (2011). Emotional state talk and emotion understanding: A training study with preschool children. *Journal of Child Language, 38*(5), 1124-1139.

Grazzani, I., & Ornaghi, V. (2012). How do use and comprehension of mental-state language relate to theory of mind in middle childhood? *Cognitive Development, 27*(2), 99-111. doi:10.1016/j.cogdev.2012.03.002

Green, C. S., & Bavelier, D. (2003). Action video games modifies visual selective attention. *Nature, 423*(534-537).

Green, C. S., & Bavelier, D. (2006). Effect of action video games on the spatial distribution of visuospatial attention. *Journal of Experimental Psychology: Human Perception and Performance, 32*(6), 1465-1478. doi: 10.1037/0096-1523.32.6.1465

Green, M. F., & Horan, W. P. (2010). Social Cognition in Schizophrenia. *Current Directions in Psychological Science, 19*(4), 243-248. doi: 10.1177/0963721410377600

Gresham, F. M., Elliott, S. N., Vance, M. J., & Cook, C. R. (2011). Comparability of the Social Skills Rating System to the Social Skills Improvement System: Content and psychometric comparisons across elementary and secondary age levels. *School Psychology Quarterly, 26*(1), 27-44. doi: 10.1037/a0022662

Griffiths, M. (2003). The Therapeutic Use of Videogames in Childhood and Adolescence. *Clinical Child Psychology and Psychiatry, 8*(4), 547-554. doi:10.1177/13591045030084012

Gross, J. J., Carstensen, L. L., Pasupathi, M., Tsai, J., Gotestam Skorpen, C., & Hsu, A. Y. (1997). Emotion and aging: Experience, expression, and control. *Psychology and Aging, 12*(4), 590-599. doi:http://dx.doi.org/10.1037/0882-7974.12.4.590

Grynszpan, O., Perbal, S., Pelissolo, A., Fossati, P., Jouvent, R., Dubal, S., & Perez-Diaz, F. (2011). Efficacy and specificity of computer-assisted cognitive remediation in schizophrenia: a meta-analytical study. *Psychological Medicine, 41*(01), 163-173. doi: doi:10.1017/S0033291710000607

Guilford, J. (1967). Creativity: Yesterday, today, and tomorrow. *The Journal of Creative Behavior, 1*(1), 3-14. doi:http://dx.doi.org/10.1002/j.2162-6057.1967.tb00002.x

Gullone, E., Hughes, E. K., King, N. J., & Tonge, B. (2010). The normative development of emotion regulation strategy use in children and adolescents: A 2-year follow-up study. *Journal of Child Psychology and Psychiatry, 51*(5), 567-574. doi:http://dx.doi.org/10.1111/j.1469-7610.2009.02183.x

Gyurak, A., Gross, J. J., & Etkin, A. (2011). Explicit and implicit emotion regulation: A dual-process framework. *Cognition and Emotion, 25*(3), 400-412.

Hadwin, Julie, Baron-Cohen, Simon, Howlin, Patricia, & Hill, Katie. (1996). Can we teach children with autism to understand emotions, belief, or pretence? *Development and Psychopathology, 8*(2), 345-365.

Hadwin, J., & Perner, J. (1991). Pleased and surprised: Children's cognitive theory of emotion. *British Journal of Developmental Psychology, 9*(2), 215-234. doi:http://dx.doi.org/10.1111/j.2044-835X.1991.tb00872.x

Haidt, J. (2013). Moral psychology for the twenty-first century. *Journal of Moral Education, 42*(3), 281-297. doi:http://dx.doi.org/10.1080/03057240.2013.817327

Haith, M. M. (1998). Who put the cog in infant cognition? Is rich interpretation too costly? *Infant Behavior & Development, 21*(2), 167-179.

Happe, F. G. (1994). An advanced test of theory of mind: understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders, 24*(2), 129-154.

Harley, C. W. (2004). Norepinephrine and Dopamine as Learning Signals. *Neural Plasticity, 11*(3-4), 191-204. doi:10.1155/np.2004.191

Harlow JM (1848): Passage of an iron rod through the head. *Boston Medical and Surgical Journal, 39*, 389-393.

Harris, P. L. (1992). From simulation to folk psychology: The case for development. *Mind & Language, 7*(1-2), 120-144.

Harris, P. L., de Rosnay, M., & Ronfard, S. (2014). The mysterious emotional life of little red riding hood. In K. H. Lagattuta (Ed.), *Children and emotion: New insights into developmental affective sciences* (pp. 106–118). Switzerland: Karger. http://dx.doi.org/10.1159/000354364.

Harris, P. L., Donnelly, K., Guz, G. R., & Pitt-Watson, R. (1986). Children's understanding of the distinction between real and apparent emotion. *Child Develoment, 57*(4), 895-909.

Harris, P. L., Guz, G. R., Lipian, M. S., & Man-Shu, Z. (1985). Insight into the time course of emotion among Western and Chinese children. *Child Development, 56*(4), 972-988. doi:http://dx.doi.org/10.2307/1130109

Harris, P. L., Johnson, C. N., Hutton, D., Andrews, G., & Cooke, T. (1989). Young children's theory of mind and emotion. *Cognition and Emotion, 3*(4), 379-400.

Harris, P. L., Lillard, A., & Perner, J. (1994). Commentary: Triangulating pretence and belief. In C. Lewis & P. Mitchell [Eds], *Children's early understanding of mind: Origins and development* (pp 287-293). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc

Harris, P. L., Olthof, T., & Terwogt, M. M. (1981). Children's knowledge of emotion. *Child Psychology & Psychiatry & Allied Disciplines, 22*(3), 247-261. doi:http://dx.doi.org/10.1111/j.1469-7610.1981.tb00550.x

Harris, P. L., & Saarni, C. (1989). Children's understanding of emotion: An introduction. In *Children's understanding of emotion* (pp. 3-24). New York, NY: Cambridge University Press; US.

Haslinger, H. (2014). *Computerized test of emotion comprehension* (Diplomarbeit), University of Vienna, Vienna. Retrieved from http://othes.univie.ac.at/34908/

Haslinger, H., & Leyrer, J. (2013). CTRUE: Computerized Task of Recognizing and Understanding Emotions [Measurement instrument].

Harter, S., & Buddin, B. J. (1987). Children's understanding of the simultaneity of two emotions: A five-stage developmental acquisition sequence. *Developmental Psychology, 23*(3), 388-399.

Hartley, H. O. (1950). The use of range in analysis of variance. *Biometrika, 37*, 271-280.

Harvey, P. D., McGurk, S. R., Mahncke, H., & Wykes, T. (2018). Controversies in Computerized Cognitive Training. *Biological Psychiatry Cognitive Neuroscience and Neuroimaging, 3*(11), 907-915. doi:10.1016/j.bpsc.2018.06.008

Hayward, E. O. (2011). *Measurement of advanced theory of mind in school-age children: Investigating the validity of a unified construct* (Unpublished doctoral dissertation). New York University, New York, NY.

Hayward, E. O., & Homer, B. D. (2017). Reliability and validity of advanced theory-of-mind measures in middle childhood and adolescence. *British Journal of Developmental Psychology, 35*(3), 454-462. doi:http://dx.doi.org/10.1111/bjdp.12186

Hayward, E. O., Homer, B. D., & Sprung, M. (2016). Developmental trends in flexibility and automaticity of social cognition. *Child Development*, No Pagination Specified. doi:http://dx.doi.org/10.1111/cdev.12705

Heal, J. (1986). Replication and functionalism. In J. Butterfield (Ed.), *Language, Mind, and Logic* (Vol. 14, pp. 135--150): Cambridge University Press.

Heckhausen, H. & Kuhl, J. (1985). From wishes to action: The dead ends and short cuts on the long way to action. In M. Frese & J. Sabini (Eds.), *Goal-directed behavior: The concept of action in psychology* (pp. 134–159). Hillsdale (NJ): Lawrence Erlbaum.

Heider, F. (1958). The psychology of interpersonal relations. *(1958) The psychology of interpersonal relations ix, 326 pp Hoboken, NJ, US: John Wiley & Sons Inc; US*.

Heidrich, S. M., & Denney, N. W. (1994). Does social problem solving differ from other types of problem solving during the adult years? *Experimental Aging Research, 20*(2), 105-126.

Henning, A., Spinath, F. M., & Aschersleben, G. (2011). The link between preschoolers' executive function and theory of mind and the role of epistemic states. *Journal of Experimental Child Psychology, 108*(3), 513-531. doi: 10.1016/j.jecp.2010.10.006

Henry, J. D., Phillips, L. H., Ruffman, T., & Bailey, P. E. (2013). A meta-analytic review of age differences in theory of mind. *Psychology and Aging, 28*(3), 826-839.

Herbet, G., Lafargue, G., Bonnetblanc, F., Moritz-Gasser, S., Menjot de Champfleur, N., & Duffau, H. (2014). Inferring a dual-stream model of mentalizing from associative white matter fibres disconnection. *Brain, 137*(3), 944.

Hess, M., Scheithauer, H., Kleiber, D., Wille, N., Erhart, M., & Ravens-Sieberer, U. (2013). The Parent Version of the Preschool Social Skills Rating System: Psychometric Analysis and Adaptation with a German Preschool Sample. *Journal of Psychoeducational Assessment, 32*(3), 216-226. doi: 10.1177/0734282913483977

Hill, A. L., Degnan, K. A., Calkins, S. D., & Keane, S. P. (2006). Profiles of externalizing behavior problems for boys and girls across preschool: The roles of emotion regulation and inattention. *Developmental Psychology, 42*(5), 913-928. doi:http://dx.doi.org/10.1037/0012-1649.42.5.913

Hilty, D. M., Ferrer, D. C., Parish, M. B., Johnston, B., Callahan, E. J., & Yellowlees, P. M. (2013). The effectiveness of Telemental health: A 2013 review. *Telemedicine and e-Health, 19*(6), 444-454. doi:http://dx.doi.org/10.1089/tmj.2013.0075

Hochberg, Y., & Benjamini, Y. (1990). More powerful procedures for multiple significance testing. *Statistics in Medicine, 9*(7), 811-818. doi: 10.1002/sim.4780090710

Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research, 67*(1), 88-140. doi:http://dx.doi.org/10.2307/1170620

Hogrefe, G., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: A developmental lag in attribution of epistemic states. *Child Development, 57*(3), 567-582.

Hooker, C. I., Bruce, L., Fisher, M., Verosky, S. C., Miyakawa, A., D'Esposito, M., & Vinogradov, S. (2013). The influence of combined cognitive plus social-cognitive training on amygdala response during face emotion recognition in schizophrenia. *Psychiatry Res, 213*(2), 99-107. doi: 10.1016/j.pscychresns.2013.04.001

Horan, W. P., Kern, R. S., Shokat-Fadai, K., Sergi, M. J., Wynn, J. K., & Green, M. F. (2009). Social cognitive skills training in schizophrenia: an initial efficacy study of stabilized outpatients. *Schizophrenia Research, 107*(1), 47-54. doi: 10.1016/j.schres.2008.09.006

Horan, W. P., Kern, R. S., Tripp, C., Hellemann, G., Wynn, J. K., Bell, M., . . . Green, M. F. (2011). Efficacy and specificity of social cognitive skills training for outpatients with psychotic disorders. *Journal of Psychiatric Research, 45*(8), 1113-1122. doi: 10.1016/j.jpsychires.2011.01.015

Howell, D.C. (2002, November 02). *Multiple Comparisons with Repeated Measures*. Retrieved from https://www.uvm.edu/~dhowell/StatPages/More_Stuff/RepMeasMultComp/RepMeasMultComp.html

Howell, G. T. & Lacroix, G., L. (2012). Decomposing interactions using GLM in combination with the COMPARE, LMATRIX and MMATRIX subcommands in SPSS. *Tutorials in Quantitative Methods for Psychology, 8*(1), 1-22.

Hu, T., Zhang, D., Wang, J., Mistry, R., Ran, G., & Wang, X. (2014). Relation between emotion regulation and mental health: A meta-analysis review. *Psychological Reports, 114*(2), 341-362. doi:http://dx.doi.org/10.2466/03.20.PR0.114k22w4

Huelle, J. O., Sack, B., Broer, K., Komlewa, I., & Anders, S. (2014). Unsupervised learning of facial emotion decoding skills. *Front Hum Neurosci, 8*, 77. doi: 10.3389/fnhum.2014.00077

Huether, G. (1998). Stress and the adaptive self-organization of neuronal connectivity during early childhood. *International Journal of Developmental Neuroscience, 16*(3-4), 297-306.

Hughes, C., Deater-Deckard, K., & Cutting, A. L. (1999). "Speak roughly to your little boy?" Sex differences in the relations between parenting and preschoolers' understanding of mind. *Social Development, 8*(2), 143-160. doi:http://dx.doi.org/10.1111/1467-9507.00088

Hughes, C., & Dunn, J. (1998). Understanding mind and emotion: Longitudinal associations with mental-state talk between young friends. *Developmental Psychology, 34*(5), 1026-1037. doi:http://dx.doi.org/10.1037/0012-1649.34.5.1026

Hughes, C., Dunn, J., & White, A. (1998). Trick or treat? Uneven understanding of mind and emotion and executive dysfunction in "hard-to-manage" preschoolers. *Journal of Child Psychology and Psychiatry, 39*(7), 981-994.

Hughes, C., & Ensor, R. (2007). Executive function and theory of mind: Predictive relations from ages 2 to 4. *Developmental Psychology, 43*(6), 1447-1459. doi:http://dx.doi.org/10.1037/0012-1649.43.6.1447

Hughes, C., Jaffee, S. R., Happe, F., Taylor, A., Caspi, A., & Moffitt, T. E. (2005). Origins of Individual Differences in Theory of Mind: From Nature to Nurture? *Child Development, 76*(2), 356-370.

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: developmental trends and a latent variable analysis. *Neuropsychologia, 44*(11), 2017-2036. doi:10.1016/j.neuropsychologia.2006.01.010

Huizinga, M., & van der Molen, M. W. (2007). Age-Group Differences in Set-Switching and Set-Maintenance on the Wisconsin Card Sorting Task. *Developmental Neuropsychology, 31*(2), 193-215. doi:10.1080/87565640701190817

Hyler, S. E., & Gangure, D. P. (2003). A review of the costs of telepsychiatry. *Psychiatric Services, 54*(7), 976-980. doi:http://dx.doi.org/10.1176/appi.ps.54.7.976

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the Intentions of Others with One's Own Mirror Neuron System. *PLoS Biol, 3*(3), e79. doi:10.1371/journal.pbio.0030079

Institut für Jugendkulturforschung. (2008). Computerspiele im Alltag Jugendlicher: Gamer-Segmente und Gamer-Kulturen in der Altersgruppe der 11- bis 18-Jährigen (p. 47). Wien. Retrieved from http://jugendkultur.at/medien-und-kommunikation/

Institute for Health Metrics and Evaluation (IHME). GBD Compare Data Visualization. Seattle, WA: IHME, University of Washington, 2017. Available from http://vizhub.healthdata.org/gbd-compare. (Accessed 19.12.2017)

Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science, 2*(3), 260-280.

Izard, C., Fine, S., Schultz, D., Mostow, A., Ackerman, B., & Youngstrom, E. (2001). Emotion Knowledge as a Predictor of Social Behavior and Academic Competence in Children at Risk. *Psychological Science, 12*(1), 18-23. doi: 10.1111/1467-9280.00304

Jackson, A. (2001). Language facility and theory of mind development in deaf children. *Journal of Deaf Studies and Deaf Education, 6*(3), 161-176.

Jahn, A., Nee, D. E., Alexander, W. H., & Brown, J. W. (2014). Distinct regions of anterior cingulate cortex signal prediction and outcome evaluation. *NeuroImage, 95*, 80-89.

Jak, A. J., Seelye, A. M., & Jurick, S. M. (2013). Crosswords to computers: a critical review of popular approaches to cognitive enhancement. *Neuropsychology Review, 23*(1), 13-26. doi: 10.1007/s11065-013-9226-5

Jenkins-Guarnieri, M. A., Pruitt, L. D., Luxton, D. D., & Johnson, K. (2015). Patient perceptions of telemental health: Systematic review of direct comparisons to in-person psychotherapeutic treatments. *Telemedicine and e-Health, 21*(8), 652-660. doi:http://dx.doi.org/10.1089/tmj.2014.0165

Joshi, M. S., & MacLean, M. (1994). Indian and English children's understanding of the distinction between real and apparent emotion. *Child Development, 65*(5), 1372-1384.

Kaland, N., Moller-Nielsen, A., Callesen, K., Mortensen, E. L., Gottlieb, D., & Smith, L. (2002). A new "advanced" test of theory of mind: Evidence from children and adolescents with Asperger syndrome. *Journal of Child Psychology and Psychiatry, 43*(4), 517-528.

Kaland, N., Moller-Nielsen, A., Smith, L., Mortensen, E. L., Callesen, K., & Gottlieb, D. (2005). The strange stories test: A replication study of children and adolescents with Asperger syndrome. European Child & Adolescent Psychiatry, 14(2), 73-82.

Kaland, N., Callesen, K., Moller-Nielsen, A., Mortensen, E. L., & Smith, L. (2008). Performance of children and adolescents with Asperger syndrome or high-functioning autism on advanced theory of mind tasks. *Journal of Autism and Developmental Disorders, 38*(6), 1112-1123.

Kaltenthaler, E., Brazier, J., de Nigris, E., Tumur, I., & Ferriter, M. (2006). Computerised cognitive behaviour therapy for depression and anxiety update: a systematic review and economic evaluation. *Health Technology Assessment, 10*(33), 186. doi:10.3310/hta10330

Karlsson, M. P., Tervo, D. G., & Karpova, A. Y. (2012). Network resets in medial prefrontal cortex mark the onset of behavioral uncertainty. *Science, 338*(6103), 135-139.

Kestenbaum, R., & Gelman, S. A. (1995). Preschool children's identification and understanding of mixed emotions. *Cognitive Development, 10*(3), 443-458. doi:http://dx.doi.org/10.1016/0885-2014%2895%2990006-3

Kestenbaum, R., & Nelson, C. A. (1992). Neural and behavioral correlates of emotion recognition in children and adults. *Journal of Experimental Child Psychology, 54*(1), 1-18.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science, 11*(1), 32-38.

Kidd, D. C., & Castano, E. (2013). Reading literary fiction improves theory of mind. *Science, 342*(6156), 377-380. doi:http://dx.doi.org/10.1126/science.1239918

Klann-Delius, G. (1998). *Kodiermanual.*

Klinnert, M. D. (1984). The regulation of infant behavior by maternal facial expression. *Infant Behavior & Development, 7*(4), 447-465. doi: http://dx.doi.org/10.1016/S0163-6383%2884%2980005-3

Kliegel, M., Jager, T., & Phillips, L. H. (2007). Emotional development across adulthood: Differential age-related emotional reactivity and emotion regulation in a negative mood induction procedure. *The International Journal of Aging & Human Development, 64*(3), 217-244. doi:http://dx.doi.org/10.2190/U48Q-0063-3318-1175

Kloo, D., Perner, J., & Giritzer, T. (2010). Object-based set-shifting in preschoolers: Relations to theory of mind *Self and social regulation: Social interaction and the development of social understanding and executive functions* (pp. 193-217). New York, NY: Oxford University Press; US.

Kocsis-Bogar, K., Kotulla, S., Maier, S., Voracek, M., & Hennig-Fast, K. (2017). Cognitive correlates of different mentalizing abilities in individuals with high and low trait schizotypy: Findings from an extreme-group design. *Frontiers in Psychology Vol 8 2017, ArtID 922, 8*. doi:http://dx.doi.org/10.3389/fpsyg.2017.00922

Kolko, D. J., & Kazdin, A. E. (1993). Emotional/behavioral problems in clinic and nonclinic children: Correspondence among child, parent and teacher reports. *Child Psychology & Psychiatry & Allied Disciplines, 34*(6), 991-1006.

Kolling, N., Behrens, T., Wittmann, M., & Rushworth, M. (2016). Multiple signals in anterior cingulate cortex. *Current Opinion in Neurobiology, 37*, 36-43.

Kreutzer, M. A., Leonard, C., & Flavell, J. H. (1975). An interview study of children's knowledge about memory. *Monographs of the Society for Research in Child Development, 40*(1), 1-60.

Kristen, S., Thoermer, C., Hofer, T., Aschersleben, G., & Sodian, B. (2006). Validation of the "Theory of Mind" Scale. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 38*(4), 186-195.

LaBounty, J., Wellman, H. M., Olson, S., Lagattuta, K., & Liu, D. (2008). Mothers'andFathers' Use of Internal State Talk with their Young Children. *Social Development, 17*(4), 757-775. doi: 10.1111/j.1467-9507.2007.00450.x

Ladd, G. W. (1990). Having friends, Keeping friends, making friends, and being liked by peers in the classroom: Predictors of children's early school adjustment? *Child Development, 61*(4), 1081-1100.

Lagattuta, K. H. (2005). When You Shouldn't Do What You Want to Do: Young Children's Understanding of Desires, Rules, and Emotions. *Child Development, 76*(3), 713-733. doi:http://dx.doi.org/10.1111/j.1467-8624.2005.00873.x

Lagattuta, K. H., Kramer, H. J., Kennedy, K., Hjortsvang, K., Goldfarb, D., & Tashjian, S. (2015). Beyond Sally's missing marble: further development in children's understanding of mind and emotion in middle childhood. *Advances in Child Development and Behavior, 48*, 185-217. doi:10.1016/bs.acdb.2014.11.005

Lagattuta, K. H., Sayfan, L., & Blattman, A. J. (2010). Forgetting common ground: Six- to seven-year-olds have an overinterpretive theory of mind. *Developmental Psychology, 46*(6), 1417-1432.

Lagattuta, K. H., & Sayfan, L. (2013). Not all past events are equal: Biased attention and emerging heuristics in children's past-to-future forecasting. *Child Development, 84*(6), 2094-2111. doi:http://dx.doi.org/10.1111/cdev.12082

Lagattuta, K. H., Sayfan, L., & Harvey, C. (2014). Beliefs about thought probability: evidence for persistent errors in mindreading and links to executive control. *Child Development, 85*(2), 659-674. doi:10.1111/cdev.12154

Lagattuta, K. H., & Wellman, H. M. (2001). Thinking about the past: Early knowledge about links between prior experience, thinking, and emotion. *Child Development, 72*(1), 82-102.

Lagattuta, K. H., Wellman, H. M., & Flavell, J. H. (1997). Preschoolers' understanding of the link between thinking and feeling: Cognitive cuing and emotional change. *Child Development, 68*(6), 1081-1104.

Lake, N., Lane, S., & Harris, P. L. (1995). The expectation of guilt and resistance to temptation. *Early Development and Parenting, 4*(2), 63-73. doi:doi:10.1002/edp.2430040203

Lalonde, C. E., & Chandler, M. J. (1995). False belief understanding goes to school: On the social-emotional consequences of coming early or late to a first theory of mind. *Cognition and Emotion, 9*(2-3), 167-185.

Lampit, A., Hallock, H., & Valenzuela, M. (2014). Computerized Cognitive Training in Cognitively Healthy Older Adults: A Systematic Review and Meta-Analysis of Effect Modifiers. *PLoS Medicine, 11*(11), e1001756. doi:10.1371/journal.pmed.1001756

Laurent, J., Catanzaro, S. J., Joiner, T. E., Jr., Rudolph, K. D., Potter, K. I., Lambert, S., . . . Gathright, T. (1999). A measure of positive and negative affect for children: Scale development and preliminary validation. *Psychological Assessment, 11*(3), 326-338.

Lecce, S., Bianco, F., Devine, R. T., & Hughes, C. (2017). Relations between theory of mind and executive function in middle childhood: A short-term longitudinal study. *Journal of Experimental Child Psychology, 163*, 69-86. doi:http://dx.doi.org/10.1016/j.jecp.2017.06.011

Lecce, S., Caputi, M., Pagnin, A., & Banerjee, R. (2017). Theory of mind and school achievement: The mediating role of social competence. *Cognitive Development, 44*, 85-97.

Lee, K., Eskritt, M., Symons, L. A., & Muir, D. (1998). Children's use of triadic eye gaze information for "mind reading". *Developmental Psychology, 34*(3), 525-539.

Lench, H. C., Bench, S. W., & Flores, S. A. (2013). Searching for evidence, not a war: Reply to Lindquist, Siegel, Quigley, and Barrett (2013). *Psychological Bulletin, 139*(1), 264-268. doi:http://dx.doi.org/10.1037/a0029296

Lench, H. C., Flores, S. A., & Bench, S. W. (2011). Discrete emotions predict changes in cognition, judgment, experience, behavior, and physiology: A meta-analysis of experimental emotion elicitations. *Psychological Bulletin, 137*(5), 834-855.

Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind.". *Psychological Review, 94*(4), 412-426. doi:http://dx.doi.org/10.1037/0033-295X.94.4.412

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in 'theory of mind'. *Trends in Cognitive Sciences, 8*(12), 529-533.

Leslie, A. M., German, T. P., & Polizzi, P. (2005). Belief-desire reasoning as a process of selection. *Cognitive Psychology, 50*(1), 45-85. doi:http://dx.doi.org/10.1016/j.cogpsych.2004.06.002

Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition, 43*(3), 225-251. doi:http://dx.doi.org/10.1016/0010-0277%2892%2990013-8

Lever, A. G., & Geurts, H. M. (2016). Age-related differences in cognition across the adult lifespan in autism spectrum disorder. *Autism Research, 9*(6), 666-676.

Lewis, C., Freeman, N. H., Kyriakidou, C., Maridaki-Kassotaki, K., & Berridge, D. M. (1996). Social influences on false belief access: Specific sibling influences or general apprenticeship? *Child Development, 67*(6), 2930-2947. doi:http://dx.doi.org/10.2307/1131760

Lezak, M. D., Howieson, D. B., Loring, D. W., Hannay, H., & Fischer, J. S. (2004). Neuropsychological assessment (4th ed.). *(2004) Neuropsychological assessment (4th ed) xiv, 1016 pp New York, NY, US: Oxford University Press; US.*

Lindsey, E. W., & Colwell, M. J. (2003). Preschoolers' emotional competence: Links to pretend and physical play. *Child Study Journal, 33*(1), 39-52.

Lindquist, K. A., Siegel, E. H., Quigley, K. S., & Barrett, L. F. (2013). The hundred-year emotion war: Are emotions natural kinds or psychological constructions? Comment on Lench, Flores, and Bench (2011) Are emotions natural kinds? *Psychological Bulletin, 139*(1), 255-263.

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: a meta-analytic review. *Behavioral and Brain Sciences, 35*(3), 121-143. doi:10.1017/S0140525X11000446

Lisman, J., Grace, A. A., & Duzel, E. (2011). A neoHebbian framework for episodic memory; role of dopamine-dependent late LTP. *Trends in Neurosciences, 34*(10), 536-547.

Littlewood, E., Duarte, A., Hewitt, C., Knowles, S., Palmer, S., Walker, S., . . . Gilbody, S. (2015). A randomised controlled trial of computerised cognitive behaviour therapy for the treatment of depression in primary care: the Randomised Evaluation of the Effectiveness and Acceptability of Computerised Therapy (REEACT) trial. *Health Technology Assessment, 19*(101), 1-174. doi:10.3310/hta191010

Lohmann, H., & Tomasello, M. (2003). The Role of Language in the Development of False Belief Understanding: A Training Study. *Child Development, 74*(4), 1130-1144. doi:10.1111/1467-8624.00597

Lord, F. M. (1967). A Paradox in the Interpretation of Group Comparisons. *Psychological Bulletin, 68*(5), 304-305. doi: http://dx.doi.org/10.1037/h0025105

Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska Directed Emotional Faces - KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, ISBN 91-630-7164-9.

Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A., & Sweeney, J. A. (2004). Maturation of Cognitive Processes from Late Childhood to Adulthood. *Child Development, 75*(5), 1357-1372. doi:10.1111/j.1467-8624.2004.00745.x

Lynch, B. (2002). Historical review of computer-assisted cognitive retraining. *The Journal of Head Trauma Rehabilitation, 17*(5), 446-457.

Macmillan, M. (2000). Restoring Phineas Gage: A 150th retrospective. *Journal of the History of the Neurosciences, 9*(1), 46-66.

MacWhinney, B., & Snow, C. (1985). The child language data exchange system. *Journal of Child Language, 12*, 271-295. doi:10.1017/So3o5000900006449

Mahy, C. E., Moses, L. J., & Pfeifer, J. H. (2014). How and where: Theory-of-mind in the brain. *Developmental Cognitive Neuroscience, 9*, 68-81.

Maier, J. (2012). *Entwicklung einer Aufgabenbatterie zum kindlichen Verständnis mentaler Aktivität.* (Diploma thesis), University of Vienna, Vienna. Retrieved from http://othes.univie.ac.at/24600/

Malone, T. W. (1980). *What makes things fun to learn? Heuristics for designing instructional computer games*. Paper presented at the Proceedings of the 3rd ACM SIGSMALL symposium and the first SIGPC symposium on Small systems, Palo Alto, California, USA.

Malti, T., Gummerum, M., Keller, M., & Buchmann, M. (2009). Children's moral motivation, sympathy, and prosocial behavior. *Child Development, 80*(2), 442-460.

Mar, R. A., Oatley, K., Hirsh, J., dela Paz, J., & Peterson, J. B. (2006). Bookworms versus nerds: Exposure to fiction versus non-fiction, divergent associations with social ability, and the simulation of fictional social worlds. *Journal of Research in Personality, 40*(5), 694-712. doi:http://dx.doi.org/10.1016/j.jrp.2005.08.002

Marcovitch, S., O'Brien, M., Calkins, S. D., Leerkes, E. M., Weaver, J. M., & Levine, D. W. (2015). A Longitudinal Assessment of the Relation between Executive Function and Theory of Mind at 3, 4, and 5 Years. *Cognitive Development, 33*, 40-55. doi: 10.1016/j.cogdev.2014.07.001

Marder, S. R., & Fenton, W. (2004). Measurement and Treatment Research to Improve Cognition in Schizophrenia: NIMH MATRICS initiative to support the development of agents for improving cognition in schizophrenia. *Schizophrenia Research, 72*(1), 5-9. doi: 10.1016/j.schres.2004.09.010

Marraffa, M. (n.d.). Theory of Mind. In *Internet Encyclopedia of Philosophy*. Retrieved from https://www.iep.utm.edu/theomind/

Massimiliano, P. (2015). The effects of age on divergent thinking and creative objects production: A cross-sectional study. *High Ability Studies, 26*(1), 93-104. doi:http://dx.doi.org/10.1080/13598139.2015.1029117

Mauss, I. B., Bunge, S. A., & Gross, J. J. (2007). Automatic emotion regulation. *Social and Personality Psychology Compass, 1*(1), 146-167.

Martins, E. C., Osorio, A., Verissimo, M., & Martins, C. (2016). Emotion understanding in preschool children: The role of executive functions. *International Journal of Behavioral Development, 40*(1), 1-10.

Max Planck Institute for Human Development and Stanford Center on Longevity. (October 20, 2014). A Consensus on the Brain Training Industry from the Scientific Community. Retrieved from http://longevity.stanford.edu/a-consensus-on-the-brain-training-industry-from-the-scientific-community-2/

Mayer, J. D., Salovey, P., & Caruso, D. R. (2002). *Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) Item Booklet*. Toronto: MHS Publ.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2012). The Validity of the MSCEIT: Additional Analyses and Evidence. *Emotion Review, 4*(4), 403-408. doi: 10.1177/1754073912445815

McCleery, J. P., Surtees, A. D., Graham, K. A., Richards, J. E., & Apperly, I. A. (2011). The neural and cognitive time course of theory of mind. *Journal of Neuroscience, 31*(36), 12849-12854. doi: 10.1523/JNEUROSCI.1392-11.2011

Medienpädagogischer Forschungsverbund Südwest. (2011). KIM-Studie 2010: Kinder + Medien, Computer + Internet (p. 49). Stuttgart. Retrieved from: http://www.mpfs.de/fileadmin/KIM-pdf10/KIM2010.pdf

Meins, E., Fernyhough, C., Wainwright, R., Clark-Carter, D., Gupta, M. D., Fradley, E., & Tuckey, M. (2003). Pathways to Understanding Mind: Construct Validity and Predictive Validity of Maternal Mind-Mindedness. *Child Development, 74*(4), 1194-1211. doi:http://dx.doi.org/10.1111/1467-8624.00601

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology, 31*(5), 838-850.

Meltzoff, A. N., & Brooks, R. (2008). Self-experience as a mechanism for learning about others: A training study in social cognition. *Developmental Psychology, 44*(5), 1257-1265.

Meristo, M., & Hjelmquist, E. (2009). Executive functions and theory-of-mind among deaf children: Different routes to understanding other minds? *Journal of Cognition and Development, 10*(1-2), 67-91.

Meusburger, K. (2014). *Game Engine for web-based Serious Games* (master’s thesis), University of Vienna, Vienna. Retrieved from http://othes.univie.ac.at/34693

Mewborn, C. M., Lindbergh, C. A., & Miller, L. (2017). Cognitive interventions for cognitively healthy, mildly impaired, and mixed samples of older adults: A systematic review and meta-analysis of randomized-controlled trials. *Neuropsychology Review, 27*(4), 403-439. doi:http://dx.doi.org/10.1007/s11065-017-9350-8

Mikolajczak, M., Avalosse, H., Vancorenland, S., Verniest, R., Callens, M., van Broeck, N., . . . Mierop, A. (2015). A nationally representative study of emotional competence and health. *Emotion, 15*(5), 653-667. doi: 10.1037/emo0000034

Mill, A., Allik, J., Realo, A., & Valk, R. (2009). Age-related differences in emotion recognition ability: a cross-sectional study. *Emotion, 9*(5), 619-630. doi:10.1037/a0016562

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology, 110*(1), 40-48. doi: http://dx.doi.org/10.1037/0021-843X.110.1.40

Miller, S. A. (2009). Children's understanding of second-order mental states. *Psychological Bulletin, 135*(5), 749-773. doi:10.1037/a0016854

Miller, S. A. (2012). *Theory of mind: Beyond the preschool years*. New York, NY: Psychology Press; US.

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development, 78*(2), 622-646.

Mitchell, J. E., Crosby, R. D., Wonderlich, S. A., Crow, S., Lancaster, K., Simonich, H., . . . Myers, T. C. (2008). A randomized trial comparing the efficacy of cognitive-behavioral therapy for bulimia nervosa delivered via telemedicine versus face-to-face. *Behaviour Research and Therapy, 46*(5), 581-592. doi:http://dx.doi.org/10.1016/j.brat.2008.02.004

Mitchell, P., & Lacohee, H. (1991). Children's early understanding of false belief. *Cognition, 39*(2), 107-127. doi:http://dx.doi.org/10.1016/0010-0277%2891%2990040-B

Mitroff, S. R., Sobel, D. M., & Gopnik, A. (2006). Reversing How to Think about Ambiguous Figure Reversals: Spontaneous Alternating by Uninformed Observers. Perception, 35(5), 709-715. doi:10.1068/p5520

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. *Cogn Psychol, 41*(1), 49-100. doi: 10.1006/cogp.1999.0734

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science, 21*(1), 8-14.

Moll, H., & Tomasello, M. (2007). How 14- and 18-month-olds know what others have experienced. *Developmental Psychology, 43*(2), 309-317.

Moll, H., Carpenter, M., & Tomasello, M. (2014). Two- and 3-year-olds know what others have and have not heard. *Journal of Cognition and Development, 15*(1), 12-21.

Montagne, B., Kessels, R. P. C., De Haan, E. H. F., & Perrett, D. I. (2007). The Emotion Recognition Task: A paradigm to measure the perception of facial emotional expressions at different intensities. *Perceptual and Motor Skills, 104*(2), 589-598.

Montgomery, D. E. (1997). Wittgenstein's private language argument and children's understanding of the mind. *Developmental Review, 17*(3), 291-320. doi:http://dx.doi.org/10.1006/drev.1997.0436

Moran, J. M. (2013). Lifespan development: the effects of typical aging on theory of mind. *Behavioral Brain Research, 237*, 32-40. doi:10.1016/j.bbr.2012.09.020

Moscovitch, M. (1994). Cognitive resources and dual-task interference effects at retrieval in normal people: the role of the frontal lobes and medial temporal cortex. *Neuropsychology* 8, 524–534. doi: 10.1037/0894-4105.8.4.524

Moses, L. J. (2001). Executive Accounts of Theory-of-Mind Development. *Child Development, 72*(3), 688-690. doi:10.1111/1467-8624.00306

Mostow, A. J., Izard, C. E., Fine, S., & Trentacosta, C. J. (2002). Modeling emotional, cognitive, and behavioral predictors of peer acceptance. *Child Development, 73*(6), 1775-1787.

Muller, U., Liebermann-Finestone, D. P., Carpendale, J. I., Hammond, S. I., & Bibok, M. B. (2012). Knowing minds, controlling actions: the developmental relations between theory of mind and executive function from 2 to 4 years of age. *Journal of Experimental Child Psychology, 111*(2), 331-348. doi: 10.1016/j.jecp.2011.08.014

Muma, J. R. (1965). Peer evaluation and academic performance. *Personnel & Guidance Journal, 44*(4), 405-409.

Myers, K., & Cain, S. (2008). Practice Parameter for Telepsychiatry With Children and Adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry, 47*(12), 1468-1483. doi:10.1097/CHI.0b013e31818b4e13

Nahum, M., Fisher, M., Loewy, R., Poelke, G., Ventura, J., Nuechterlein, K. H., . . . Vinogradov, S. (2014). A novel, online social cognitive training program for young adults with schizophrenia: A pilot study. *Schizophrenia Research: Cognition, 1*(1), e11-e19. doi: http://dx.doi.org/10.1016/j.scog.2014.01.003

Nelson, C. A. (1987). The recognition of facial expressions in the first two years of life: Mechanisms of development. *Child Development, 58*(4), 889-909. doi:http://dx.doi.org/10.2307/1130530

Neudorfer, J. (2014). Der FASC, ein Verfahren zur Messung sozialer Kognition, und sein Zusammenhang mit verwandten Konstrukten. (Diploma thesis), University of Vienna, Vienna. Retrieved from http://othes.univie.ac.at/32223/

Nilsson, K. K., & de Lopez, K. J. (2016). Theory of mind in children with specific language impairment: A systematic review and meta-analysis. *Child Development, 87*(1), 143-153.

Nixon, C. L., & Watson, A. C. (2001). Family experiences and early emotion understanding. *Merrill-Palmer Quarterly, 47*(2), 300-322.

Norman, W., & Shallice, T. (1986). Attention to action. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self regulation: Advances in research and theory. vol. 4.* (pp. 1-18). New York: Plenum.

Nunner-Winkler, G., & Sodian, B. (1988). Children's understanding of moral emotions. *Child Development, 59*(5), 1323-1328. doi:http://dx.doi.org/10.2307/1130495

O'Brien, M., Weaver, J. M., Nelson, J. A., Calkins, S., Leerkes, E., & Marcovitch, S. (2011). Longitudinal Associations between Children's Understanding of Emotions and Theory of Mind. *Cognition and Emotion, 25*(6), 1074-1086. doi:10.1080/02699931.2010.518417

O'Hare, A. E., Bremner, L., Nash, M., Happe, F., & Pettigrew, L. M. (2009). A clinical assessment tool for advanced theory of mind performance in 5 to 12 year olds. *Journal of Autism and Developmental Disorders, 39*(6), 916-928. doi:http://dx.doi.org/10.1007/s10803-009-0699-2

O'Reilly, J., & Peterson, C. C. (2014). Theory of mind at home: Linking authoritative and authoritarian parenting styles to children's social understanding. *Early Child Development and Care, 184*(12), 1934-1947. doi:http://dx.doi.org/10.1080/03004430.2014.894034

Olson, D. R. (1988). On the origins of beliefs and other intentional states in children. In J. W. Astington, P. L. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 414-426). New York: Cambridge University Press.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*(5719), 255-258. doi:10.1126/science.1107621

Ornaghi, V., Brockmeier, J., & Grazzani, I. (2014). Enhancing social cognition by training children in emotion understanding: A primary school study. *Journal of Experimental Child Psychology, 119*, 26-39. doi:http://dx.doi.org/10.1016/j.jecp.2013.10.005

Ornaghi, V., & Grazzani, I. (2013). The relationship between emotional-state language and emotion understanding: a study with school-age children. *Cognition and Emotion, 27*(2), 356-366. doi: 10.1080/02699931.2012.711745

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review, 97*(3), 315-331. doi:http://dx.doi.org/10.1037/0033-295X.97.3.315

Panksepp, J. (1992). A critical role for "affective neuroscience" in resolving what is basic about basic emotions. *Psychological Review, 99*(3), 554-560. doi:http://dx.doi.org/10.1037/0033-295X.99.3.554

Panksepp, J. (2007). Neurologizing the Psychology of Affects: How Appraisal-Based Constructivism and Basic Emotion Theory Can Coexist. *Perspectives on Psychological Science, 2*(3), 281-296. doi:10.1111/j.1745-6916.2007.00045.x

Panksepp, J., & Watt, D. (2011). What is basic about basic emotions? Lasting lessons from affective neuroscience. *Emotion Review, 3*(4), 387-396. doi:http://dx.doi.org/10.1177/1754073911410741

Panyan, M. V. (1984). Computer technology for autistic students. *Journal of Autism and Developmental Disorders, 14*(4), 375-382.

Patel, V. (2012). Global mental health: from science to action. *Harvard Review of Psychiatry, 20*(1), 6-12. doi: 10.3109/10673229.2012.649108

Pears, K. C., & Moses, L. J. (2003). Demographics, parenting, and theory of mind in preschool children. *Social Development, 12*(1), 1-19. doi:http://dx.doi.org/10.1111/1467-9507.00219

Penades, R., Catalan, R., Rubia, K., Andres, S., Salamero, M., & Gasto, C. (2007). Impaired response inhibition in obsessive compulsive disorder. *European Psychiatry, 22*(6), 404-410. doi:10.1016/j.eurpsy.2006.05.001

Peng, Manli, Johnson, Carl, Pollock, John, Glasspool, Rosalind, & Harris, Paul. (1992). Training young children to acknowledge mixed emotions. *Cognition and Emotion, 6*(5), 387-401.

Pennant, M. E., Loucas, C. E., Whittington, C., Creswell, C., Fonagy, P., Fuggle, P., . . . Expert Advisory, G. (2015). Computerised therapies for anxiety and depression in children and young people: a systematic review and meta-analysis. *Behaviour Research and Therapy, 67*, 1-18. doi:10.1016/j.brat.2015.01.009

Perchtold, C. M., Papousek, I., Koschutnig, K., Rominger, C., Weber, H., Weiss, E. M., & Fink, A. (2018). Affective creativity meets classic creativity in the scanner. *Human Brain Mapping, 39*(1), 393-406. doi:http://dx.doi.org/10.1002/hbm.23851

Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.

Perner, J., Baker, S., & Hutton, D. (1994). Prelief: The conceptual origins of belief and pretence. In C. Lewis, Charlie & P. Mitchell [Eds], *Children's early understanding of* *mind: Origins and development* (pp 261-286). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc

Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Science, 3*(9), 337-344.

Perner, J., & Ruffman, T. (2005). Psychology. Infants' insight into the mind: how deep? *Science, 308*(5719), 214-216. doi:10.1126/science.1111656

Perner, J., Ruffman, T., & Leekam, S. R. (1994). Theory of mind is contagious: You catch it from your sibs. *Child Development, 65*(4), 1228-1238.

Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that": Attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology, 39*(3), 437-471.

Peskin, J., & Ardino, V. (2003). Representing the Mental World in Children's Social Behavior: Playing Hide-and-Seek and Keeping a Secret. *Social Development, 12*(4), 496-512.

Petermann, F., & Petermann, U. (Eds.). (2011). *Wechsler Intelligence Scale for Children (WISC-IV)*. Frankfurt: Pearson Assessment.

Petermann, F., & Petermann, U. (Eds.). (2011). *Wechsler Intelligence Scale for Children - Fourth Edition. Deutschsprachige Adaption nach D. Wechsler. Manual 1: Grundlagen, Testauswertung und Interpretation*. Frankfurt: Pearson Assessment.

Peterson, C. C. (2000). Influence of siblings' perspectives on theory of mind. *Cognitive Development, 15*(4), 435-455. doi:http://dx.doi.org/10.1016/S0885-2014%2801%2900040-5

Peterson, C. C., & Siegal, M. (1999). Representing inner worlds: Theory of mind in autistic, deaf, and normal hearing children. *Psychological Science, 10*(2), 126-129.

Peterson, C. C., Wellman, H. M., & Liu, D. (2005). Steps in Theory-of-Mind Development for Children with Deafness or Autism. *Child Development, 76*(2), 502-517.

Peterson, E., & Miller, S. F. (2012). The Eyes Test as a Measure of Individual Differences: How much of the Variance Reflects Verbal IQ? *Frontiers in Psychology*, 3, 220. doi:10.3389/fpsyg.2012.00220

Phillips, A. T., & Wellman, H. M. (2005). Infants' understanding of object-directed action. *Cognition, 98*(2), 137-155.

Piaget, J., & Inhelder, B. (1956). *The child's conception of space*. London, England: Routledge & Kegan Paul.

Piaget, J. & Inhelder, B. (1969). *The Psychology of the child*. New York: Basic Books, Inc.

Pillow, B. H. (1991). Children's understanding of biased social cognition. *Developmental Psychology, 27*(4), 539-551. doi: 10.1037/0012-1649.27.4.539

Pillow, B. H., & Weed, S. T. (1995). Children's understanding of biased interpretation: Generality and limitations. *British Journal of Developmental Psychology, 13*(4), 347-366. doi: 10.1111/j.2044-835X.1995.tb00685.x

Platt, J. J., Spivack, G., 1975. *Manual for the means–ends problemsolving procedure (MEPS): a measure of interpersonal cognitive problem-solving skills*. Hahnemann Community Mental Health/Mental Retardation Center, Philadelphia.

Polak, A., & Harris, P. L. (1999). Deception by young children following noncompliance. *Developmental Psychology, 35*(2), 561-568.

Pons, F., Harris, P., & Doudin, P-A. (2002). Teaching emotion understanding. *European Journal of Psychology of Education, 17*(3), 293-304.

Pons, F., Lawson, J., Harris, P. L., & de Rosnay, M. (2003). Individual differences in children's emotion understanding: effects of age and language. *Scandinavian Journal of Psychology, 44*(4), 347-353.

Pons, F., Harris, P. L., & de Rosnay, M. (2004). Emotion comprehension between 3 and 11 years: Developmental periods and hierarchical organization. *European Journal of Developmental Psychology, 1*(2), 127-152. doi: 10.1080/17405620344000022

Pons, F., & Harris, P. L. (2005). Longitudinal change and longitudinal stability of individual differences in children's emotion understanding. *Cognition and Emotion, 19*(8), 1158-1174.

Popova, P., Popov, T. G., Wienbruch, C., Carolus, A. M., Miller, G. A., & Rockstroh, B. S. (2014). Changing facial affect recognition in schizophrenia: effects of training on brain dynamics. *Neuroimage Clinical, 6*, 156-165. doi: 10.1016/j.nicl.2014.08.026

Portnoy, D. B., Scott-Sheldon, L. A., Johnson, B. T., & Carey, M. P. (2008). Computer-delivered interventions for health promotion and behavioral risk reduction: A meta-analysis of 75 randomized controlled trials, 1988-2007. *Preventive Medicine: An International Journal Devoted to Practice and Theory, 47*(1), 3-16.

Powers, K. L., Brooks, P. J., Aldrich, N. J., Palladino, M. A., & Alfieri, L. (2013). Effects of video-game play on information processing: A meta-analytic investigation. *Psychonomic Bulletin & Review*. doi: 10.3758/s13423-013-0418-z

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(4), 515-526.

Qualter, P., Barlow, A., & Stylianou, M. S. (2011). Investigating the relationship between trait and ability emotional intelligence and theory of mind. *British Journal of Developmental Psychology, 29*(3), 437-454.

Quintanilla, L., & Gimenez-Dasi, M. (2017). Children's understanding of depreciation in scenarios of envy and modesty. *European Journal of Developmental Psychology, 14*(3), 281-294. doi:http://dx.doi.org/10.1080/17405629.2016.1200029

Rada, P. V., Mark, G. P., & Hoebel, B. G. (1998). Dopamine release in the nucleus accumbens by hypothalamic stimulation-escape behavior. *Brain Research, 782*(1-2), 228-234.

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences, 20*(2), 110-122. doi:10.1016/j.lindif.2009.10.005

Ramdoss, S., Machalicek, W., Rispoli, M., Mulloy, A., Lang, R., & O'Reilly, M. (2012). Computer-based interventions to improve social and emotional skills in individuals with autism spectrum disorders: a systematic review. *Developmental Neurorehabilitation, 15*(2), 119-135. doi: 10.3109/17518423.2011.651655

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*(3), 510-531.

Rawana, J. S., Flett, G. L., McPhie, M. L., Nguyen, H. T., & Norwood, S. J. (2014). Developmental trends in emotion regulation: A systematic review with implications for community mental health. *Canadian Journal of Community Mental Health, 33*(1), 31-44. doi:http://dx.doi.org/10.7870/cjcmh-2014-004

Razza, R. A., & Blair, C. (2009). Associations among false-belief understanding, executive function, and social competence: A longitudinal analysis. *Journal of Applied Developmental Psychology, 30*(3), 332-343.

Reimer, J. (2005, December 15). Total share: 30 years of personal computer market share figures [website]. Retrieve from http://arstechnica.com/features/2005/12/total-share/.

Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology, 33*(1), 12-21. doi:http://dx.doi.org/10.1037/0012-1649.33.1.12

Riccobono, J. A. (1985). School Utilization Study: Availability, Use, and Support of Instructional Media. 1982-83 Final Report (pp. 96): Corporation for Public Broadcasting, 1111 Sixteenth Street, NW, Washington, DC 20036.

Richards, D., & Richardson, T. (2012). Computer-based psychological treatments for depression: a systematic review and meta-analysis. *Clinical Psychology Review, 32*(4), 329-342. doi:10.1016/j.cpr.2012.02.004

Rieffe, C., Meerum Terwogt, M., & Kotronopoulou, K. (2007). Awareness of single and multiple emotions in high-functioning children with autism. *Journal of Autism and Developmental Disorders, 37*(3), 455-465. doi:10.1007/s10803-006-0171-5

Riggs, N. R., Blair, C. B., & Greenberg, M. T. (2003). Concurrent and 2-year longitudinal relations between executive function and the behavior of 1st and 2nd grade children. *Child Neuropsychology, 9*(4), 267-276. doi:10.1076/chin.9.4.267.23513

Roberts, W., & Strayer, J. (1996). Empathy, Emotional Expressiveness, and Prosocial Behavior. *Child Development, 67*(2), 449-470. doi:10.1111/j.1467-8624.1996.tb01745.x

Röcke, C., & Grühm, D. (2003). *German Translation of the PANAS-X*. Unpublished manuscript. Free University Berlin. Retrieved from: http://www4.ncsu.edu/~dgruehn/page7/page10/files/panas-x-german.pdf

Ross, H. S., Recchia, H. E., & Carpendale, J. I. (2005). Making Sense of Divergent Interpretations of Conflict and Developing an Interpretive Understanding of Mind. *Journal of Cognition and Development, 6*(4), 571-592. doi: 10.1207/s15327647jcd0604_7

Rothman, K. J. (1990). No adjustments are needed for multiple comparisons. *Epidemiology, 1*(1), 43-46.

Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P., & Posner, M. I. (2004). Development of attentional networks in childhood. *Neuropsychologia, 42*(8), 1029-1040. doi: 10.1016/j.neuropsychologia.2003.12.012

Ruffman, T., Garnham, W., Import, A., & Connolly, D. (2001). Does eye gaze indicate implicit knowledge of false belief? Charting transitions in knowledge. *Journal of Experimental Child Psychology, 80*(3), 201-224.

Ruffman, T., Henry, J. D., Livingstone, V., & Phillips, L. H. (2008). A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging. *Neuroscience and Biobehavioral Reviews, 32*(4), 863-881.

Ruffman, T., Olson, D. R., & Astington, J. W. (1991). Children's understanding of visual ambiguity. *British Journal of Developmental Psychology, 9*(1), 89-102.

Ruffman, T., & Perner, J. (2005). Do infants really understand false belief? *Trends in Cognitive Sciences, 9*(10), 462-463. doi:http://dx.doi.org/10.1016/j.tics.2005.08.001

Ruffman, T., Perner, J., Naito, M., Parkin, L., & Clements, W. A. (1998). Older (but not younger) siblings facilitate false belief understanding. *Developmental Psychology, 34*(1), 161-174. doi:http://dx.doi.org/10.1037/0012-1649.34.1.161

Ruffman, T., Slade, L., & Crowe, E. (2002). The relation between children's and mothers' mental state language and theory-of-mind understanding. *Child Development, 73*(3), 734-751. doi:http://dx.doi.org/10.1111/1467-8624.00435

Ruiz-Aranda, D., Castillo, R., Salguero, J. M., Cabello, R., Fernandez-Berrocal, P., & Balluerka, N. (2012). Short- and midterm effects of emotional intelligence training on adolescent mental health. *Journal of Adolescent Health, 51*(5), 462-467. doi: 10.1016/j.jadohealth.2012.02.003

Russell, E. W. (2012). 1 - Introductory Scientific History to Neuropsychological Assessment. In W. E. Russel (Ed.), *The Scientific Foundation of Neuropsychological Assessment* (pp. 3-30). Oxford: Elsevier.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161-1178. doi:http://dx.doi.org/10.1037/h0077714

Russell, J. A. (1990). The preschooler's understanding of the causes and consequences of emotion. *Child Development, 61*(6), 1872-1881. doi:http://dx.doi.org/10.2307/1130843

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology, 76*(5), 805-819. doi:http://dx.doi.org/10.1037/0022-3514.76.5.805

Russell, J., Mauthner, N., Sharpe, S., & Tidswell, T. (1991). The "windows task" as a measure of strategic deception in preschoolers and autistic subjects. *British Journal of Developmental Psychology, 9*(2), 331-349. doi:http://dx.doi.org/10.1111/j.2044-835X.1991.tb00881.x

Russell, J. A., & Widen, S. C. (2002). Words versus faces in evoking preschool children's knowledge of the causes of emotions. *International Journal of Behavioral Development, 26*(2), 97-103.

Rutherford, M., Baron-Cohen, S., & Wheelwright, S. (2002). Reading the mind in the voice: A study with normal adults and adults with Asperger syndrome and high functioning autism. *Journal of Autism and Developmental Disorders, 32*(3), 189-194

Rydell, A.-M., Berlin, L., & Bohlin, G. (2003). Emotionality, emotion regulation, and adaptation among 5- to 8-year-old children. *Emotion, 3*(1), 30-47. doi:http://dx.doi.org/10.1037/1528-3542.3.1.30

Ryskin, R. A., & Brown-Schmidt, S. (2014). Do Adults Show a Curse of Knowledge in False-Belief Reasoning? A Robust Estimate of the True Effect Size. *PLOS ONE, 9*(3), e92406. doi:10.1371/journal.pone.0092406

Saarni, C. and Harris, P. L., editors (1989). *Children's understanding of emotion*. Cambridge University Press, New York.

Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The Development of Executive Functioning and Theory of Mind: A Comparison of Chinese and U.S. Preschoolers. *Psychological Science, 17*(1), 74-81. doi:http://dx.doi.org/10.1111/j.1467-9280.2005.01667.x

Sacks, S., Fisher, M., Garrett, C., Alexander, P., Holland, C., Rose, D., . . . Vinogradov, S. (2013). Combining computerized social cognitive training with neuroplasticity-based auditory training in schizophrenia. *Clinical Schizophrenia and Related Psychoses, 7*(2), 78-86A. doi: 10.3371/csrp.safi.012513

Salen, K., & Zimmerman, E. (2003). *Rules of Play: Game Design Fundamentals*: The MIT Press.

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance, 36*(5), 1255-1266.

Saunders, N., Downham, R., Turman, B., Kropotov, J., Clark, R., Yumash, R., & Szatmary, A. (2015). Working memory training with tDCS improves behavioral and neurophysiological symptoms in pilot group with post-traumatic stress disorder (PTSD) and with poor working memory. *Neurocase, 21*(3), 271-278. doi: 10.1080/13554794.2014.890727

Schellig, D., Drechsler, R., Heinemann, D., & Sturm, W. (2009). *Handbuch neuropsychologischer Testverfahren. Band 1: Aufmerksamkeit, Gedaechtnis, exekutive Funktionen.* (Vol. 1). Goettingen: Hogrefe.

Schneider, D., Bayliss, A. P., Becker, S. I., & Dux, P. E. (2012). Eye movements reveal sustained implicit processing of others' mental states. *Journal of Experimental Psychology: General, 141*(3), 433-438.

Schneider, J. (2012). Computerised CBT for common mental disorders: RCT of a workplace intervention. Retrieved from the British Occupational Health Research Foundation: http://www.bohrf.org.uk/downloads/Computerised_CBT-Sep2012.pdf

Schonert-Reichl, K. A., Smith, V., Zaidman-Zait, A., & Hertzman, C. (2012). Promoting children's prosocial behaviors in school: Impact of the "Roots of Empathy" program on the social and emotional competence of school-aged children. *School Mental Health, 4*(1), 1-21.

Schultz, D., Izard, C. E., Ackerman, B. P., & Youngstrom, E. A. (2001). Emotion knowledge in economically disadvantaged children: Self-regulatory antecedents and relations to social difficulties and withdrawal. *Development and Psychopathology, 13*(1), 53-67.

Schultz, D., Izard, C. E., & Bear, G. (2004). Children's emotion processing: relations to emotionality and aggression. *Developmental Psychopathology, 16*(2), 371-387.

Schultz, D., Trentacosta, C., Izard, C. E., Leaf, P., & Mostow, A. (2004). *Children's emotion processing: The development of the Assessment of Children's Emotion Skills (ACES)*. Manuscript submitted for publication.

Schutte, N. S., Malouff, J. M., & Thorsteinsson, E. B. (2013). Increasing emotional intelligence through training: Current status and future directions. *The International Journal of Emotional Education, 5*(1), 56-72.

Schweiger, N. (2015). *Serious game for diagnosis and therapy in emotion recognition* (master's thesis), University of Vienna, Vienna. Retrieved from http://othes.univie.ac.at/36930/

Sellars, W. (1956). "Empiricism and the philosophy of mind." In *Science, Perception and Reality*. London and New York, Routledge & Kegan Paul, 1963, 127–96.

Senn, T. E., Espy, K. A., & Kaufmann, P. M. (2004). Using path analysis to understand executive function organization in preschool children. *Developmental Neuropsychology, 26*(1), 445-464. doi:10.1207/s15326942dn2601_5

Shahaeian, A., Peterson, C. C., Slaughter, V., & Wellman, H. M. (2011). Culture and the sequence of steps in theory of mind development. *Developmental Psychology, 47*(5), 1239-1247.

Shakoor, S., Jaffee, S. R., Bowes, L., Ouellet-Morin, I., Andreou, P., Happe, F., . . . Arseneault, L. (2012). A prospective longitudinal study of children's theory of mind and adolescent involvement in bullying. *Journal of Child Psychology and Psychiatry, 53*(3), 254-261. doi:http://dx.doi.org/10.1111/j.1469-7610.2011.02488.x

Shallice T, Burgess P (1991) Higher-order cognitive impairments and frontal lobe lesions in man. In H.S. Levin, H.M. Eisenberg, A.L. Benton (Eds.), *Frontal lobe functions and dysfunctions* (pp 125–137). New York: Oxford University Press.

Shields, A., & Cicchetti, D. (1998). Reactive aggression among maltreated children: The contributions of attention and emotion dysregulation. *Journal of Clinical Child Psychology, 27*(4), 381-395.

Silver, M., & Oakes, P. (2001). Evaluation of a new computer intervention to teach people with autism or Asperger syndrome to recognize and predict emotions in others. *Autism, 5*(3), 299-316.

Silver, H., Goodman, C., Knoll, G., & Isakov, V. (2004). Brief emotion training improves recognition of facial emotions in chronic schizophrenia. A pilot study. *Psychiatry Research, 128*(2), 147-154. doi: 10.1016/j.psychres.2004.06.002

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. (2016). Do "brain-training" programs work? *Psychological Science in the Public Interest, 17*(3), 103-186. doi:http://dx.doi.org/10.1177/1529100616661983

Sklad, M., Diekstra, R., Ritter, M. D., Ben, J., & Gravesteijn, C. (2012). Effectiveness of school-based universal social, emotional, and behavioral programs: Do they enhance students' development in the area of skill, behavior, and adjustment? *Psychology in the Schools, 49*(9), 892-909. doi: 10.1002/pits.21641

Slaughter, V. (1998). Children's understanding of pictorial and mental representations. *Child Development*, *69*, 321–332.

Slaughter, V., Dennis, M. J., & Pritchard, M. (2002). Theory of mind and peer acceptance in preschool children. *British Journal of Developmental Psychology, 20*(4), 545-564.

Smith, H. (2011). The effects of a drama-based language intervention on the development of theory of mind and executive function in urban kindergarten children. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 72*(5-B), 3125.

Smith, J. P., Glass, D. J., & Fireman, G. (2015). The understanding and experience of mixed emotions in 3-5-year-old children. *The Journal of Genetic Psychology: Research and Theory on Human Development, 176*(2), 65-81. doi:http://dx.doi.org/10.1080/00221325.2014.1002750

Solomon, M., Goodlin-Jones, B. L., & Anders, T. F. (2004). A Social Adjustment Enhancement Intervention for High Functioning Autism, Asperger's Syndrome, and Pervasive Developmental Disorder NOS. *Journal of Autism and Developmental Disorders, 34*(6), 649-668.

Sommerville, J. A., Bernstein, D. M., & Meltzoff, A. N. (2013). Measuring beliefs in centimeters: Private knowledge biases preschoolers' and adults' representation of others' beliefs. *Child Development, 84*(6), 1846-1854.

Sorce, J. F., Emde, R. N., Campos, J. J., & Klinnert, M. D. (1985). Maternal emotional signaling: Its effect on the visual cliff behavior of 1-year-olds. *Developmental Psychology, 21*(1), 195-200. doi:http://dx.doi.org/10.1037/0012-1649.21.1.195

Southam-Gerow, M. A., & Kendall, P. C. (2002). Emotion regulation and understanding: Implications for child psychopathology and therapy. *Clinical Psychology Review, 22*(2), 189-222.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science, 18*, 587 - 592.

Spek, A. A., Scholte, E. M., & Van Berckelaer-Onnes, I. A. (2010). Theory of mind in adults with HFA and Asperger syndrome. *Journal of Autism and Developmental Disorders, 40*(3), 280-289.

Sporer, S. L. (2004). .Reality monitoring and the detection of deception. In P.-A. Granhag and L. Stromwall, *Deception Detection in Forensic Contexts* (pp. 64–102). Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511490071

Sprung, M. (2008). Unwanted intrusive thoughts and cognitive functioning in kindergarten and young elementary school-age children following Hurricane Katrina. *Journal of Clinical Child and Adolescent Psychology, 37*(3), 575-587. doi:10.1080/15374410802148236

Sprung, M., & Harris, P. L. (2010). Intrusive thoughts and young children's knowledge about thinking following a natural disaster. *Journal of Child Psychology and Psychiatry, 51*(10), 1115-1124. doi:10.1111/j.1469-7610.2010.02273.x

Sprung, M., Eder, A., Helmreich, V., Maier, J., Markova, G., & Leyrer, J. (2012). *Manual for the Test of Understanding of Cognitive Activities (TUCA)*. Unpublished manuscript, Division of Clinical Child and Adolescent Psychology, Faculty of Psychology, University of Vienna, Vienna, Austria.

Sprung, M., Münch, H. M., Harris, P. L., Ebesutani, C., & Hofmann, S. G. (2015). Children's emotion understanding: A meta-analysis of training studies. *Developmental Review*. doi:http://dx.doi.org/10.1016/j.dr.2015.05.001

Sprung, M., Leyrer, J., Hayward, E., Ganglmayer, K., Neudorfer, J., Schmidt, A. K., Walczak, A., Homer, B. (2012). *Flexibility and Automaticity of Social Cognition*

(FASC) Vienna. Unpublished manuscript, Division of Clinical Child and Adolescent Psychology, Faculty of Psychology, University of Vienna, Vienna, Austria.

Statistik Austria (2016, February 10). Retrieved from http://www.statistik.at/web_de/statistiken/ menschen_und_gesellschaft/ bevoelkerung/ sterbetafeln/index.html

Steadman, J., Boska, C., Lee, C., Lim, X.-S., & Nichols, N. (2014). Using Popular Commercial Video Games in Therapy with Children and Adolescents. *Journal of Technology in Human Services, 32*(3), 201-219. doi:10.1080/15228835.2014.930680

Steele, H., Steele, M., Croft, C., & Fonagy, P. (1999). Infant-mother attachment at one year predicts children's understanding of mixed emotions at six years. *Social Development, 8*(2), 161-178.

Stein, N. L., & Jewett, J. L. (1986). A conceptual analysis of the meaning of negative emotions: Implications for a theory of development. In C. E. Izard & P. B. Read [Eds], *Measuring emotions in infants and children*, *Vol* 2 (pp 238-267). New York, NY: Cambridge University Press

Stein, N. L., & Levine, L. (1987). Thinking about feelings: The development and organization of emotional knowledge. In R. E. Snow & M. Farr (Eds.), *Aptitude, learning, and instruction: Vol. 3. Cognition, conation, and affect* (pp.165–198). Hillsdale, NJ: Erlbaum.

Sternheimer, K. (2007). Do Video Games Kill? *Contexts, 6*(1), 13-17. doi:10.1525/ctx.2007.6.1.13

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience, 10*(5), 640-656.

Stuss, D. T., & Alexander, M. P. (2007). Is there a dysexecutive syndrome? *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 362*(1481), 901-915. doi:10.1098/rstb.2007.2096

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science, 18*(7), 580-586. doi:http://dx.doi.org/10.1111/j.1467-9280.2007.01943.x

Surtees, A. D., Butterfill, S. A., & Apperly, I. A. (2012). Direct and indirect measures of level-2 perspective-taking in children and adults. *British Journal of Developmental Psychology, 30*(1), 75-86.

Sutton, J., Smith, P., & Swettenham, J. (1999a). Social cognition and bullying: Social inadequacy or skilled manipulation? *British Journal of Developmental Psychology, 17*(3), 435-450.

Sutton, J., Smith, P. K., & Swettenham, J. (1999b). Bullying and "theory of mind': A critique of the "social skills deficit' view of anti-social behaviour. *Social Development, 8*(1), 117-127. doi:http://dx.doi.org/10.1111/1467-9507.00083

Sutton, J., Smith, P. K., & Swettenham, J. (1999c). Socially undesirable need not be incompetent: A response to Crick and Dodge. *Social Development, 8*(1), 132-134. doi:http://dx.doi.org/10.1111/1467-9507.00085

Swettenham, J. (1996). Can children with autism be taught to understand false belief using computers? *Journal of Child Psychology and Psychiatry, 37*(2), 157-165.

Talwar, V., & Lee, K. (2008). Social and cognitive correlates of children's lying behavior. *Child Development, 79*(4), 866-881.

Taumoepeau, M., & Ruffman, T. (2006). Mother and infant talk about mental states relates to desire language and emotion understanding. *Child Development, 77*(2), 465-481. doi: 10.1111/j.1467-8624.2006.00882.x

Taylor Tavares, J. V., Clark, L., Cannon, D. M., Erickson, K., Drevets, W. C., & Sahakian, B. J. (2007). Distinct profiles of neurocognitive function in unmedicated unipolar depression and bipolar II depression. *Biological Psychiatry, 62*(8), 917-924. doi:10.1016/j.biopsych.2007.05.034

Tenenbaum, H. R., Alfieri, L., Brooks, P. J., & Dunne, G. (2008). The effects of explanatory conversations on children's emotion understanding. *British Journal of Developmental Psychology, 26*(2), 249-263.

Teper, R., Segal, Z. V., & Inzlicht, M. (2013). Inside the Mindful Mind: How Mindfulness Enhances Emotion Regulation Through Improvements in Executive Control. *Current Directions in Psychological Science, 22*(6), 449-454. doi: Doi 10.1177/0963721413495869

Terwogt, M. M., Koops, W., Oosterhoff, T., & Olthof, T. (1986). Development in Processing of Multiple Emotional Situations. *The Journal of General Psychology, 113*(2), 109-119. doi:10.1080/00221309.1986.9710548

Teuber, H. L. (1972). Unity and diversity of frontal lobe functions. *Acta Neurobiologiae Experimentalis, 32*(2), 615-656.

Tombaugh, T.N., McIntyre, N.J. (1992). The mini-mental Status Examination: A comprehensive Review. *Journal of the American Geriatrics Society. 40, 922–935.* Doi:10.1111/j.1532-5415.1992.tb01992.x.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., . . . Nelson, C. (2009). The NimStim set of facial expressions: judgments from untrained research participants. Psychiatry Res, 168(3), 242-249. doi:10.1016/j.psychres.2008.05.006

Trentacosta, C. J., & Izard, C. E. (2007). Kindergarten children's emotion competence as a predictor of their academic competence in first grade. *Emotion, 7*(1), 77-88. doi: 10.1037/1528-3542.7.1.77

Tsujimoto, S. (2008). The Prefrontal Cortex: Functional Neural Development During Early Childhood. *The Neuroscientist, 14*(4), 345-358.

Tukey, J. W., Ciminera, J. L., & Heyse, J. F. (1985). Testing the Statistical Certainty of a Response to Increasing Doses of a Drug. *Biometrics, 41*(1), 295-301. doi: 10.2307/2530666

Turner, T. J., & Ortony, A. (1992). Basic emotions: Can conflicting criteria converge? *Psychological Review, 99*(3), 566-571. doi:http://dx.doi.org/10.1037/0033-295X.99.3.566

Undeutsch, U. (1967). *Beurteilung der Glaubhaftigkeit von Aussagen [Assessing the Credibility of Testimonies] Handbuch der Psychologie: Forensische Psychologie.* Göttingen: Hogrefe, 26–181.

van Beugen, S., Ferwerda, M., Hoeve, D., Rovers, M. M., Spillekom-van Koulil, S., van Middendorp, H., & Evers, A. W. (2014). Internet-based cognitive behavioral therapy for patients with chronic somatic conditions: a meta-analytic review. *Journal of Medical Internet Research, 16*(3), e88. doi:10.2196/jmir.2777

Velleman, S., Stallard, P., & Richardson, T. (2010). A review and meta-analysis of computerized cognitive behaviour therapy for the treatment of pain in children and adolescents. *Child: Care, Health and Development, 36*(4), 465-472.

Verlinden, M., Veenstra, R., Ghassabian, A., Jansen, P. W., Hofman, A., Jaddoe, V. W. V., . . . Tiemeier, H. (2014). Executive Functioning and Non-Verbal Intelligence as Predictors of Bullying in Early Elementary School. *Journal of Abnormal Child Psychology, 42*(6), 953-966. doi:10.1007/s10802-013-9832-y

Von Salisch, M., Haenel, M., & Freund, P. A. (2013). Emotion understanding and cognitive abilities in young children. *Learning and Individual Differences, 26*, 15-19. doi: http://dx.doi.org/10.1016/j.lindif.2013.04.001

Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes. Cambridge*: MA, Harvard University Press.

Vytal, K., & Hamann, S. (2010). Neuroimaging Support for Discrete Neural Correlates of Basic Emotions: A Voxel-based Meta-analysis. *Journal of Cognitive Neuroscience, 22*(12), 2864-2885.

Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: Differentiation of multimodal information. *Psychological Bulletin, 121*(3), 437-456. doi:http://dx.doi.org/10.1037/0033-2909.121.3.437

Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J., . . . Gershon, R. C. (2013). Cognition assessment using the NIH Toolbox. *Neurology, 80*(11 Suppl 3), S54-64. doi: 10.1212/WNL.0b013e3182872ded

Wellman, H. M. (2014). Making minds: How theory of mind develops. *(2014) Making minds: How theory of mind develops viii, 366 pp New York, NY, US: Oxford University Press; US.*

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development, 72*(3), 655-684.

Wellman, H. M., & Lagattuta, K. H. (2004). Theory of mind for learning and teaching: the nature and role of explanation. *Cognitive Development, 19*(4), 479-497. doi: http://dx.doi.org/10.1016/j.cogdev.2004.09.003

Wellman, H. M., Fang, F., Liu, D., Zhu, L., & Liu, G. (2006). Scaling of Theory-of-Mind Understandings in Chinese Children. *Psychological Science, 17*(12), 1075-1081.

Wellman, H. M., Fang, F., & Peterson, C. C. (2011). Sequential progressions in a theory-of-mind scale: Longitudinal perspectives. *Child Development, 82*(3), 780-792.

Wellman, H. M., Lopez-Duran, S., LaBounty, J., & Hamilton, B. (2008). Infant attention to intentional action predicts preschool theory of mind. *Developmental Psychology, 44*(2), 618-623

Wellman, H. M., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition, 35*(3), 245-275.

WHO. (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. Geneva: World Health Organization.

WHO (2002). *Prevention and Promotion in Mental Health*. Geneva: World Health Organization.

WHO (2004a). *Promoting mental health: concepts, emerging evidence, practice*. Geneva: World Health Organization.

WHO. (2004b). *Prevention of mental disorders: Effective interventions and policy options* . Geneva: World Health Organization.

WHO. (2008a). *Task shifting: global recommendations and guidelines.* Geneva, Switzerland: WHO Press.

WHO. (2008b). *mhGAP Mental Health Gap Action Programme: Scaling up care for mental, neurological, and substance use disorders*. Geneva: Switzerland: WHO Press.

Widen, S. C., & Russell, J. A. (2008). Young children's understanding of others' emotions. In *Handbook of emotions, 3rd ed* (pp. 348-363). New York, NY: Guilford Press; US.

Widen, S. C., & Russell, J. A. (2011). In building a script for an emotion, do preschoolers add its cause before its behavior consequence? *Social Development, 20*(3), 471-485. doi:http://dx.doi.org/10.1111/j.1467-9507.2010.00594.x

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103-128

Wintre, M. G., & Vallance, D. D. (1994). A developmental sequence in the comprehension of emotions: Intensity, multiple emotions, and valence. *Developmental Psychology, 30*(4), 509-514.

Winerman, L. (2012, February). *'A machine for jumping to conclusions'*. Retrieved from http://www.apa.org/monitor/2012/02/conclusions.aspx

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69*(1), 1-34.

Wolwer, W., & Frommann, N. (2011). Social-cognitive remediation in schizophrenia: generalization of effects of the Training of Affect Recognition (TAR). *Schizophrenia Bulletin, 37 Suppl 2*, S63-70. doi: 10.1093/schbul/sbr071

Wright, J. C. (2012). Children's and adolescents' tolerance for divergent beliefs: Exploring the cognitive and affective dimensions of moral conviction in our youth. *British Journal of Developmental Psychology, 30*(4), 493-510. doi:http://dx.doi.org/10.1111/j.2044-835X.2011.02058.x

Yamashita, C., Mizuno, M., Nemoto, T., & Kashima, H. (2005). Social cognitive problem-solving in schizophrenia: Associations with fluency and verbal memory. *Psychiatry Research, 134*(2), 123-129. doi:http://dx.doi.org/10.1016/j.psychres.2004.06.018

Yeh, M., & Weisz, J. R. (2001). Why are we here at the clinic? Parent-child (dis)agreement on referral problems at outpatient treatment entry. *Journal of Consulting and Clinical Psychology, 69*(6), 1018-1025.

Young, A., Perrett, D., Calder, A., Sprengelmeyer, R., Ekman, P. (2002). *Facial Expressions of Emotion – Stimuli and Tests (FEEST). Psychology manual v 1.0.* Bury St Edmunds: Thames Valley Test Company; England.

Youngblade, L. M., & Dunn, J. (1995). Individual differences in young children's pretend play with mother and sibling: Links to relationships and understanding of other people's feelings and beliefs. *Child Development, 66*(5), 1472-1492. doi:http://dx.doi.org/10.2307/1131658

Zaitchik, D. (1990). When representations conflict with reality: The preschooler's problem with false beliefs and "false" photographs. *Cognition, 35*(1), 41-68.

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): a method of assessing executive function in children. *Nature Protocols, 1*(1), 297-301.

Zelazo, P. D., Muller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function in early childhood: VI. The development of executive function: Cognitive complexity and control--revised. *Monographs of the Society for Research in Child Development, 68*(3), 93-119. doi:http://dx.doi.org/10.1111/j.0037-976X.2003.00266.x

Zimmermann, P., & Iwanski, A. (2014). Emotion regulation from early adolescence to emerging adulthood and middle adulthood: Age differences, gender differences, and emotion-specific developmental variations. *International Journal of Behavioral Development, 38*(2), 182-194. doi:http://dx.doi.org/10.1177/0165025413515405

Appendix A: Examples of FASC vignettes and subjects' responses

For the purpose of giving examples the original responses were translated from german and slightly adjusted for readability. All response examples for each vignette below are from the same subject for each age group. Multiple justifications are separated by a slash.

Figure A 1 "Birthday present", an unambiguous nonverbal vignette

Child (8y, female): She is having a birthday party, then a boy comes and she thinks he has a gift for her and she thinks there is a doll inside but instead there is a teddy inside and then she is sad.

Adolescent (12y, female): Well this one girl has a birthday party and she wants a doll for a present. But she gets a teddy from a boy and then she was a bit vexed about it. / And that she didn't get from another kid what she wanted.

Adult (25y, male): Maybe she doesn't want to be impolite, because she is not really happy about the the present and because it is not what she expected. / Oder she was simply surprised, becaue it is something else than she thought or at the end she likes it anyway.

Older adult (85y, female): Because it's not the right present, she wanted something else and now she is disappointed, because the present is not what she wanted.

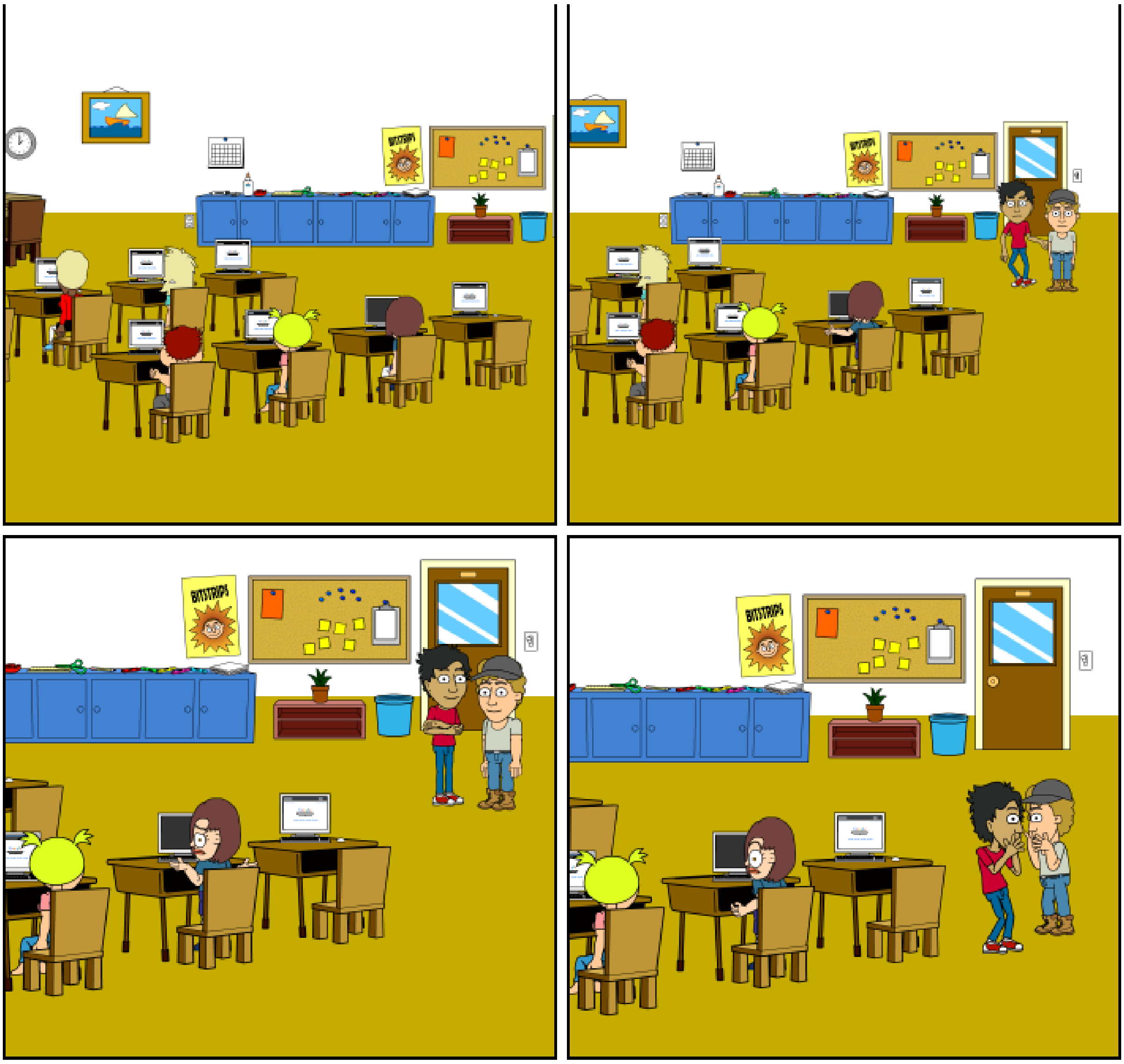

Figure A 2 "PC class", an ambiguous nonverbal vignette

Child (8y, female): Because they talk and I think it is something funny.

Adolescent (12y, female): I think the two boys scoff at someone or so and now they slander about someone else, about a girl. / Maybe someone has a crush on somebody.

Adult (25y, male): I think she is sitting in front of a computer with the screen turned of or somethiing like that and maybe they find that funny. / Or maybe, they are only two, so there are only two of these guys and maybe they want two computers and she is occupying the one that does not work. I don't know, she adressed him with that. / Maybe they bitch about the the person, maybe they know wher and don't like her.

Older adult (85y, female): They are supposed to say or recite something but they are don't agree on it and now they whisper secretly to find out what is was about, because the one sitting on the bench and putting her hand up, I don't know why they whisper secretly.

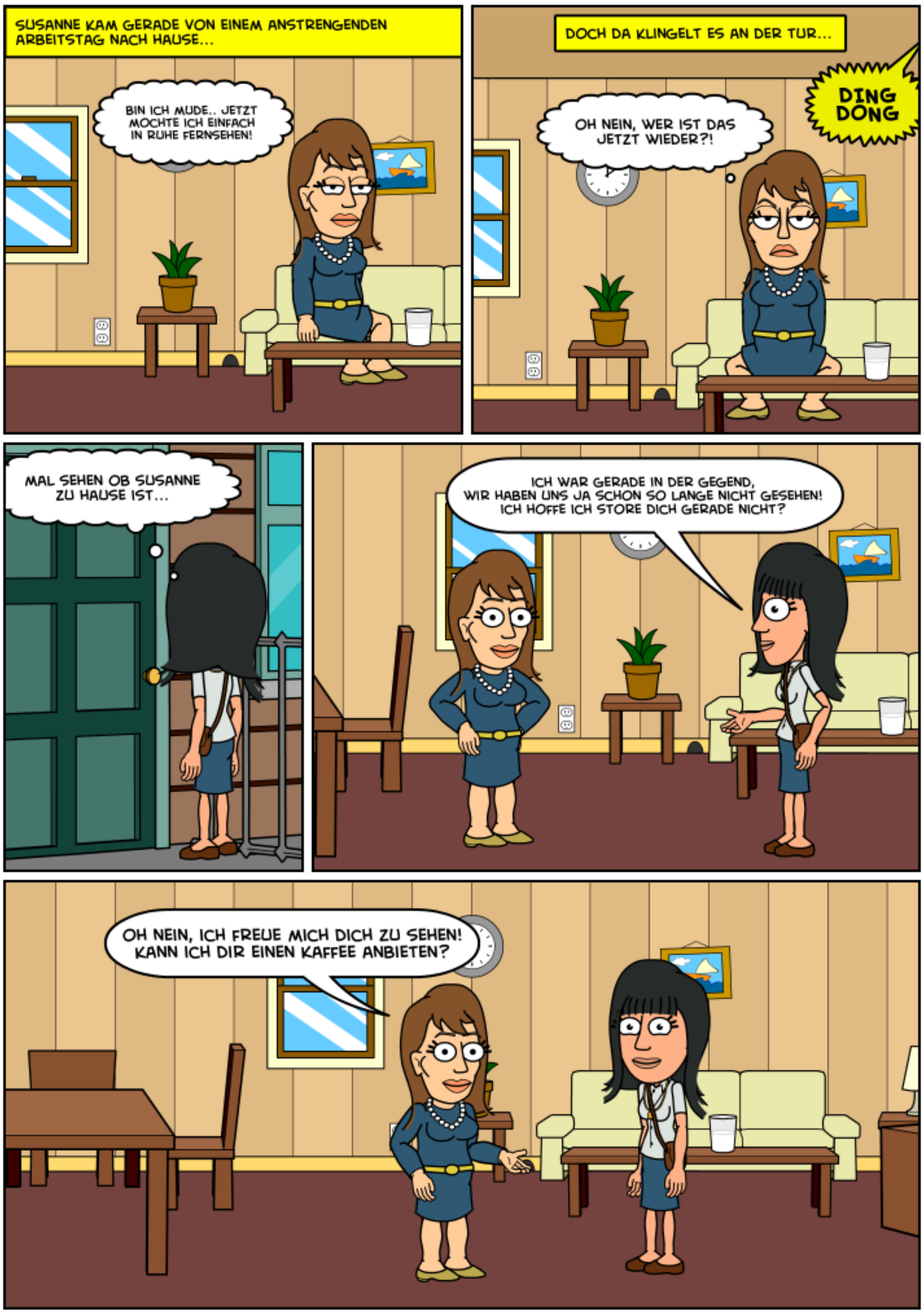

Figure A 3 “Visit”, an unambiguous verbal vignette

Child (8y, female): Because she comes home from a day of work and then she is really tired and then the doorbell rings, I think it's her friend. And then Susannah kindly offers her friend a coffee.

Adolescent (12y, female): Because she comes home from a day of hard work and then the doorbell rings and then they stand together at the door and she asks her, well she was around and asks her, well she offers her coffe. / Because maybe she was tired.

Adult (25y, male): Maybe there are two possibilities. The one that she is really happy, that she hasn't seen the friend for a long time and it doesn't matter that she really wanted to be alone. / Or she doesn't want to be impolite.

Older adult (85y, female): She is just tired and doesn't feel like talking with this woman, but out of courtesy she offers her coffee, but not with pleasure. / First I would say, what might be another reason, well actually the only reason is, although she is a friend, but at that time she is wrong here. But she is friendly and offers her coffee.

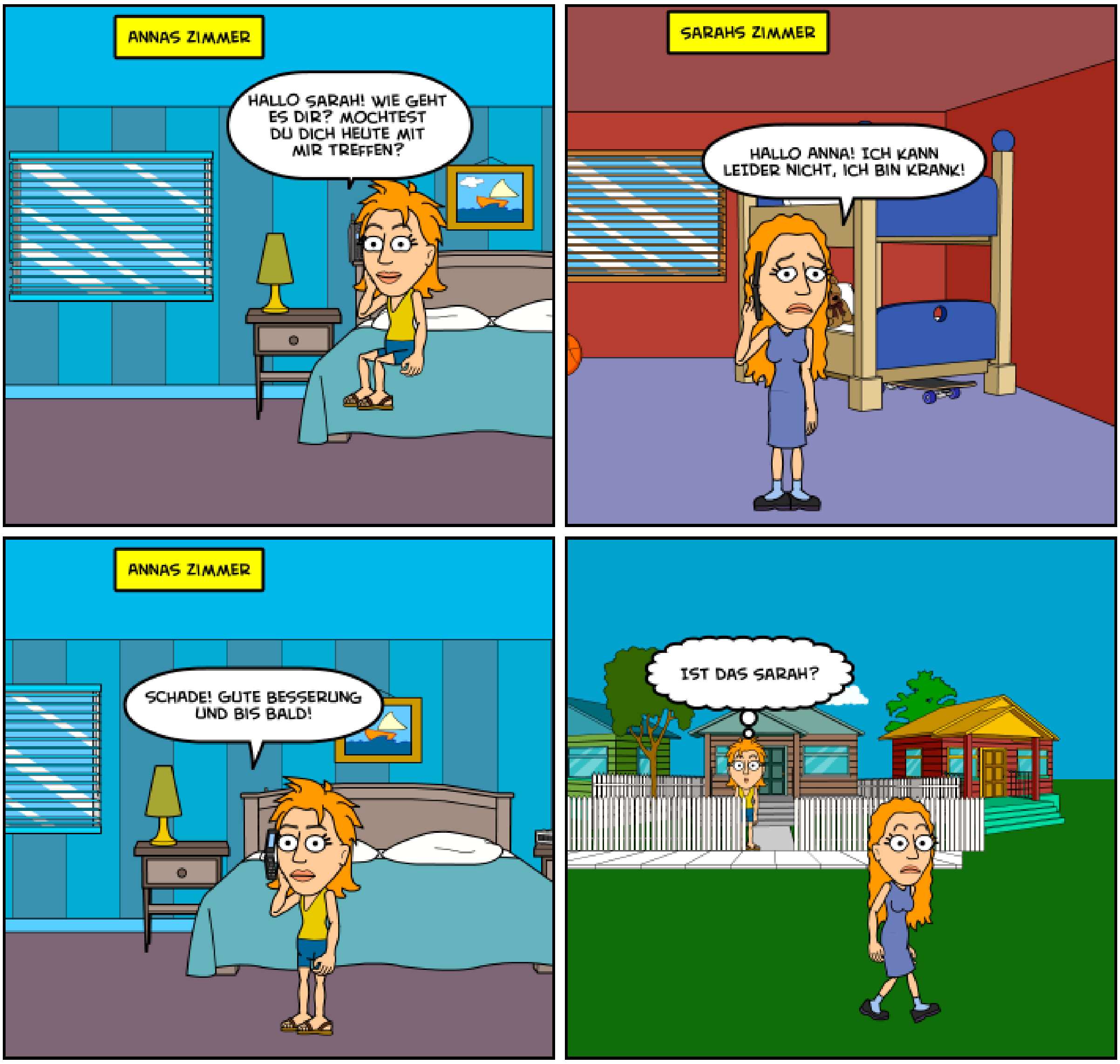

Figure A 4 “Phone call”, an ambiguous verbal vignette

Child (8y, female): She is sick and then suddenly Anna sees her.

Adolescent (12y, female): I think Sarah does not want to meet with Anna. So Anna calls her, if they want to meet and Sarah says no, she is sick and afterwards Anna was angry and she passes by her, well she passes by her house.

Adult (25y, male): Maybe she just doesn't feel like meeting somone. / Or she has another plan she does not want to talk about. / Or she is really sick and only goes to the pharmcy. / Maybe she just has another plan, something secret.

Older adult (85y, female): Because the girl doesn't come, because she has another route. / She doesn't feel like it.

Appendix B: References for the literature reviews in study 1

Strange Stories

Achim, A. M., Ouellet, R., Roy, M.-A., & Jackson, P. L. (2012). Mentalizing in first-episode psychosis. *Psychiatry Research, 196*(2-3), 207-213. doi:http://dx.doi.org/10.1016/j.psychres.2011.10.011

Adrian, J. E., Clemente, R. A., & Villanueva, L. (2007). Mothers' use of cognitive state verbs in picture-book reading and the development of children's understanding of mind: A longitudinal study. *Child Development, 78*(4), 1052-1067. doi:http://dx.doi.org/10.1111/j.1467-8624.2007.01052.x

Ahmed, F. S., & Miller, L. (2011). Executive function mechanisms of theory of mind. *Journal of Autism and Developmental Disorders, 41*(5), 667-678.

Ahmed, F. S., & Miller, L. S. (2013). Relationship between theory of mind and functional independence is mediated by executive function. *Psychology and Aging, 28*(2), 293-303. doi:10.1037/a0031365

Barragan, M., Laurens, K. R., Navarro, J. B., & Obiols, J. E. (2011). 'Theory of Mind', psychotic-like experiences and psychometric schizotypy in adolescents from the general population. *Psychiatry Research, 186*(2-3), 225-231. doi:http://dx.doi.org/10.1016/j.psychres.2010.07.051

Botting, N., & Conti-Ramsden, G. (2008). The role of language, social cognition, and social skill in the functional social outcomes of young adolescents with and without a history of SLI. *British Journal of Developmental Psychology, 26*(2), 281-300.

Brent, E., Rios, P., Happe, F., & Charman, T. (2004). Performance of children with autism spectrum disorder on advanced theory of mind tasks. *Autism, 8*(3), 283-299.

Castelli, I., Pini, A., Alberoni, M., Liverta-Sempio, O., Baglio, F., Massaro, D., . . . Nemni, R. (2011). Mapping levels of theory of mind in Alzheimer's disease: A preliminary study. *Aging & Mental Health, 15*(2), 157-168. doi:http://dx.doi.org/10.1080/13607863.2010.513038

Castellino, N., Bosco, F. M., Marshall, W. L., Marshall, L. E., & Veglia, F. (2011). Mindreading abilities in sexual offenders: An analysis of theory of mind processes. *Consciousness and Cognition: An International Journal, 20*(4), 1612-1624. doi:http://dx.doi.org/10.1016/j.concog.2011.08.011

Cavallini, E., Lecce, S., Bottiroli, S., Palladino, P., & Pagnin, A. (2013). Beyond false belief: Theory of mind in young, young-old, and old-old adults. *The International Journal of Aging & Human Development, 76*(3), 181-198.

Charlton, R. A., Barrick, T. R., Markus, H. S., & Morris, R. G. (2009). Theory of mind associations with other cognitive functions and brain imaging in normal aging. *Psychology and Aging, 24*(2), 338-348. doi:10.1037/a0015225

Charman, T., Carroll, F., & Sturge, C. (2001). Theory of mind, executive function and social competence in boys with ADHD. *Emotional & Behavioural Difficulties, 6*(1), 31-49. doi:http://dx.doi.org/10.1177/136327520100600104

Colvert, E., Rutter, M., Kreppner, J., Beckett, C., Castle, J., Groothues, C., . . . Sonuga-Barke, E. J. (2008). Do theory of mind and executive function deficits underlie the adverse outcomes associated with profound early deprivation?: Findings from the English and Romanian adoptees study. *Journal of Abnormal Child Psychology, 36*(7), 1057-1068. doi:http://dx.doi.org/10.1007/s10802-008-9232-x

Crane, L., Goddard, L., & Pring, L. (2013). Autobiographical memory in adults with autism spectrum disorder: The role of depressed mood, rumination, working memory and theory of mind. *Autism, 17*(2), 205-219. doi:http://dx.doi.org/10.1177/1362361311418690

de Achaval, D., Costanzo, E. Y., Villarreal, M., Jauregui, I. O., Chiodi, A., Castro, M. N., . . . Guinjoan, S. M. (2010). Emotion processing and theory of mind in schizophrenia patients and their unaffected first-degree relatives. *Neuropsychologia, 48*(5), 1209-1215. doi:10.1016/j.neuropsychologia.2009.12.019

Devine, R. T., & Hughes, C. (2013). Silent films and strange stories: theory of mind, gender, and social experiences in middle childhood. *Child Dev, 84*(3), 989-1003. doi:10.1111/cdev.12017

Dyck, M., Ferguson, K., & Shochet, I. (2001). Do autism spectrum disorders differ from each other and from non-spectrum disorders on emotion recognition tests? *European Child & Adolescent Psychiatry, 10*(2), 105-116. doi:http://dx.doi.org/10.1007/s007870170033

Dyck, M. J., Piek, J. P., Hay, D., Smith, L., & Hallmayer, J. (2006). Are Abilities Abnormally Interdependent in Children With Autism? *Journal of Clinical Child and Adolescent Psychology, 35*(1), 20-33. doi:http://dx.doi.org/10.1207/s15374424jccp3501_3

Filippova, E., & Astington, J. W. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child Dev, 79*(1), 126-138.

Gillott, A., Furniss, F., & Walter, A. (2004). Theory of mind ability in children with specific language impairment. *Child Language Teaching and Therapy, 20*(1), 1-11. doi:http://dx.doi.org/10.1191/0265659004ct260oa

Happe, F. G. (1994). An advanced test of theory of mind: understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders, 24*(2), 129-154.

Happe, F. G., Winner, E., & Brownell, H. (1998). The getting of wisdom: Theory of mind in old age. *Developmental Psychology, 34*(2), 358-362.

Hayward, E. O. (2012). Measurement of advanced theory of mind in school-age children: Investigating the validity of a unified construct. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 73*(5-B), 3300.

Henderson, H. A., Zahka, N. E., Kojkowski, N. M., Inge, A. P., Schwartz, C. B., Hileman, C. M., . . . Mundy, P. C. (2009). Self-referenced memory, social cognition, and symptom presentation in autism. *Journal of Child Psychology and Psychiatry, 50*(7), 853-861. doi:http://dx.doi.org/10.1111/j.1469-7610.2008.02059.x

Hileman, C. M., Henderson, H., Mundy, P., Newell, L., & Jaime, M. (2011). Developmental and individual differences on the P1 and N170 ERP components in children with and without autism. *Developmental Neuropsychology, 36*(2), 214-236. doi:http://dx.doi.org/10.1080/87565641.2010.549870

Jolliffe, T., & Baron-Cohen, S. (1999). The Strange Stories Test: A replication with high-functioning adults with autism or Asperger syndrome. *Journal of Autism and Developmental Disorders, 29*(5), 395-406. doi:http://dx.doi.org/10.1023/A:1023082928366

Kaland, N., Moller-Nielsen, A., Smith, L., Mortensen, E. L., Callesen, K., & Gottlieb, D. (2005). The Strange Stories test: A replication study of children and adolescents with Asperger syndrome. *European Child & Adolescent Psychiatry, 14*(2), 73-82.

Loth, E., Happe, F., & Gomez, J. C. (2010). Variety is not the spice of life for people with autism spectrum disorders: Frequency ratings of central, variable and inappropriate aspects of common real-life events. *Journal of Autism and Developmental Disorders, 40*(6), 730-742. doi:http://dx.doi.org/10.1007/s10803-009-0929-7

Meins, E., Fernyhough, C., Johnson, F., & Lidstone, J. (2006). Mind-mindedness in children: Individual differences in internal-state talk in middle childhood. *British Journal of Developmental Psychology, 24*(1), 181-196. doi:http://dx.doi.org/10.1348/026151005X80174

O'Hare, A. E., Bremner, L., Nash, M., Happe, F., & Pettigrew, L. M. (2009). A clinical assessment tool for advanced theory of mind performance in 5 to 12 year olds. *Journal* *of Autism and Developmental Disorders, 39*(6), 916-928. doi:http://dx.doi.org/10.1007/s10803-009-0699-2

Rakoczy, H., Harder-Kasten, A., & Sturm, L. (2012). The decline of theory of mind in old age is (partly) mediated by developmental changes in domain-general abilities. *British Journal of Psychology, 103*(1), 58-72.

Reid, C., Davis, H., Horlin, C., Anderson, M., Baughman, N., & Campbell, C. (2013). The Kids' Empathic Development Scale (KEDS): A multi-dimensional measure of empathy in primary school-aged children. *British Journal of Developmental Psychology, 31*(2), 231-256. doi:http://dx.doi.org/10.1111/bjdp.12002

Rogers, K., Dziobek, I., Hassenstab, J., Wolf, O. T., & Convit, A. (2007). Who cares? Revisiting empathy in Asperger syndrome. *Journal of Autism and Developmental Disorders, 37*(4), 709-715. doi:http://dx.doi.org/10.1007/s10803-006-0197-8

Schneider, D., Slaughter, V. P., Bayliss, A. P., & Dux, P. E. (2013). A temporally sustained implicit theory of mind deficit in autism spectrum disorders. *Cognition, 129*(2), 410-417. doi:http://dx.doi.org/10.1016/j.cognition.2013.08.004

Senju, A., Southgate, V., White, S., & Frit, U. (2009). Mindblind eyes: An absence of spontaneous theory of mind in Asperger syndrome. *Science, 325*(5942), 883-885. doi:http://dx.doi.org/10.1126/science.1176170

Shaked, M., Gamliel, I., & Yirmiya, N. (2006). Theory of mind abilities in young siblings of children with autism. *Autism, 10*(2), 173-187. doi:http://dx.doi.org/10.1177/1362361306062023

Slessor, G., Phillips, L. H., & Bull, R. (2007). Exploring the specificity of age-related differences in theory of mind tasks. *Psychology and Aging, 22*(3), 639-643. doi:http://dx.doi.org/10.1037/0882-7974.22.3.639

Sobel, D. M., Capps, L. M., & Gopnik, A. (2005). Ambiguous figure perception and theory of mind understanding in children with autistic spectrum disorders. *British Journal of Developmental Psychology, 23*(2), 159-174. doi:http://dx.doi.org/10.1348/026151004X20694

Spek, A. A., Scholte, E. M., & Van Berckelaer-Onnes, I. A. (2010). Theory of mind in adults with HFA and Asperger syndrome. *Journal of Autism and Developmental Disorders, 40*(3), 280-289.

Sullivan, S., & Ruffman, T. (2004). Social understanding: How does it fare with advancing years? *British Journal of Psychology, 95*(1), 1-18. doi:http://dx.doi.org/10.1348/000712604322779424

Turkstra, L. S., Williams, W., Tonks, J., & Frampton, I. (2008). Measuring social cognition in adolescents: Implications for students with TBI returning to school. *NeuroRehabilitation, 23*(6), 501-509.

Wang, Y., & Su, Y. (2006). Theory of mind in old adults: The performance on Happe's stories and faux pas stories. *Psychologia: An International Journal of Psychology in the Orient, 49*(4), 228-237. doi:http://dx.doi.org/10.2117/psysoc.2006.228

White, S., Hill, E., Happé, F., & Frith, U. (2009). Revisiting the Strange Stories: Revealing Mentalizing Impairments in Autism. *Child Development, 80*(4), 1097-1117.

Wiffen, B. D., O'Connor, J. A., Gayer-Anderson, C., Marques, T. R., McQueen, G., Happe, F., . . . David, A. S. (2013). "I am sane but he is mad": Insight and illness attributions to self and others in psychosis. *Psychiatry Research, 207*(3), 173-178. doi:http://dx.doi.org/10.1016/j.psychres.2013.01.020

Faux Pas

Ahmed, F. S., & Miller, L. (2011). Executive function mechanisms of theory of mind. *Journal of Autism and Developmental Disorders, 41*(5), 667-678.

Ahmed, F. S., & Miller, L. S. (2013). Relationship between theory of mind and functional independence is mediated by executive function. *Psychology and Aging, 28*(2), 293-303. doi:10.1037/a0031365

Banati, M., Sandor, J., Mike, A., Illes, E., Bors, L., Feldmann, A., . . . Illes, Z. (2010). Social cognition and Theory of Mind in patients with relapsing-remitting multiple sclerosis. *European Journal of Neurology, 17*(3), 426-433. doi:http://dx.doi.org/10.1111/j.1468-1331.2009.02836.x

Banerjee, R. (2000). The development of an understanding of modesty. *British Journal of Developmental Psychology, 18*(4), 499-517.

Banerjee, R., Watling, D., & Caputi, M. (2011). Peer relations and the understanding of faux pas: Longitudinal evidence for bidirectional associations. *Child Development, 82*(6), 1887-1905. doi:http://dx.doi.org/10.1111/j.1467-8624.2011.01669.x

Barlow, A., Qualter, P., & Stylianou, M. (2010). Relationships between Machiavellianism, emotional intelligence and theory of mind in children. *Personality and Individual Differences, 48*(1), 78-82. doi:http://dx.doi.org/10.1016/j.paid.2009.08.021

Baron-Cohen, S., O'Riordan, M., Stone, V., Jones, R., & Plaisted, K. (1999). Recognition of faux pas by normally developing children and children with Asperger syndrome or high-functioning autism. *Journal of Autism and Developmental Disorders, 29*(5), 407-418.

Broicher, S. D., Kuchukhidze, G., Grunwald, T., Kramer, G., Kurthen, M., & Jokeit, H. (2012). "Tell me how do I feel" - Emotion recognition and theory of mind in symptomatic mesial temporal lobe epilepsy. *Neuropsychologia, 50*(1), 118-128. doi:http://dx.doi.org/10.1016/j.neuropsychologia.2011.11.005

Channon, S., Drury, H., Gafson, L., Stern, J., & Robertson, M. M. (2012). Judgements of social inappropriateness in adults with Tourette's syndrome. *Cognitive Neuropsychiatry, 17*(3), 246-261. doi:http://dx.doi.org/10.1080/13546805.2011.590689

Costa, A., Peppe, A., Martini, M., Coletta, K., Oliveri, M., Caltagirone, C., & Carlesimo, G. A. (2013). Parkinsonian patients with deficits in the dysexecutive spectrum are impaired on theory of mind tasks. *Behavioural Neurology, 27*(4), 523-533. doi:http://dx.doi.org/10.1155/2013/903127

Costa, A., Torriero, S., Oliveri, M., & Caltagirone, C. (2008). Prefrontal and temporo-parietal involvement in taking others' perspective: TMS evidence. *Behavioural Neurology, 19*(1-2), 71-74. doi:http://dx.doi.org/10.1155/2008/694632

de Sampaio, F. T. P., Soneira, S., Aulicino, A., & Allegri, R. F. (2013). Theory of mind in eating disorders and their relationship to clinical profile. *European Eating Disorders Review, 21*(6), 479-487. doi:http://dx.doi.org/10.1002/erv.2247

Dolan, M., & Fullam, R. (2004). Theory of mind and mentalizing ability in antisocial personality disorders with and without psychopathy. *Psychological Medicine, 34*(6), 1093-1102. doi:http://dx.doi.org/10.1017/S0033291704002028

Farrant, A., Morris, R. G., Russell, T., Elwes, R., Akanuma, N., Alarcon, G., & Koutroumanidis, M. (2005). Social cognition in frontal lobe epilepsy. *Epilepsy & Behavior, 7*(3), 506-516. doi:http://dx.doi.org/10.1016/j.yebeh.2005.07.018

Filippova, E., & Astington, J. W. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child Development, 79*(1), 126-138.

Geraci, A., Signorelli, M., & Aguglia, E. (2012). Reading the mind: A comparative study of out- and inpatients. *Journal of Psychopathology / Giornale di Psicopatologia, 18*(3), 247-250.

Giovagnoli, A. R., Parente, A., Villani, F., Franceschetti, S., & Spreafico, R. (2013). Theory of mind and epilepsy: What clinical implications? *Epilepsia, 54*(9), 1639-1646. doi:http://dx.doi.org/10.1111/epi.12255

Gregory, C., Lough, S., Stone, V., Erzinclioglu, S., Martin, L., Baron-Cohen, S., & Hodges, J. R. (2002). Theory of mind in patients with frontal variant frontotemporal dementia and Alzheimer's disease: Theoretical and practical implications. *Brain: A Journal of Neurology, 125*(4), 752-764. doi:http://dx.doi.org/10.1093/brain/awf079

Hayward, E. O. (2012). Measurement of advanced theory of mind in school-age children: Investigating the validity of a unified construct. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 73*(5-B), 3300.

Herold, R., Feldmann, A., Simon, M., Tenyi, T., Kover, F., Nagy, F., . . . Fekete, S. (2009). Regional gray matter reduction and theory of mind deficit in the early phase of schizophrenia: A voxel-based morphometric study. *Acta Psychiatrica Scandinavica, 119*(3), 199-208. doi:http://dx.doi.org/10.1111/j.1600-0447.2008.01297.x

Hooker, C. I., Bruce, L., Lincoln, S. H., Fisher, M., & Vinogradov, S. (2011). Theory of mind skills are related to gray matter volume in the ventromedial prefrontal cortex in schizophrenia. *Biological Psychiatry, 70*(12), 1169-1178. doi:http://dx.doi.org/10.1016/j.biopsych.2011.07.027

Huepe, D., Riveros, R., Manes, F., Couto, B., Hurtado, E., Cetkovich, M., . . . Ibanez, A. (2012). The relationship of clinical, cognitive and social measures in schizophrenia: A preliminary finding combining measures in probands and relatives. *Behavioural Neurology, 25*(2), 137-150. doi:http://dx.doi.org/10.1155/2012/194159

Kobayakawa, M., Tsuruya, N., & Kawamura, M. (2012). Theory of mind impairment in adult-onset myotonic dystrophy type 1. *Neuroscience Research, 72*(4), 341-346. doi:http://dx.doi.org/10.1016/j.neures.2012.01.005

Li, X., Wang, K., Wang, F., Tao, Q., Xie, Y., & Cheng, Q. (2013). Aging of theory of mind: The influence of educational level and cognitive processing. *International Journal of Psychology, 48*(4), 715-727. doi:http://dx.doi.org/10.1080/00207594.2012.673724

Li, Y.-H., Chiu, M.-J., Yeh, Z.-T., Liou, H.-H., Cheng, T.-W., & Hua, M.-S. (2013). Theory of mind in patients with temporal lobe epilepsy. *Journal of the International Neuropsychological Society, 19*(5), 594-600. doi:http://dx.doi.org/10.1017/S1355617713000143

MacPherson, S. E., Phillips, L. H., & Della Sala, S. (2002). Age, executive function and social decision making: A dorsolateral prefrontal theory of cognitive aging. *Psychology and Aging, 17*(4), 598-609. doi:http://dx.doi.org/10.1037/0882-7974.17.4.598

Martino, D. J., Bucay, D., Butman, J. T., & Allegri, R. F. (2007). Neuropsychological frontal impairments and negative symptoms in schizophrenia. *Psychiatry Research, 152*(2-3), 121-128. doi:http://dx.doi.org/10.1016/j.psychres.2006.03.002

Martino, D. J., Strejilevich, S. A., Fassi, G., Marengo, E., & Igoa, A. (2011). Theory of mind and facial emotion recognition in euthymic bipolar I and bipolar II disorders. *Psychiatry Research, 189*(3), 379-384. doi:http://dx.doi.org/10.1016/j.psychres.2011.04.033

Meier, S. L., Charleston, A. J., & Tippett, L. J. (2010). Cognitive and behavioural deficits associated with the orbitomedial prefrontal cortex in amyotrophic lateral sclerosis. *Brain: A Journal of Neurology, 133*(11), 3444-3457. doi:http://dx.doi.org/10.1093/brain/awq254

Meristo, M., Falkman, K. W., Hjelmquist, E., Tedoldi, M., Surian, L., & Siegal, M. (2007). Language access and theory of mind reasoning: Evidence from deaf children in bilingual and oralist environments. *Developmental Psychology, 43*(5), 1156-1169. doi:http://dx.doi.org/10.1037/0012-1649.43.5.1156

Milders, M., Fuchs, S., & Crawford, J. R. (2003). Neuropsychological impairments and changes in emotional and social behaviour following severe traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology, 25*(2), 157-172. doi:http://dx.doi.org/10.1076/jcen.25.2.157.13642

Milders, M., Ietswaart, M., Crawford, J. R., & Currie, D. (2006). Impairments in theory of mind shortly after traumatic brain injury and at 1-year follow-up. *Neuropsychology, 20*(4), 400-408. doi:http://dx.doi.org/10.1037/0894-4105.20.4.400

Muller, F., Simion, A., Reviriego, E., Galera, C., Mazaux, J.-M., Barat, M., & Joseph, P.-A. (2010). Exploring theory of mind after severe traumatic brain injury. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior, 46*(9), 1088-1099. doi:http://dx.doi.org/10.1016/j.cortex.2009.08.014

Peron, J., Vicente, S., Leray, E., Drapier, S., Drapier, D., Cohen, R., . . . Verin, M. (2009). Are dopaminergic pathways involved in theory of mind? A study in Parkinson's disease. *Neuropsychologia, 47*(2), 406-414. doi:http://dx.doi.org/10.1016/j.neuropsychologia.2008.09.008

Schacher, M., Winkler, R., Grunwald, T., Kraemer, G., Kurthen, M., Reed, V., & Jokeit, H. (2006). Mesial temporal lobe epilepsy impairs advanced social cognition. *Epilepsia, 47*(12), 2141-2146. doi:http://dx.doi.org/10.1111/j.1528-1167.2006.00857.x

Spek, A. A., Scholte, E. M., & Van Berckelaer-Onnes, I. A. (2010). Theory of mind in adults with HFA and Asperger syndrome. *Journal of Autism and Developmental Disorders, 40*(3), 280-289.

Symington, S. H., Paul, L. K., Symington, M. F., Ono, M., & Brown, W. S. (2010). Social cognition in individuals with agenesis of the corpus callosum. *Social Neuroscience, 5*(3), 296-308. doi:http://dx.doi.org/10.1080/17470910903462419

Thoma, P., Winter, N., Juckel, G., & Roser, P. (2013). Mental state decoding and mental state reasoning in recently detoxified alcohol-dependent individuals. *Psychiatry Research, 205*(3), 232-240. doi:http://dx.doi.org/10.1016/j.psychres.2012.08.042

Torralva, T., Kipps, C. M., Hodges, J. R., Clark, L., Bekinschtein, T., Roca, M., . . . Manes, F. (2007). The relationship between affective decision-making and theory of mind in the frontal variant of fronto-temporal dementia. *Neuropsychologia, 45*(2), 342-349. doi:http://dx.doi.org/10.1016/j.neuropsychologia.2006.05.031

Turkstra, L. S., Williams, W., Tonks, J., & Frampton, I. (2008). Measuring social cognition in adolescents: Implications for students with TBI returning to school. *NeuroRehabilitation, 23*(6), 501-509.

Xi, C., Zhu, Y., Niu, C., Zhu, C., Lee, T. M., Tian, Y., & Wang, K. (2011). Contributions of subregions of the prefrontal cortex to the theory of mind and decision making. *Behavioural Brain Research, 221*(2), 587-593. doi:http://dx.doi.org/10.1016/j.bbr.2010.09.031

Xi, C., Zhu, Y., Zhu, C., Song, D., Wang, Y., & Wang, K. (2013). Deficit of theory of mind after temporal lobe cerebral infarction. *Behavioral and Brain Functions Vol 9 2013, ArtID 15, 9*. doi:http://dx.doi.org/10.1186/1744-9081-9-15

Yu, R.-L., Wu, R.-M., Chiu, M.-J., Tai, C.-H., Lin, C.-H., & Hua, M.-S. (2012). Advanced theory of mind in patients at early stage of Parkinson's disease. *Parkinsonism & Related Disorders, 18*(1), 21-24. doi:http://dx.doi.org/10.1016/j.parkreldis.2011.08.003

Zalla, T., Barlassina, L., Buon, M., & Leboyer, M. (2011). Moral judgment in adults with autism spectrum disorders. *Cognition, 121*(1), 115-126. doi:http://dx.doi.org/10.1016/j.cognition.2011.06.004

Zhang, X., Fung, H. H., Stanley, J. T., Isaacowitz, D. M., & Ho, M. Y. (2013). Perspective taking in older age revisited: A motivational perspective. *Developmental Psychology, 49*(10), 1848-1858. doi:http://dx.doi.org/10.1037/a0031211

Zhu, C.-Y., Lee, T. M., Li, X.-S., Jing, S.-C., Wang, Y.-G., & Wang, K. (2007). Impairments of social cues recognition and social functioning in Chinese people with schizophrenia. *Psychiatry and Clinical Neurosciences, 61*(2), 149-158. doi:http://dx.doi.org/10.1111/j.1440-1819.2007.01630.x

Reading the Mind in the Eyes

Ahmed, F. S., & Miller, L. (2011). Executive function mechanisms of theory of mind. *Journal of Autism and Developmental Disorders, 41*(5), 667-678.

Bailey, P. E., & Henry, J. D. (2008). Growing less empathic with age: Disinhibition of the self-perspective. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 63*(4), P219-P226. doi:http://dx.doi.org/10.1093/geronb/63.4.P219

Bailey, P. E., Henry, J. D., & Von Hippel, W. (2008). Empathy and social functioning in late adulthood. *Aging & Mental Health, 12*(4), 499-503. doi:http://dx.doi.org/10.1080/13607860802224243

Banati, M., Sandor, J., Mike, A., Illes, E., Bors, L., Feldmann, A., . . . Illes, Z. (2010). Social cognition and Theory of Mind in patients with relapsing-remitting multiple sclerosis. *European Journal of Neurology, 17*(3), 426-433. doi:http://dx.doi.org/10.1111/j.1468-1331.2009.02836.x

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism. *Journal of Child Psychology and Psychiatry, 42*(2), 241-251.

Botting, N., & Conti-Ramsden, G. (2008). The role of language, social cognition, and social skill in the functional social outcomes of young adolescents with and without a history of SLI. *British Journal of Developmental Psychology, 26*(2), 281-300.

Brent, E., Rios, P., Happe, F., & Charman, T. (2004). Performance of children with autism spectrum disorder on advanced theory of mind tasks. *Autism, 8*(3), 283-299.

Broicher, S. D., Kuchukhidze, G., Grunwald, T., Kramer, G., Kurthen, M., & Jokeit, H. (2012). "Tell me how do I feel" - Emotion recognition and theory of mind in symptomatic mesial temporal lobe epilepsy. *Neuropsychologia, 50*(1), 118-128. doi:http://dx.doi.org/10.1016/j.neuropsychologia.2011.11.005

Caletti, E., Paoli, R. A., Fiorentini, A., Cigliobianco, M., Zugno, E., Serati, M., . . . Altamura, A. C. (2013). Neuropsychology, social cognition and global functioning among bipolar, schizophrenic patients and healthy controls: preliminary data. *Frontiers in Human Neuroscience, 7*, 661. doi:10.3389/fnhum.2013.00661

Castelli, I., Pini, A., Alberoni, M., Liverta-Sempio, O., Baglio, F., Massaro, D., . . . Nemni, R. (2011). Mapping levels of theory of mind in Alzheimer's disease: A preliminary study. *Aging & Mental Health, 15*(2), 157-168. doi:http://dx.doi.org/10.1080/13607863.2010.513038

Craig, J. S., Hatton, C., Craig, F. B., & Bentall, R. P. (2004). Persecutory beliefs, attributions and theory of mind: Comparison of patients with paranoid delusions, Asperger's syndrome and healthy controls. *Schizophrenia Research, 69*(1), 29-33. doi:http://dx.doi.org/10.1016/S0920-9964%2803%2900154-3

de Achaval, D., Costanzo, E. Y., Villarreal, M., Jauregui, I. O., Chiodi, A., Castro, M. N., . . . Guinjoan, S. M. (2010). Emotion processing and theory of mind in schizophrenia patients and their unaffected first-degree relatives. *Neuropsychologia, 48*(5), 1209-1215. doi:10.1016/j.neuropsychologia.2009.12.019

de Sampaio, F. T. P., Soneira, S., Aulicino, A., & Allegri, R. F. (2013). Theory of mind in eating disorders and their relationship to clinical profile. *European Eating Disorders Review, 21*(6), 479-487. doi:http://dx.doi.org/10.1002/erv.2247

Deshpande, G., Libero, L. E., Sreenivasan, K. R., Deshpande, H. D., & Kana, R. K. (2013). Identification of neural connectivity signatures of autism using machine learning. *Frontiers in Human Neuroscience, 7*.

Dorris, L., Espie, C., Knott, F., & Salt, J. (2004). Mind-reading difficulties in the siblings of people with Asperger's syndrome: evidence for a genetic influence in the abnormal development of a specific cognitive domain. *Journal of Child Psychology and Psychiatry, 45*(2), 412-418. doi:http://dx.doi.org/10.1111/j.1469-7610.2004.00232.x

Farrant, A., Morris, R. G., Russell, T., Elwes, R., Akanuma, N., Alarcon, G., & Koutroumanidis, M. (2005). Social cognition in frontal lobe epilepsy. *Epilepsy & Behavior, 7*(3), 506-516. doi:http://dx.doi.org/10.1016/j.yebeh.2005.07.018

Geraci, A., Signorelli, M., & Aguglia, E. (2012). Reading the mind: A comparative study of out- and inpatients. *Journal of Psychopathology / Giornale di Psicopatologia, 18*(3), 247-250.

Golan, O., Baron-Cohen, S., Hill, J. J., & Rutherford, M. (2007). The 'Reading the Mind in the Voice' Test-Revised: A study of complex emotion recognition in adults with and without autism spectrum conditions. *Journal of Autism and Developmental Disorders, 37*(6), 1096-1106.

Gonzalez-Gadea, M. L., Baez, S., Torralva, T., Castellanos, F. X., Rattazzi, A., Bein, V., . . . Ibanez, A. (2013). Cognitive variability in adults with ADHD and AS: disentangling the roles of executive functions and social cognition. *Research in Developmental Disabilities, 34*(2), 817-830. doi:10.1016/j.ridd.2012.11.009

Gregory, C., Lough, S., Stone, V., Erzinclioglu, S., Martin, L., Baron-Cohen, S., & Hodges, J. R. (2002). Theory of mind in patients with frontal variant frontotemporal dementia and Alzheimer's disease: Theoretical and practical implications. *Brain: A Journal of Neurology, 125*(4), 752-764. doi:http://dx.doi.org/10.1093/brain/awf079

Harrison, A., Tchanturia, K., & Treasure, J. (2010). Attentional bias, emotion recognition, and emotion regulation in anorexia: State or trait? *Biological Psychiatry, 68*(8), 755-761. doi:http://dx.doi.org/10.1016/j.biopsych.2010.04.037

Hayward, E. O. (2012). Measurement of advanced theory of mind in school-age children: Investigating the validity of a unified construct. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 73*(5-B), 3300.

Henderson, H. A., Zahka, N. E., Kojkowski, N. M., Inge, A. P., Schwartz, C. B., Hileman, C. M., . . . Mundy, P. C. (2009). Self-referenced memory, social cognition, and symptom presentation in autism. *Journal of Child Psychology and Psychiatry, 50*(7), 853-861. doi:http://dx.doi.org/10.1111/j.1469-7610.2008.02059.x

Hileman, C. M., Henderson, H., Mundy, P., Newell, L., & Jaime, M. (2011). Developmental and individual differences on the P1 and N170 ERP components in children with and without autism. *Developmental Neuropsychology, 36*(2), 214-236. doi:http://dx.doi.org/10.1080/87565641.2010.549870

Huepe, D., Riveros, R., Manes, F., Couto, B., Hurtado, E., Cetkovich, M., . . . Ibanez, A. (2012). The relationship of clinical, cognitive and social measures in schizophrenia: A preliminary finding combining measures in probands and relatives. *Behavioural Neurology, 25*(2), 137-150. doi:http://dx.doi.org/10.1155/2012/194159

Ibanez, A., Urquina, H., Petroni, A., Baez, S., Lopez, V., do Nascimento, M., . . . Manes, F. (2012). Neural processing of emotional facial and semantic expressions in euthymic bipolar disorder (BD) and its association with theory of mind (ToM). *PLoS ONE, 7*(10). doi:http://dx.doi.org/10.1371/journal.pone.0046877

Kaland, N., Callesen, K., Moller-Nielsen, A., Mortensen, E. L., & Smith, L. (2008). Performance of children and adolescents with Asperger syndrome or high-functioning autism on advanced theory of mind tasks. *Journal of Autism and Developmental Disorders, 38*(6), 1112-1123.

Kandalaft, M. R., Didehbani, N., Cullum, C., Krawczyk, D. C., Allen, T. T., Tamminga, C. A., & Chapman, S. B. (2012). The Wechsler ACS Social Perception subtest: A preliminary comparison with other measures of social cognition. *Journal of Psychoeducational Assessment, 30*(5), 455-465. doi:http://dx.doi.org/10.1177/0734282912436411

Kobayakawa, M., Tsuruya, N., & Kawamura, M. (2012). Theory of mind impairment in adult-onset myotonic dystrophy type 1. *Neuroscience Research, 72*(4), 341-346. doi:http://dx.doi.org/10.1016/j.neures.2012.01.005

Krajmer, P., Spajdel, M., Celec, P., & Ostatnikova, D. (2011). Relationship between salivary testosterone levels and empathizing/systemizing in Slovak boys with Asperger syndrome. *Studia Psychologica, 53*(3), 293-305.

Li, X., Wang, K., Wang, F., Tao, Q., Xie, Y., & Cheng, Q. (2013). Aging of theory of mind: The influence of educational level and cognitive processing. *International Journal of Psychology, 48*(4), 715-727. doi:http://dx.doi.org/10.1080/00207594.2012.673724

Mar, R. A., Oatley, K., Hirsh, J., dela Paz, J., & Peterson, J. B. (2006). Bookworms versus nerds: Exposure to fiction versus non-fiction, divergent associations with social ability, and the simulation of fictional social worlds. *Journal of Research in Personality, 40*(5), 694-712. doi:http://dx.doi.org/10.1016/j.jrp.2005.08.002

Martino, D. J., Strejilevich, S. A., Fassi, G., Marengo, E., & Igoa, A. (2011). Theory of mind and facial emotion recognition in euthymic bipolar I and bipolar II disorders. *Psychiatry Research, 189*(3), 379-384. doi:http://dx.doi.org/10.1016/j.psychres.2011.04.033

Milders, M., Fuchs, S., & Crawford, J. R. (2003). Neuropsychological impairments and changes in emotional and social behaviour following severe traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology, 25*(2), 157-172. doi:http://dx.doi.org/10.1076/jcen.25.2.157.13642

Moor, B. G., Op de Macks, Z. A., Guroglu, B., Rombouts, S. A., Van der Molen, M. W., & Crone, E. A. (2012). Neurodevelopmental changes of reading the mind in the eyes. *Social Cognitive and Affective Neuroscience, 7*(1), 44-52. doi:http://dx.doi.org/10.1093/scan/nsr020

Muller, F., Simion, A., Reviriego, E., Galera, C., Mazaux, J.-M., Barat, M., & Joseph, P.-A. (2010). Exploring theory of mind after severe traumatic brain injury. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior, 46*(9), 1088-1099. doi:http://dx.doi.org/10.1016/j.cortex.2009.08.014

Nietlisbach, G., Maercker, A., Rossler, W., & Haker, H. (2010). Are empathic abilities impaired in posttraumatic stress disorder? *Psychological Reports, 106*(3), 832-844. doi:http://dx.doi.org/10.2466/pr0.106.3.832-844

Norenzayan, A., Gervais, W. M., & Trzesniewski, K. H. (2012). Mentalizing deficits constrain belief in a personal God. *PLoS ONE, 7*(5). doi:http://dx.doi.org/10.1371/journal.pone.0036880

Pardini, M., & Nichelli, P. F. (2009). Age-related decline in mentalizing skills across adult life span. *Experimental Aging Research, 35*(1), 98-106. doi:http://dx.doi.org/10.1080/03610730802545259

Peron, J., Vicente, S., Leray, E., Drapier, S., Drapier, D., Cohen, R., . . . Verin, M. (2009). Are dopaminergic pathways involved in theory of mind? A study in Parkinson's disease. *Neuropsychologia, 47*(2), 406-414. doi:http://dx.doi.org/10.1016/j.neuropsychologia.2008.09.008

Phillips, L. H., MacLean, R. D., & Allen, R. (2002). Age and the understanding of emotions: Neuropsychological and sociocognitive perspectives. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 57*(6), P526-P530.

Riveros, R., Hurtado, E., Escobar, M., Ibanez, A., Manes, F., Cetkovich, M., & Reyes, M. (2010). Context-sensitive social cognition is impaired in schizophrenic patients and their healthy relatives. *Schizophrenia Research, 116*(2-3), 297-298. doi:http://dx.doi.org/10.1016/j.schres.2009.10.017

Roca, M., Torralva, T., Gleichgerrcht, E., Chade, A., Arevalo, G. G., Gershanik, O., & Manes, F. (2010). Impairments in social cognition in early medicated and unmedicated Parkinson disease. *Cognitive and Behavioral Neurology, 23*(3), 152-158. doi:http://dx.doi.org/10.1097/WNN.0b013e3181e078de

Shur, S., Shamay-Tsoory, S. G., & Levkovitz, Y. (2008). Integration of emotional and cognitive aspects of theory of mind in schizophrenia and its relation to prefrontal neurocognitive performance. *Cognitive Neuropsychiatry, 13*(6), 472-490. doi:http://dx.doi.org/10.1080/13546800802490034

Slessor, G., Phillips, L. H., & Bull, R. (2007). Exploring the specificity of age-related differences in theory of mind tasks. *Psychology and Aging, 22*(3), 639-643. doi:http://dx.doi.org/10.1037/0882-7974.22.3.639

Spek, A. A., Scholte, E. M., & Van Berckelaer-Onnes, I. A. (2010). Theory of mind in adults with HFA and Asperger syndrome. *Journal of Autism and Developmental Disorders, 40*(3), 280-289.

Stanford, A. D., Messinger, J., Malaspina, D., & Corcoran, C. M. (2011). Theory of Mind in patients at clinical high risk for psychosis. *Schizophrenia Research, 131*(1-3), 11-17. doi:http://dx.doi.org/10.1016/j.schres.2011.06.005

Sutton, J., Reeves, M., & Keogh, E. (2000). Disruptive behaviour, avoidance of responsibility and theory of mind. *British Journal of Developmental Psychology, 18*(Pt 1), 1-11. doi:http://dx.doi.org/10.1348/026151000165517

Thoma, P., Winter, N., Juckel, G., & Roser, P. (2013). Mental state decoding and mental state reasoning in recently detoxified alcohol-dependent individuals. *Psychiatry Research, 205*(3), 232-240. doi:http://dx.doi.org/10.1016/j.psychres.2012.08.042

Torralva, T., Kipps, C. M., Hodges, J. R., Clark, L., Bekinschtein, T., Roca, M., . . . Manes, F. (2007). The relationship between affective decision-making and theory of mind in the frontal variant of fronto-temporal dementia. *Neuropsychologia, 45*(2), 342-349. doi:http://dx.doi.org/10.1016/j.neuropsychologia.2006.05.031

Torralva, T., Roca, M., Gleichgerrcht, E., Bekinschtein, T., & Manes, F. (2009). A neuropsychological battery to detect specific executive and social cognitive impairments in early frontotemporal dementia. *Brain: A Journal of Neurology, 132*(5), 1299-1309. doi:http://dx.doi.org/10.1093/brain/awp041

Vetter, N. C., Leipold, K., Kliegel, M., Phillips, L. H., & Altgassen, M. (2013). Ongoing development of social cognition in adolescence. *Child Neuropsychology, 19*(6), 615-629. doi:http://dx.doi.org/10.1080/09297049.2012.718324

Voracek, M., & Dressler, S. G. (2006). Lack of correlation between digit ratio (2D:4D) and Baron-Cohen's "Reading the Mind in the Eyes" test, empathy, systemising, and autism-spectrum quotients in a general population sample. *Personality and Individual Differences, 41*(8), 1481-1491. doi:http://dx.doi.org/10.1016/j.paid.2006.06.009

Wolkenstein, L., Schonenberg, M., Schirm, E., & Hautzinger, M. (2011). I can see what you feel, but I can't deal with it: Impaired theory of mind in depression. *Journal of Affective Disorders, 132*(1-2), 104-111. doi:http://dx.doi.org/10.1016/j.jad.2011.02.010

Xi, C., Zhu, Y., Zhu, C., Song, D., Wang, Y., & Wang, K. (2013). Deficit of theory of mind after temporal lobe cerebral infarction. *Behavioral and Brain Functions Vol 9 2013, ArtID 15, 9*. doi:http://dx.doi.org/10.1186/1744-9081-9-15

Appendix C: Additional results for study 1

Table C 1
Distribution statistics of variables used in the analyses over all subjects

							Shapiro-Wilk		
Variable	Skew	*SE*-Skew	Kurt	*SE*-Kurt	*Z*-Skew	*Z*-Kurt	*W*	*df*	*p*
Vocabulary raw	0.14	0.19	-0.93	0.39	0.74	-2.42	0.97	157	0.002
Vocabulary raw OLc	0.18	0.19	-0.91	0.39	0.94	-2.36	0.97	157	0.003
Vocabulary raw OLc (z)	0.02	0.19	-0.35	0.39	0.11	-0.91	0.99	157	0.347
Vocabulary (S)	0.57	0.19	0.80	0.39	2.94	2.07	0.96	157	0.000
Vocabulary (S) OLc	0.43	0.19	0.35	0.39	2.23	0.91	0.97	157	0.001
Digit Span raw	0.60	0.19	0.69	0.39	3.11	1.80	0.97	157	0.002
Digit Span raw OLc	0.58	0.19	0.63	0.39	2.99	1.62	0.97	157	0.002
Digit Span raw OLc (z)	0.30	0.20	0.17	0.39	1.54	0.43	0.99	154	0.082
Digit Span (S)	0.35	0.19	1.03	0.39	1.82	2.68	0.96	157	0.000
Digit Span (S) OLc	0.22	0.19	0.36	0.39	1.11	0.94	0.98	157	0.006
DCCS Score	-0.92	0.19	0.31	0.39	-4.72	0.80	0.92	157	0.000
DCCS Score OLc	-0.71	0.19	-0.27	0.39	-3.68	-0.70	0.93	157	0.000
DCCS Score OLc (z)	-0.28	0.19	-0.31	0.39	-1.44	-0.81	0.99	157	0.109
DCCS (IQ)	-0.66	0.19	-0.54	0.39	-3.41	-1.40	0.93	157	0.000
DCCS (IQ) OLc	-0.64	0.19	-0.58	0.39	-3.28	-1.50	0.94	157	0.000
Flanker Score	-0.67	0.19	-0.65	0.39	-3.43	-1.68	0.91	156	0.000
Flanker Score OLc	-0.70	0.19	-0.57	0.39	-3.60	-1.47	0.91	156	0.000
Flanker Score OLc (z)	-0.46	0.19	-0.27	0.39	-2.39	-0.69	0.97	156	0.003
Flanker (IQ)	-0.56	0.19	-0.91	0.39	-2.86	-2.36	0.92	156	0.000
Flanker (IQ) OLc	-0.55	0.19	-0.90	0.39	-2.82	-2.33	0.93	156	0.000
ASEBA self raw OLc (z)	0.71	0.20	-0.39	0.39	3.58	-0.99	0.93	152	0.000
ASEBA self (T)	0.77	0.20	1.35	0.39	3.91	3.45	0.97	152	0.001
ASEBA self (T) OLc	0.41	0.20	0.00	0.39	2.08	0.00	0.98	152	0.072
ASEBA other raw OLc (z)	-0.01	0.21	0.05	0.41	-0.06	0.11	0.98	137	0.088
ASEBA other (T)	-0.01	0.21	0.05	0.41	-0.06	0.11	0.98	137	0.088
ASEBA other (T) OLc	0.63	0.21	-0.69	0.41	3.02	-1.68	0.92	137	0.000
MSJ	1.78	0.19	5.91	0.39	9.15	15.36	0.88	157	0.000
MSJ v-u	2.17	0.19	7.12	0.39	11.16	18.48	0.82	157	0.000
MSJ v-a	2.04	0.19	9.97	0.39	10.54	25.90	0.87	157	0.000
MSJ n-u	1.46	0.19	3.20	0.39	7.54	8.30	0.89	157	0.000
MSJ n-a	1.13	0.19	1.89	0.39	5.81	4.90	0.92	157	0.000
MSJ OLc	0.98	0.19	1.29	0.39	5.06	3.35	0.94	157	0.000
MSJ v-u OLc	1.04	0.19	1.31	0.39	5.35	3.40	0.92	157	0.000
MSJ v-a OLc	0.69	0.19	0.59	0.39	3.57	1.53	0.96	157	0.000
MSJ n-u OLc	0.97	0.19	1.09	0.39	5.02	2.84	0.93	157	0.000
MSJ n-a OLc	1.07	0.19	2.03	0.39	5.49	5.27	0.93	157	0.000
MSJ OLc (z)	0.50	0.19	-0.43	0.39	2.56	-1.11	0.97	157	0.001
MSJ v-u OLc (z)	0.81	0.19	0.50	0.39	4.18	1.29	0.94	157	0.000
MSJ v-a OLc (z)	0.60	0.19	-0.09	0.39	3.08	-0.24	0.96	157	0.000
MSJ n-u OLc (z)	0.80	0.19	0.65	0.39	4.12	1.68	0.95	157	0.000
MSJ n-a OLc (z)	0.59	0.19	0.38	0.39	3.04	0.99	0.97	157	0.003

Variable	Skew	*SE*-Skew	Kurt	*SE*-Kurt	*Z*-Skew	*Z*-Kurt	Shapiro-Wilk *W*	*df*	*p*
uIST	1.82	0.19	3.84	0.39	9.39	9.97	0.83	157	0.000
uIST v-u	1.57	0.19	2.97	0.39	8.11	7.70	0.87	157	0.000
uIST v-a	2.04	0.19	6.62	0.39	10.51	17.18	0.84	157	0.000
uIST n-u	2.28	0.19	6.52	0.39	11.76	16.93	0.78	157	0.000
uIST n-a	1.75	0.19	3.13	0.39	9.03	8.13	0.82	157	0.000
uIST OLc	1.67	0.19	3.24	0.39	8.58	8.40	0.85	157	0.000
uIST v-u OLc	1.21	0.19	1.42	0.39	6.25	3.69	0.90	157	0.000
uIST v-a OLc	1.27	0.19	1.93	0.39	6.56	5.00	0.91	157	0.000
uIST n-u OLc	2.00	0.19	4.79	0.39	10.29	12.45	0.81	157	0.000
uIST n-a OLc	1.81	0.19	3.86	0.39	9.31	10.02	0.83	157	0.000
uIST OLc (z)	0.83	0.19	-0.24	0.39	4.29	-0.63	0.92	157	0.000
uIST v-u OLc (z)	0.77	0.19	-0.26	0.39	3.96	-0.68	0.93	157	0.000
uIST v-a OLc (z)	0.80	0.19	0.12	0.39	4.11	0.31	0.94	157	0.000
uIST n-u OLc (z)	0.84	0.19	0.17	0.39	4.35	0.44	0.94	157	0.000
uIST n-a OLc (z)	0.62	0.19	-0.30	0.39	3.17	-0.77	0.95	157	0.000
ISTr	2.16	0.19	6.18	0.39	11.12	16.06	0.81	157	0.000
ISTr v-u	1.79	0.19	3.81	0.39	9.22	9.88	0.84	157	0.000
ISTr v-a	3.60	0.19	21.60	0.39	18.58	56.11	0.72	157	0.000
ISTr n-u	2.81	0.20	10.17	0.39	14.42	26.14	0.71	154	0.000
ISTr n-a	4.34	0.20	23.58	0.39	22.03	60.14	0.56	151	0.000
ISTr OLc	1.84	0.19	4.70	0.39	9.50	12.21	0.85	157	0.000
ISTr v-u OLc	0.84	0.19	0.51	0.39	4.32	1.33	0.95	157	0.000
ISTr v-a OLc	1.28	0.19	1.86	0.39	6.61	4.84	0.90	157	0.000
ISTr n-u OLc	1.32	0.20	1.66	0.39	6.77	4.26	0.88	154	0.000
ISTr n-a OLc	1.73	0.20	3.82	0.39	8.79	9.74	0.85	151	0.000
ISTr OLc (z)	0.67	0.19	-0.22	0.39	3.46	-0.58	0.95	157	0.000
ISTr v-u OLc (z)	0.57	0.19	-0.18	0.39	2.95	-0.47	0.97	157	0.001
ISTr v-a OLc (z)	0.50	0.19	-0.58	0.39	2.59	-1.50	0.95	157	0.000
ISTr n-u OLc (z)	0.91	0.20	0.06	0.39	4.65	0.16	0.92	154	0.000
ISTr n-a OLc (z)	0.54	0.20	-0.83	0.39	2.73	-2.11	0.93	151	0.000
IRT	3.22	0.19	18.09	0.39	16.59	46.98	0.75	157	0.000
IRT v-u	2.45	0.19	7.65	0.39	12.64	19.87	0.75	157	0.000
IRT v-a	3.91	0.19	21.78	0.39	20.13	56.57	0.64	157	0.000
IRT n-u	3.27	0.20	14.00	0.39	16.75	36.00	0.66	154	0.000
IRT n-a	2.29	0.20	9.55	0.39	11.58	24.23	0.81	150	0.000
IRT (lg10)	0.00	0.19	0.26	0.39	-0.01	0.67	1.00	157	0.985
IRT v-u (lg10)	0.10	0.19	-0.26	0.39	0.52	-0.68	0.99	157	0.542
IRT v-a (lg10)	0.22	0.19	0.06	0.39	1.11	0.16	0.99	157	0.794
IRT n-u (lg10)	-0.06	0.20	0.80	0.39	-0.31	2.07	0.99	154	0.191
IRT n-a (lg10)	-0.50	0.20	0.80	0.39	-2.54	2.03	0.98	150	0.018
IRT (lg10) OLc	0.12	0.19	-0.01	0.39	0.62	-0.02	1.00	157	0.950
IRT v-u (lg10) OLc	0.14	0.19	-0.26	0.39	0.70	-0.68	0.99	157	0.507
IRT v-a (lg10) OLc	0.27	0.19	0.01	0.39	1.39	0.03	0.99	157	0.347
IRT n-u (lg10) OLc	0.30	0.20	-0.16	0.39	1.55	-0.40	0.99	154	0.372

Variable	Skew	*SE*-Skew	Kurt	*SE*-Kurt	*Z*-Skew	*Z*-Kurt	Shapiro-Wilk *W*	*df*	*p*
IRT n-a (lg10) OLc	-0.16	0.20	-0.28	0.39	-0.82	-0.71	0.99	150	0.407
IRT (lg10) OLc (z)	-0.06	0.19	-0.24	0.39	-0.29	-0.63	1.00	157	0.965
IRT v-u (lg10) OLc (z)	0.02	0.19	-0.36	0.39	0.08	-0.94	0.99	157	0.485
IRT v-a (lg10) OLc (z)	-0.08	0.19	-0.33	0.39	-0.43	-0.86	0.99	157	0.397
IRT n-u (lg10) OLc (z)	0.17	0.20	-0.36	0.39	0.89	-0.93	0.99	154	0.575
IRT n-a (lg10) OLc (z)	-0.14	0.20	-0.13	0.39	-0.69	-0.34	1.00	150	0.948

Note. MSJ = mean mental state justifications, uIST = mean unique internal state terms, IRT = mean initial response time, ISTr = ratio of sum of overall response time to sum of IST, v-u = verbal unambiguous, v-a = verbal ambiguous, n-u = nonverbal unambiguous, n-a = nonverbal ambiguous, OLc = outlier-corrected, r = raw score, S = Wechsler scaled score (*M*=10, *SD*=6), IQ = IQ-normed score (*M*=100, *SD*=15), T = T-score (*M*=50, *SD*=10)

Table C 2
Distribution statistics of variables used in the analyses for children

Variable	Skew	*SE*-Skew	Kurt	*SE*-Kurt	*Z*-Skew	*Z*-Kurt	Shapiro-Wilk *W*	*df*	*p*
Vocabulary raw	1.17	0.37	1.98	0.73	3.13	2.70	0.92	40	0.006
Vocabulary raw OLc	0.22	0.37	-0.45	0.73	0.58	-0.61	0.97	40	0.440
Vocabulary raw OLc (z)	0.22	0.37	-0.45	0.73	0.58	-0.61	0.97	40	0.440
Vocabulary (S)	1.37	0.37	3.96	0.73	3.67	5.40	0.88	40	0.000
Vocabulary (S) OLc	0.34	0.37	0.96	0.73	0.90	1.31	0.95	40	0.086
Digit Span raw	0.63	0.37	1.01	0.73	1.68	1.38	0.95	40	0.075
Digit Span raw OLc	0.23	0.37	-0.19	0.73	0.63	-0.25	0.96	40	0.171
Digit Span raw OLc (z)	0.27	0.38	-0.12	0.75	0.70	-0.16	0.96	38	0.152
Digit Span (S)	0.61	0.37	1.78	0.73	1.62	2.43	0.95	40	0.074
Digit Span (S) OLc	0.34	0.37	0.01	0.73	0.91	0.02	0.97	40	0.289
DCCS Score	-1.61	0.37	2.53	0.73	-4.30	3.46	0.83	40	0.000
DCCS Score OLc	-0.66	0.37	-0.07	0.73	-1.78	-0.10	0.92	40	0.009
DCCS Score OLc (z)	-0.66	0.37	-0.07	0.73	-1.78	-0.10	0.92	40	0.009
DCCS (IQ)	-0.75	0.37	-0.67	0.73	-2.00	-0.92	0.90	40	0.002
DCCS (IQ) OLc	-0.75	0.37	-0.67	0.73	-2.00	-0.92	0.90	40	0.002
Flanker Score	-0.15	0.37	-0.51	0.73	-0.41	-0.70	0.98	40	0.697
Flanker Score OLc	-0.15	0.37	-0.52	0.73	-0.40	-0.70	0.98	40	0.701
Flanker Score OLc (z)	-0.15	0.37	-0.52	0.73	-0.40	-0.70	0.98	40	0.701
Flanker (IQ)	-0.36	0.37	-1.16	0.73	-0.97	-1.58	0.93	40	0.020
Flanker (IQ) OLc	-0.36	0.37	-1.16	0.73	-0.97	-1.58	0.93	40	0.020
ASEBA self raw OLc (z)	0.65	0.38	-0.49	0.74	1.71	-0.66	0.94	39	0.030
ASEBA self (T)	0.06	0.38	-0.46	0.74	0.17	-0.62	0.98	39	0.817
ASEBA self (T) OLc	0.06	0.38	-0.46	0.74	0.17	-0.62	0.98	39	0.817
ASEBA other raw OLc (z)	0.00	0.39	-1.04	0.77	0.01	-1.35	0.97	36	0.337
ASEBA other (T)	0.00	0.39	-1.04	0.77	0.01	-1.35	0.97	36	0.337
ASEBA other (T) OLc	0.64	0.39	-0.63	0.77	1.63	-0.83	0.92	36	0.012
MSJ	1.70	0.37	3.60	0.73	4.54	4.91	0.83	40	0.000

Variable	Skew	*SE*-Skew	Kurt	*SE*-Kurt	*Z*-Skew	*Z*-Kurt	Shapiro-Wilk *W*	*df*	*p*
MSJ v-u	1.83	0.37	4.11	0.73	4.89	5.61	0.79	40	0.000
MSJ v-a	1.26	0.37	2.90	0.73	3.37	3.96	0.90	40	0.001
MSJ n-u	1.59	0.37	4.68	0.73	4.26	6.38	0.85	40	0.000
MSJ n-a	1.27	0.37	2.26	0.73	3.40	3.08	0.89	40	0.001
MSJ OLc	0.32	0.37	-0.77	0.73	0.86	-1.05	0.94	40	0.037
MSJ v-u OLc	0.59	0.37	0.86	0.73	1.58	1.17	0.93	40	0.018
MSJ v-a OLc	0.60	0.37	0.14	0.73	1.60	0.19	0.93	40	0.022
MSJ n-u OLc	1.04	0.37	2.15	0.73	2.78	2.93	0.89	40	0.001
MSJ n-a OLc	0.73	0.37	0.35	0.73	1.94	0.48	0.94	40	0.025
MSJ OLc (z)	0.32	0.37	-0.77	0.73	0.86	-1.05	0.94	40	0.037
MSJ v-u OLc (z)	0.59	0.37	0.86	0.73	1.58	1.17	0.93	40	0.018
MSJ v-a OLc (z)	0.60	0.37	0.14	0.73	1.60	0.19	0.93	40	0.022
MSJ n-u OLc (z)	1.04	0.37	2.15	0.73	2.78	2.93	0.89	40	0.001
MSJ n-a OLc (z)	0.73	0.37	0.35	0.73	1.94	0.48	0.94	40	0.025
uIST	1.24	0.37	1.39	0.73	3.32	1.90	0.88	40	0.001
uIST v-u	0.96	0.37	0.36	0.73	2.57	0.49	0.89	40	0.001
uIST v-a	1.32	0.37	1.95	0.73	3.52	2.66	0.88	40	0.001
uIST n-u	1.61	0.37	2.70	0.73	4.30	3.68	0.84	40	0.000
uIST n-a	0.99	0.37	0.44	0.73	2.63	0.60	0.90	40	0.002
uIST OLc	0.96	0.37	0.16	0.73	2.57	0.22	0.90	40	0.001
uIST v-u OLc	0.96	0.37	0.36	0.73	2.57	0.49	0.89	40	0.001
uIST v-a OLc	1.07	0.37	0.86	0.73	2.87	1.18	0.90	40	0.002
uIST n-u OLc	1.09	0.37	0.85	0.73	2.92	1.15	0.90	40	0.002
uIST n-a OLc	0.99	0.37	0.44	0.73	2.63	0.60	0.90	40	0.002
uIST OLc (z)	0.96	0.37	0.16	0.73	2.57	0.22	0.90	40	0.001
uIST v-u OLc (z)	0.96	0.37	0.36	0.73	2.57	0.49	0.89	40	0.001
uIST v-a OLc (z)	1.07	0.37	0.86	0.73	2.87	1.18	0.90	40	0.002
uIST n-u OLc (z)	1.09	0.37	0.85	0.73	2.92	1.15	0.90	40	0.002
uIST n-a OLc (z)	0.99	0.37	0.44	0.73	2.63	0.60	0.90	40	0.002
ISTr	1.27	0.37	1.47	0.73	3.40	2.00	0.89	40	0.001
ISTr v-u	1.23	0.37	0.92	0.73	3.29	1.25	0.87	40	0.000
ISTr v-a	2.81	0.37	10.91	0.73	7.52	14.88	0.73	40	0.000
ISTr n-u	1.55	0.38	1.87	0.75	4.05	2.49	0.82	38	0.000
ISTr n-a	2.90	0.40	9.42	0.78	7.29	12.11	0.65	35	0.000
ISTr OLc	0.96	0.37	0.12	0.73	2.56	0.17	0.90	40	0.002
ISTr v-u OLc	0.39	0.37	-0.69	0.73	1.04	-0.94	0.94	40	0.027
ISTr v-a OLc	0.78	0.37	-0.55	0.73	2.08	-0.75	0.89	40	0.001
ISTr n-u OLc	0.72	0.38	-0.57	0.75	1.87	-0.76	0.90	38	0.002
ISTr n-a OLc	0.84	0.40	-0.50	0.78	2.10	-0.64	0.88	35	0.001
ISTr OLc (z)	0.96	0.37	0.12	0.73	2.56	0.17	0.90	40	0.002
ISTr v-u OLc (z)	0.39	0.37	-0.69	0.73	1.04	-0.94	0.94	40	0.027
ISTr v-a OLc (z)	0.78	0.37	-0.55	0.73	2.08	-0.75	0.89	40	0.001
ISTr n-u OLc (z)	0.72	0.38	-0.57	0.75	1.87	-0.76	0.90	38	0.002
ISTr n-a OLc (z)	0.84	0.40	-0.50	0.78	2.10	-0.64	0.88	35	0.001

Variable	Skew	*SE*-Skew	Kurt	*SE*-Kurt	*Z*-Skew	*Z*-Kurt	Shapiro-Wilk *W*	*df*	*p*
IRT	1.74	0.37	3.41	0.73	4.64	4.65	0.84	40	0.000
IRT v-u	2.13	0.37	5.22	0.73	5.69	7.12	0.78	40	0.000
IRT v-a	2.95	0.37	11.03	0.73	7.88	15.05	0.69	40	0.000
IRT n-u	2.38	0.38	7.25	0.75	6.21	9.67	0.74	38	0.000
IRT n-a	1.08	0.40	0.11	0.78	2.72	0.14	0.86	35	0.000
IRT (lg10)	-0.09	0.37	0.71	0.73	-0.24	0.97	0.98	40	0.768
IRT v-u (lg10)	0.20	0.37	-0.26	0.73	0.53	-0.35	0.99	40	0.977
IRT v-a (lg10)	0.56	0.37	0.59	0.73	1.49	0.80	0.98	40	0.555
IRT n-u (lg10)	0.28	0.38	-0.18	0.75	0.73	-0.23	0.98	38	0.648
IRT n-a (lg10)	-0.17	0.40	-0.36	0.78	-0.42	-0.46	0.97	35	0.519
IRT (lg10) OLc	0.40	0.37	-0.24	0.73	1.06	-0.32	0.97	40	0.490
IRT v-u (lg10) OLc	0.20	0.37	-0.26	0.73	0.54	-0.36	0.99	40	0.977
IRT v-a (lg10) OLc	0.29	0.37	-0.18	0.73	0.78	-0.24	0.98	40	0.656
IRT n-u (lg10) OLc	0.28	0.38	-0.18	0.75	0.73	-0.24	0.98	38	0.641
IRT n-a (lg10) OLc	-0.18	0.40	-0.33	0.78	-0.44	-0.43	0.97	35	0.520
IRT (lg10) OLc (z)	0.40	0.37	-0.24	0.73	1.06	-0.32	0.97	40	0.490
IRT v-u (lg10) OLc (z)	0.20	0.37	-0.26	0.73	0.54	-0.36	0.99	40	0.977
IRT v-a (lg10) OLc (z)	0.29	0.37	-0.18	0.73	0.78	-0.24	0.98	40	0.656
IRT n-u (lg10) OLc (z)	0.28	0.38	-0.18	0.75	0.73	-0.24	0.98	38	0.641
IRT n-a (lg10) OLc (z)	-0.18	0.40	-0.33	0.78	-0.44	-0.43	0.97	35	0.520

Note. MSJ = mean mental state justifications, uIST = mean unique internal state terms, IRT = mean initial response time, ISTr = ratio of sum of overall response time to sum of IST, v-u = verbal unambiguous, v-a = verbal ambiguous, n-u = nonverbal unambiguous, n-a = nonverbal ambiguous, OLc = outlier-corrected, r = raw score, S = Wechsler scaled score (*M*=10, *SD*=6), IQ = IQ-normed score (*M*=100, *SD*=15), T = T-score (*M*=50, *SD*=10)

Table C 3
Distribution statistics of variables used in the analyses for adolescents

Variable	Skew	*SE*-Skew	Kurt	*SE*-Kurt	*Z*-Skew	*Z*-Kurt	Shapiro-Wilk *W*	*df*	*p*
Vocabulary raw	-0.35	0.37	2.80	0.73	-0.93	3.81	0.95	40	0.067
Vocabulary raw OLc	0.01	0.37	0.02	0.73	0.02	0.02	0.98	40	0.753
Vocabulary raw OLc (z)	0.01	0.37	0.02	0.73	0.02	0.02	0.98	40	0.753
Vocabulary (S)	0.68	0.37	4.05	0.73	1.82	5.52	0.90	40	0.002
Vocabulary (S) OLc	0.24	0.37	0.74	0.73	0.64	1.01	0.96	40	0.201
Digit Span raw	-0.22	0.37	0.21	0.73	-0.59	0.29	0.97	40	0.444
Digit Span raw OLc	-0.22	0.37	0.21	0.73	-0.59	0.29	0.97	40	0.444
Digit Span raw OLc (z)	-0.28	0.38	0.33	0.74	-0.74	0.44	0.97	39	0.350
Digit Span (S)	-0.29	0.37	0.42	0.73	-0.77	0.58	0.97	40	0.440
Digit Span (S) OLc	-0.29	0.37	0.42	0.73	-0.77	0.58	0.97	40	0.440
DCCS Score	0.66	0.37	-0.15	0.73	1.76	-0.21	0.93	40	0.021
DCCS Score OLc	0.66	0.37	-0.15	0.73	1.76	-0.21	0.93	40	0.021
DCCS Score OLc (z)	0.66	0.37	-0.15	0.73	1.76	-0.21	0.93	40	0.021

Variable	Skew	*SE-* Skew	Kurt	*SE-* Kurt	*Z-* Skew	*Z-* Kurt	Shapiro-Wilk *W*	*df*	*p*
DCCS (IQ)	0.50	0.37	-0.40	0.73	1.32	-0.55	0.96	40	0.216
DCCS (IQ) OLc	0.50	0.37	-0.40	0.73	1.32	-0.55	0.96	40	0.217
Flanker Score	-2.15	0.37	7.15	0.73	-5.74	9.75	0.82	40	0.000
Flanker Score OLc	-0.86	0.37	0.31	0.73	-2.31	0.43	0.92	40	0.008
Flanker Score OLc (z)	-0.86	0.37	0.31	0.73	-2.31	0.43	0.92	40	0.008
Flanker (IQ)	-1.36	0.37	1.98	0.73	-3.63	2.70	0.88	40	0.000
Flanker (IQ) OLc	-0.96	0.37	0.14	0.73	-2.56	0.19	0.90	40	0.002
ASEBA self raw OLc (z)	0.79	0.39	0.01	0.77	2.00	0.01	0.92	36	0.012
ASEBA self (T)	0.93	0.39	0.68	0.77	2.37	0.88	0.93	36	0.023
ASEBA self (T) OLc	0.83	0.39	0.40	0.77	2.11	0.52	0.94	36	0.035
ASEBA other raw OLc (z)	0.77	0.40	0.67	0.79	1.90	0.86	0.94	34	0.063
ASEBA other (T)	0.77	0.40	0.67	0.79	1.90	0.86	0.94	34	0.063
ASEBA other (T) OLc	0.99	0.40	-0.06	0.79	2.46	-0.08	0.86	34	0.000
MSJ	1.39	0.37	3.33	0.73	3.71	4.55	0.90	40	0.002
MSJ v-u	2.54	0.37	9.18	0.73	6.79	12.52	0.73	40	0.000
MSJ v-a	0.77	0.37	0.65	0.73	2.06	0.88	0.93	40	0.011
MSJ n-u	1.10	0.37	1.26	0.73	2.94	1.72	0.89	40	0.001
MSJ n-a	0.85	0.37	2.25	0.73	2.28	3.06	0.92	40	0.009
MSJ OLc	0.26	0.37	-0.35	0.73	0.70	-0.48	0.97	40	0.322
MSJ v-u OLc	1.12	0.37	0.72	0.73	2.99	0.99	0.85	40	0.000
MSJ v-a OLc	0.47	0.37	0.00	0.73	1.26	0.00	0.94	40	0.046
MSJ n-u OLc	1.10	0.37	1.26	0.73	2.94	1.72	0.89	40	0.001
MSJ n-a OLc	0.23	0.37	0.78	0.73	0.63	1.07	0.96	40	0.186
MSJ OLc (z)	0.26	0.37	-0.35	0.73	0.70	-0.48	0.97	40	0.322
MSJ v-u OLc (z)	1.12	0.37	0.72	0.73	2.99	0.99	0.85	40	0.000
MSJ v-a OLc (z)	0.47	0.37	0.00	0.73	1.26	0.00	0.94	40	0.046
MSJ n-u OLc (z)	1.10	0.37	1.26	0.73	2.94	1.72	0.89	40	0.001
MSJ n-a OLc (z)	0.23	0.37	0.78	0.73	0.63	1.07	0.96	40	0.186
uIST	1.24	0.37	1.79	0.73	3.31	2.44	0.91	40	0.004
uIST v-u	1.74	0.37	4.96	0.73	4.65	6.76	0.87	40	0.000
uIST v-a	0.82	0.37	0.60	0.73	2.19	0.82	0.94	40	0.046
uIST n-u	0.44	0.37	-0.07	0.73	1.17	-0.10	0.97	40	0.374
uIST n-a	1.44	0.37	3.04	0.73	3.85	4.15	0.89	40	0.001
uIST OLc	0.71	0.37	-0.19	0.73	1.89	-0.26	0.93	40	0.017
uIST v-u OLc	0.64	0.37	-0.38	0.73	1.70	-0.52	0.94	40	0.041
uIST v-a OLc	0.77	0.37	0.44	0.73	2.05	0.59	0.95	40	0.050
uIST n-u OLc	0.44	0.37	-0.07	0.73	1.17	-0.10	0.97	40	0.374
uIST n-a OLc	0.08	0.37	-0.30	0.73	0.22	-0.41	0.98	40	0.761
uIST OLc (z)	0.71	0.37	-0.19	0.73	1.89	-0.26	0.93	40	0.017
uIST v-u OLc (z)	0.64	0.37	-0.38	0.73	1.70	-0.52	0.94	40	0.041
uIST v-a OLc (z)	0.77	0.37	0.44	0.73	2.05	0.59	0.95	40	0.050
uIST n-u OLc (z)	0.44	0.37	-0.07	0.73	1.17	-0.10	0.97	40	0.374
uIST n-a OLc (z)	0.08	0.37	-0.30	0.73	0.22	-0.41	0.98	40	0.761
ISTr	1.80	0.37	3.98	0.73	4.80	5.43	0.84	40	0.000

Variable	Skew	*SE-* Skew	Kurt	*SE-* Kurt	*Z-* Skew	*Z-* Kurt	Shapiro-Wilk *W*	*df*	*p*
ISTr v-u	1.27	0.37	2.57	0.73	3.39	3.50	0.90	40	0.001
ISTr v-a	1.36	0.37	2.03	0.73	3.62	2.76	0.88	40	0.001
ISTr n-u	2.48	0.37	7.20	0.73	6.64	9.83	0.73	40	0.000
ISTr n-a	4.42	0.37	22.45	0.73	11.80	30.62	0.49	40	0.000
ISTr OLc	0.63	0.37	-0.31	0.73	1.68	-0.43	0.93	40	0.015
ISTr v-u OLc	0.67	0.37	0.71	0.73	1.78	0.96	0.95	40	0.082
ISTr v-a OLc	0.65	0.37	-0.21	0.73	1.75	-0.28	0.91	40	0.005
ISTr n-u OLc	0.96	0.37	0.48	0.73	2.57	0.66	0.91	40	0.004
ISTr n-a OLc	0.61	0.37	-0.62	0.73	1.63	-0.85	0.92	40	0.008
ISTr OLc (z)	0.63	0.37	-0.31	0.73	1.68	-0.43	0.93	40	0.015
ISTr v-u OLc (z)	0.67	0.37	0.71	0.73	1.78	0.96	0.95	40	0.082
ISTr v-a OLc (z)	0.65	0.37	-0.21	0.73	1.75	-0.28	0.91	40	0.005
ISTr n-u OLc (z)	0.96	0.37	0.48	0.73	2.57	0.66	0.91	40	0.004
ISTr n-a OLc (z)	0.61	0.37	-0.62	0.73	1.63	-0.85	0.92	40	0.008
IRT	0.53	0.37	-0.42	0.73	1.43	-0.58	0.96	40	0.153
IRT v-u	1.06	0.37	1.06	0.73	2.83	1.44	0.92	40	0.008
IRT v-a	1.05	0.37	1.00	0.73	2.81	1.36	0.91	40	0.005
IRT n-u	2.18	0.37	6.07	0.73	5.83	8.29	0.79	40	0.000
IRT n-a	1.68	0.38	2.37	0.74	4.43	3.20	0.80	39	0.000
IRT (lg10)	-0.47	0.37	-0.01	0.73	-1.26	-0.01	0.97	40	0.444
IRT v-u (lg10)	-0.21	0.37	-0.58	0.73	-0.55	-0.80	0.98	40	0.518
IRT v-a (lg10)	-0.63	0.37	0.37	0.73	-1.70	0.50	0.96	40	0.155
IRT n-u (lg10)	0.18	0.37	0.07	0.73	0.47	0.10	0.99	40	0.993
IRT n-a (lg10)	-0.83	0.38	2.32	0.74	-2.19	3.13	0.93	39	0.020
IRT (lg10) OLc	-0.47	0.37	-0.06	0.73	-1.24	-0.08	0.97	40	0.433
IRT v-u (lg10) OLc	-0.20	0.37	-0.58	0.73	-0.54	-0.79	0.98	40	0.538
IRT v-a (lg10) OLc	-0.45	0.37	-0.14	0.73	-1.19	-0.19	0.96	40	0.171
IRT n-u (lg10) OLc	0.18	0.37	0.06	0.73	0.48	0.08	0.99	40	0.992
IRT n-a (lg10) OLc	0.30	0.38	-0.20	0.74	0.79	-0.27	0.97	39	0.302
IRT (lg10) OLc (z)	-0.47	0.37	-0.06	0.73	-1.24	-0.08	0.97	40	0.433
IRT v-u (lg10) OLc (z)	-0.20	0.37	-0.58	0.73	-0.54	-0.79	0.98	40	0.538
IRT v-a (lg10) OLc (z)	-0.45	0.37	-0.14	0.73	-1.19	-0.19	0.96	40	0.171
IRT n-u (lg10) OLc (z)	0.18	0.37	0.06	0.73	0.48	0.08	0.99	40	0.992
IRT n-a (lg10) OLc (z)	0.30	0.38	-0.20	0.74	0.79	-0.27	0.97	39	0.302

Note. MSJ = mean mental state justifications, uIST = mean unique internal state terms, IRT = mean initial response time, ISTr = ratio of sum of overall response time to sum of IST, v-u = verbal unambiguous, v-a = verbal ambiguous, n-u = nonverbal unambiguous, n-a = nonverbal ambiguous, OLc = outlier-corrected, r = raw score, S = Wechsler scaled score (*M*=10, *SD*=6), IQ = IQ-normed score (*M*=100, *SD*=15), T = T-score (*M*=50, *SD*=10)

Table C 4

Distribution statistics of variables used in the analyses for adults

Variable	Skew	*SE-Skew*	Kurt	*SE-Kurt*	*Z-Skew*	*Z-Kurt*	Shapiro-Wilk *W*	*df*	*p*
Vocabulary raw	-1.11	0.37	2.80	0.73	-0.93	3.81	0.95	40	0.067
Vocabulary raw OLc	-0.75	0.37	0.02	0.73	0.02	0.02	0.98	40	0.753
Vocabulary raw OLc (z)	-0.75	0.37	0.02	0.73	0.02	0.02	0.98	40	0.753
Vocabulary (S)	-0.28	0.37	4.05	0.73	1.82	5.52	0.90	40	0.002
Vocabulary (S) OLc	-0.28	0.37	0.74	0.73	0.64	1.01	0.96	40	0.201
Digit Span raw	0.67	0.37	0.21	0.73	-0.59	0.29	0.97	40	0.444
Digit Span raw OLc	0.67	0.37	0.21	0.73	-0.59	0.29	0.97	40	0.444
Digit Span raw OLc (z)	0.67	0.38	0.33	0.74	-0.74	0.44	0.97	39	0.350
Digit Span (S)	0.57	0.37	0.42	0.73	-0.77	0.58	0.97	40	0.440
Digit Span (S) OLc	0.57	0.37	0.42	0.73	-0.77	0.58	0.97	40	0.440
DCCS Score	-2.62	0.37	-0.15	0.73	1.76	-0.21	0.93	40	0.021
DCCS Score OLc	-0.96	0.37	-0.15	0.73	1.76	-0.21	0.93	40	0.021
DCCS Score OLc (z)	-0.96	0.37	-0.15	0.73	1.76	-0.21	0.93	40	0.021
DCCS (IQ)	-1.17	0.37	-0.40	0.73	1.32	-0.55	0.96	40	0.216
DCCS (IQ) OLc	-0.62	0.37	-0.40	0.73	1.32	-0.55	0.96	40	0.217
Flanker Score	-1.71	0.37	7.15	0.73	-5.74	9.75	0.82	40	0.000
Flanker Score OLc	-1.14	0.37	0.31	0.73	-2.31	0.43	0.92	40	0.008
Flanker Score OLc (z)	-1.14	0.37	0.31	0.73	-2.31	0.43	0.92	40	0.008
Flanker (IQ)	-1.15	0.37	1.98	0.73	-3.63	2.70	0.88	40	0.000
Flanker (IQ) OLc	-0.98	0.37	0.14	0.73	-2.56	0.19	0.90	40	0.002
ASEBA self raw OLc (z)	0.56	0.39	0.01	0.77	2.00	0.01	0.92	36	0.012
ASEBA self (T)	0.21	0.39	0.68	0.77	2.37	0.88	0.93	36	0.023
ASEBA self (T) OLc	0.21	0.39	0.40	0.77	2.11	0.52	0.94	36	0.035
ASEBA other raw OLc (z)	0.48	0.40	0.67	0.79	1.90	0.86	0.94	34	0.063
ASEBA other (T)	0.48	0.40	0.67	0.79	1.90	0.86	0.94	34	0.063
ASEBA other (T) OLc	0.34	0.40	-0.06	0.79	2.46	-0.08	0.86	34	0.000
MSJ	2.15	0.37	3.33	0.73	3.71	4.55	0.90	40	0.002
MSJ v-u	2.28	0.37	9.18	0.73	6.79	12.52	0.73	40	0.000
MSJ v-a	2.95	0.37	0.65	0.73	2.06	0.88	0.93	40	0.011
MSJ n-u	1.39	0.37	1.26	0.73	2.94	1.72	0.89	40	0.001
MSJ n-a	0.97	0.37	2.25	0.73	2.28	3.06	0.92	40	0.009
MSJ OLc	0.78	0.37	-0.35	0.73	0.70	-0.48	0.97	40	0.322
MSJ v-u OLc	0.76	0.37	0.72	0.73	2.99	0.99	0.85	40	0.000
MSJ v-a OLc	0.78	0.37	0.00	0.73	1.26	0.00	0.94	40	0.046
MSJ n-u OLc	0.78	0.37	1.26	0.73	2.94	1.72	0.89	40	0.001
MSJ n-a OLc	0.97	0.37	0.78	0.73	0.63	1.07	0.96	40	0.186
MSJ OLc (z)	0.78	0.37	-0.35	0.73	0.70	-0.48	0.97	40	0.322
MSJ v-u OLc (z)	0.76	0.37	0.72	0.73	2.99	0.99	0.85	40	0.000
MSJ v-a OLc (z)	0.78	0.37	0.00	0.73	1.26	0.00	0.94	40	0.046
MSJ n-u OLc (z)	0.78	0.37	1.26	0.73	2.94	1.72	0.89	40	0.001
MSJ n-a OLc (z)	0.97	0.37	0.78	0.73	0.63	1.07	0.96	40	0.186
uIST	1.12	0.37	1.79	0.73	3.31	2.44	0.91	40	0.004
uIST v-u	1.03	0.37	4.96	0.73	4.65	6.76	0.87	40	0.000

Variable	Skew	*SE*-Skew	Kurt	*SE*-Kurt	*Z*-Skew	*Z*-Kurt	Shapiro-Wilk *W*	*df*	*p*
uIST v-a	1.75	0.37	0.60	0.73	2.19	0.82	0.94	40	0.046
uIST n-u	1.34	0.37	-0.07	0.73	1.17	-0.10	0.97	40	0.374
uIST n-a	0.91	0.37	3.04	0.73	3.85	4.15	0.89	40	0.001
uIST OLc	0.96	0.37	-0.19	0.73	1.89	-0.26	0.93	40	0.017
uIST v-u OLc	0.73	0.37	-0.38	0.73	1.70	-0.52	0.94	40	0.041
uIST v-a OLc	0.90	0.37	0.44	0.73	2.05	0.59	0.95	40	0.050
uIST n-u OLc	1.03	0.37	-0.07	0.73	1.17	-0.10	0.97	40	0.374
uIST n-a OLc	0.91	0.37	-0.30	0.73	0.22	-0.41	0.98	40	0.761
uIST OLc (z)	0.96	0.37	-0.19	0.73	1.89	-0.26	0.93	40	0.017
uIST v-u OLc (z)	0.73	0.37	-0.38	0.73	1.70	-0.52	0.94	40	0.041
uIST v-a OLc (z)	0.90	0.37	0.44	0.73	2.05	0.59	0.95	40	0.050
uIST n-u OLc (z)	1.03	0.37	-0.07	0.73	1.17	-0.10	0.97	40	0.374
uIST n-a OLc (z)	0.91	0.37	-0.30	0.73	0.22	-0.41	0.98	40	0.761
ISTr	0.85	0.37	3.98	0.73	4.80	5.43	0.84	40	0.000
ISTr v-u	2.42	0.37	2.57	0.73	3.39	3.50	0.90	40	0.001
ISTr v-a	0.66	0.37	2.03	0.73	3.62	2.76	0.88	40	0.001
ISTr n-u	1.62	0.37	7.20	0.73	6.64	9.83	0.73	40	0.000
ISTr n-a	0.33	0.37	22.45	0.73	11.80	30.62	0.49	40	0.000
ISTr OLc	0.85	0.37	-0.31	0.73	1.68	-0.43	0.93	40	0.015
ISTr v-u OLc	0.77	0.37	0.71	0.73	1.78	0.96	0.95	40	0.082
ISTr v-a OLc	0.00	0.37	-0.21	0.73	1.75	-0.28	0.91	40	0.005
ISTr n-u OLc	1.45	0.37	0.48	0.73	2.57	0.66	0.91	40	0.004
ISTr n-a OLc	0.33	0.37	-0.62	0.73	1.63	-0.85	0.92	40	0.008
ISTr OLc (z)	0.85	0.37	-0.31	0.73	1.68	-0.43	0.93	40	0.015
ISTr v-u OLc (z)	0.77	0.37	0.71	0.73	1.78	0.96	0.95	40	0.082
ISTr v-a OLc (z)	0.00	0.37	-0.21	0.73	1.75	-0.28	0.91	40	0.005
ISTr n-u OLc (z)	1.45	0.37	0.48	0.73	2.57	0.66	0.91	40	0.004
ISTr n-a OLc (z)	0.33	0.37	-0.62	0.73	1.63	-0.85	0.92	40	0.008
IRT	0.89	0.37	-0.42	0.73	1.43	-0.58	0.96	40	0.153
IRT v-u	1.67	0.37	1.06	0.73	2.83	1.44	0.92	40	0.008
IRT v-a	1.05	0.37	1.00	0.73	2.81	1.36	0.91	40	0.005
IRT n-u	1.45	0.37	6.07	0.73	5.83	8.29	0.79	40	0.000
IRT n-a	0.59	0.38	2.37	0.74	4.43	3.20	0.80	39	0.000
IRT (lg10)	0.00	0.37	-0.01	0.73	-1.26	-0.01	0.97	40	0.444
IRT v-u (lg10)	-0.55	0.37	-0.58	0.73	-0.55	-0.80	0.98	40	0.518
IRT v-a (lg10)	-0.58	0.37	0.37	0.73	-1.70	0.50	0.96	40	0.155
IRT n-u (lg10)	-0.07	0.37	0.07	0.73	0.47	0.10	0.99	40	0.993
IRT n-a (lg10)	-0.46	0.38	2.32	0.74	-2.19	3.13	0.93	39	0.020
IRT (lg10) OLc	0.01	0.37	-0.06	0.73	-1.24	-0.08	0.97	40	0.433
IRT v-u (lg10) OLc	-0.09	0.37	-0.58	0.73	-0.54	-0.79	0.98	40	0.538
IRT v-a (lg10) OLc	-0.42	0.37	-0.14	0.73	-1.19	-0.19	0.96	40	0.171
IRT n-u (lg10) OLc	-0.06	0.37	0.06	0.73	0.48	0.08	0.99	40	0.992
IRT n-a (lg10) OLc	-0.46	0.38	-0.20	0.74	0.79	-0.27	0.97	39	0.302
IRT (lg10) OLc (z)	0.01	0.37	-0.06	0.73	-1.24	-0.08	0.97	40	0.433

Variable	Skew	*SE*-Skew	Kurt	*SE*-Kurt	*Z*-Skew	*Z*-Kurt	Shapiro-Wilk *W*	*df*	*p*
IRT v-u (lg10) OLc (z)	-0.09	0.37	-0.58	0.73	-0.54	-0.79	0.98	40	0.538
IRT v-a (lg10) OLc (z)	-0.42	0.37	-0.14	0.73	-1.19	-0.19	0.96	40	0.171
IRT n-u (lg10) OLc (z)	-0.06	0.37	0.06	0.73	0.48	0.08	0.99	40	0.992
IRT n-a (lg10) OLc (z)	-0.46	0.38	-0.20	0.74	0.79	-0.27	0.97	39	0.302

Note. MSJ = mean mental state justifications, uIST = mean unique internal state terms, IRT = mean initial response time, ISTr = ratio of sum of overall response time to sum of IST, v-u = verbal unambiguous, v-a = verbal ambiguous, n-u = nonverbal unambiguous, n-a = nonverbal ambiguous, OLc = outlier-corrected, r = raw score, S = Wechsler scaled score (*M*=10, *SD*=6), IQ = IQ-normed score (*M*=100, *SD*=15), T = T-score (*M*=50, *SD*=10)

Table C 5
Distribution statistics of variables used in the analyses for older adults

Variable	Skew	*SE*-Skew	Kurt	*SE*-Kurt	*Z*-Skew	*Z*-Kurt	Shapiro-Wilk *W*	*df*	*p*
Vocabulary raw	0.58	0.38	-0.05	0.74	1.54	-0.07	0.96	39	0.230
Vocabulary raw OLc	0.58	0.38	-0.05	0.74	1.54	-0.07	0.96	39	0.230
Vocabulary raw OLc (z)	0.58	0.38	-0.05	0.74	1.54	-0.07	0.96	39	0.230
Vocabulary (S)	0.71	0.38	0.32	0.74	1.89	0.43	0.93	39	0.020
Vocabulary (S) OLc	0.71	0.38	0.32	0.74	1.89	0.43	0.93	39	0.020
Digit Span raw	0.85	0.38	1.53	0.74	2.24	2.06	0.93	39	0.018
Digit Span raw OLc	0.59	0.38	0.46	0.74	1.55	0.62	0.95	39	0.079
Digit Span raw OLc (z)	0.59	0.38	0.46	0.74	1.55	0.62	0.95	39	0.079
Digit Span (S)	0.11	0.38	1.82	0.74	0.28	2.45	0.93	39	0.022
Digit Span (S) OLc	-0.12	0.38	0.14	0.74	-0.31	0.18	0.95	39	0.114
DCCS Score	-0.22	0.38	-0.97	0.74	-0.59	-1.30	0.95	39	0.056
DCCS Score OLc	-0.22	0.38	-0.97	0.74	-0.59	-1.31	0.95	39	0.056
DCCS Score OLc (z)	-0.22	0.38	-0.97	0.74	-0.59	-1.31	0.95	39	0.056
DCCS (IQ)	0.59	0.38	-0.18	0.74	1.57	-0.25	0.95	39	0.105
DCCS (IQ) OLc	0.59	0.38	-0.19	0.74	1.56	-0.25	0.95	39	0.105
Flanker Score	0.25	0.38	-1.05	0.75	0.64	-1.40	0.95	38	0.064
Flanker Score OLc	0.25	0.38	-1.05	0.75	0.64	-1.40	0.95	38	0.064
Flanker Score OLc (z)	0.25	0.38	-1.05	0.75	0.64	-1.40	0.95	38	0.064
Flanker (IQ)	0.93	0.38	-0.30	0.75	2.43	-0.40	0.87	38	0.000
Flanker (IQ) OLc	0.93	0.38	-0.30	0.75	2.43	-0.40	0.87	38	0.000
ASEBA self raw OLc (z)	0.92	0.38	0.20	0.74	2.43	0.27	0.90	39	0.002
ASEBA self (T)	0.77	0.38	1.04	0.74	2.04	1.40	0.94	39	0.036
ASEBA self (T) OLc	0.19	0.38	-0.35	0.74	0.50	-0.47	0.96	39	0.146
ASEBA other raw OLc (z)	-0.11	0.43	-0.51	0.85	-0.26	-0.61	0.96	29	0.341
ASEBA other (T)	-0.11	0.43	-0.51	0.85	-0.26	-0.61	0.96	29	0.341
ASEBA other (T) OLc	0.65	0.43	-0.39	0.85	1.50	-0.46	0.94	29	0.089
MSJ	1.74	0.38	3.54	0.74	4.61	4.77	0.84	39	0.000
MSJ v-u	2.03	0.38	4.35	0.74	5.36	5.87	0.76	39	0.000
MSJ v-a	1.24	0.38	2.10	0.74	3.28	2.83	0.91	39	0.004

Variable	Skew	*SE*-Skew	Kurt	*SE*-Kurt	*Z*-Skew	*Z*-Kurt	Shapiro-Wilk		
							W	*df*	*p*
MSJ n-u	1.31	0.38	2.50	0.74	3.47	3.37	0.89	39	0.001
MSJ n-a	1.19	0.38	2.42	0.74	3.16	3.27	0.92	39	0.007
MSJ OLc	0.70	0.38	-0.19	0.74	1.84	-0.25	0.94	39	0.035
MSJ v-u OLc	0.87	0.38	1.02	0.74	2.29	1.38	0.92	39	0.010
MSJ v-a OLc	0.62	0.38	-0.10	0.74	1.65	-0.13	0.95	39	0.099
MSJ n-u OLc	0.36	0.38	0.10	0.74	0.96	0.13	0.96	39	0.247
MSJ n-a OLc	0.52	0.38	0.15	0.74	1.37	0.20	0.96	39	0.231
MSJ OLc (z)	0.70	0.38	-0.19	0.74	1.84	-0.25	0.94	39	0.035
MSJ v-u OLc (z)	0.87	0.38	1.02	0.74	2.29	1.38	0.92	39	0.010
MSJ v-a OLc (z)	0.62	0.38	-0.10	0.74	1.65	-0.13	0.95	39	0.099
MSJ n-u OLc (z)	0.36	0.38	0.10	0.74	0.96	0.13	0.96	39	0.246
MSJ n-a OLc (z)	0.52	0.38	0.15	0.74	1.37	0.20	0.96	39	0.231
uIST	1.98	0.38	5.10	0.74	5.24	6.89	0.82	39	0.000
uIST v-u	1.62	0.38	3.46	0.74	4.28	4.66	0.86	39	0.000
uIST v-a	2.15	0.38	7.02	0.74	5.68	9.48	0.82	39	0.000
uIST n-u	1.80	0.38	3.87	0.74	4.75	5.22	0.84	39	0.000
uIST n-a	1.94	0.38	4.68	0.74	5.13	6.31	0.82	39	0.000
uIST OLc	0.80	0.38	-0.21	0.74	2.12	-0.28	0.92	39	0.007
uIST v-u OLc	0.84	0.38	-0.32	0.74	2.22	-0.43	0.90	39	0.002
uIST v-a OLc	0.54	0.38	-0.07	0.74	1.42	-0.09	0.95	39	0.101
uIST n-u OLc	0.93	0.38	0.45	0.74	2.46	0.60	0.93	39	0.013
uIST n-a OLc	0.58	0.38	-0.49	0.74	1.52	-0.66	0.95	39	0.056
uIST OLc (z)	0.80	0.38	-0.21	0.74	2.12	-0.28	0.92	39	0.007
uIST v-u OLc (z)	0.84	0.38	-0.32	0.74	2.22	-0.43	0.90	39	0.002
uIST v-a OLc (z)	0.54	0.38	-0.07	0.74	1.42	-0.09	0.95	39	0.101
uIST n-u OLc (z)	0.93	0.38	0.45	0.74	2.46	0.60	0.93	39	0.013
uIST n-a OLc (z)	0.58	0.38	-0.49	0.74	1.52	-0.66	0.95	39	0.056
ISTr	3.20	0.38	13.61	0.74	8.47	18.36	0.69	39	0.000
ISTr v-u	1.14	0.38	1.62	0.74	3.01	2.18	0.92	39	0.006
ISTr v-a	2.06	0.38	5.44	0.74	5.45	7.33	0.81	39	0.000
ISTr n-u	3.77	0.38	16.40	0.75	9.84	21.86	0.57	38	0.000
ISTr n-a	2.54	0.38	7.51	0.75	6.64	10.01	0.72	38	0.000
ISTr OLc	0.32	0.38	-0.35	0.74	0.86	-0.47	0.97	39	0.309
ISTr v-u OLc	0.54	0.38	-0.12	0.74	1.44	-0.16	0.94	39	0.031
ISTr v-a OLc	0.62	0.38	-0.40	0.74	1.63	-0.54	0.94	39	0.044
ISTr n-u OLc	0.61	0.38	-0.34	0.75	1.60	-0.45	0.93	38	0.023
ISTr n-a OLc	0.46	0.38	-0.99	0.75	1.19	-1.33	0.93	38	0.015
ISTr OLc (z)	0.32	0.38	-0.35	0.74	0.86	-0.47	0.97	39	0.309
ISTr v-u OLc (z)	0.54	0.38	-0.12	0.74	1.44	-0.16	0.94	39	0.031
ISTr v-a OLc (z)	0.62	0.38	-0.40	0.74	1.63	-0.54	0.94	39	0.044
ISTr n-u OLc (z)	0.61	0.38	-0.34	0.75	1.60	-0.45	0.93	38	0.023
ISTr n-a OLc (z)	0.46	0.38	-0.99	0.75	1.19	-1.33	0.93	38	0.015
IRT	2.56	0.38	9.37	0.74	6.77	12.65	0.77	39	0.000
IRT v-u	2.00	0.38	4.13	0.74	5.29	5.57	0.77	39	0.000

Variable	Skew	*SE*-Skew	Kurt	*SE*-Kurt	*Z*-Skew	*Z*-Kurt	Shapiro-Wilk *W*	*df*	*p*
IRT v-a	2.55	0.38	7.53	0.74	6.74	10.16	0.71	39	0.000
IRT n-u	2.17	0.38	4.71	0.75	5.66	6.28	0.74	38	0.000
IRT n-a	2.92	0.38	11.46	0.75	7.62	15.28	0.72	38	0.000
IRT (lg10)	-0.17	0.38	0.38	0.74	-0.46	0.52	0.99	39	0.960
IRT v-u (lg10)	0.14	0.38	-0.46	0.74	0.37	-0.62	0.98	39	0.802
IRT v-a (lg10)	0.22	0.38	-0.54	0.74	0.57	-0.73	0.98	39	0.526
IRT n-u (lg10)	-1.07	0.38	3.64	0.75	-2.79	4.85	0.93	38	0.020
IRT n-a (lg10)	-0.24	0.38	1.13	0.75	-0.63	1.50	0.98	38	0.763
IRT (lg10) OLc	-0.17	0.38	0.37	0.74	-0.45	0.50	0.99	39	0.962
IRT v-u (lg10) OLc	0.15	0.38	-0.45	0.74	0.39	-0.61	0.98	39	0.790
IRT v-a (lg10) OLc	0.22	0.38	-0.54	0.74	0.58	-0.73	0.98	39	0.534
IRT n-u (lg10) OLc	0.32	0.38	-0.50	0.75	0.83	-0.67	0.97	38	0.418
IRT n-a (lg10) OLc	-0.24	0.38	1.13	0.75	-0.63	1.50	0.98	38	0.763
IRT (lg10) OLc (z)	-0.17	0.38	0.37	0.74	-0.45	0.50	0.99	39	0.962
IRT v-u (lg10) OLc (z)	0.15	0.38	-0.45	0.74	0.39	-0.61	0.98	39	0.790
IRT v-a (lg10) OLc (z)	0.22	0.38	-0.54	0.74	0.58	-0.73	0.98	39	0.534
IRT n-u (lg10) OLc (z)	0.32	0.38	-0.50	0.75	0.83	-0.67	0.97	38	0.418
IRT n-a (lg10) OLc (z)	-0.24	0.38	1.13	0.75	-0.63	1.50	0.98	38	0.763

Note. MSJ = mean mental state justifications, uIST = mean unique internal state terms, IRT = mean initial response time, ISTr = ratio of sum of overall response time to sum of IST, v-u = verbal unambiguous, v-a = verbal ambiguous, n-u = nonverbal unambiguous, n-a = nonverbal ambiguous, OLc = outlier-corrected, r = raw score, S = Wechsler scaled score (*M*=10, *SD*=6), IQ = IQ-normed score (*M*=100, *SD*=15), T = T-score (*M*=50, *SD*=10)

Table C 6
T-test for gender effects in adolescents and older adults

	adolescents				older adults			
	t	*df*	*p*	*r*	*t*	*df*	*p*	*r*
Vocabulary (r)	0.36	14.80	.72	.09	-0.63	16.66	.54	.15
Vocabulary (S)	0.22	22.48	.83	.05	-0.09	16.04	.93	.02
Digit Span (r)	0.96	15.94	.35	.23	0.12	15.29	.91	.03
Digit Span (S)	0.66	17.20	.52	.16	0.46	15.12	.65	.12
DCCS	-0.96	12.74	.35	.26	0.05	13.33	.96	.01
DCCS (IQ)	-1.29	13.60	.22	.33	-0.05	13.82	.96	.01
Flanker	-3.73	32.73	**.00**	.55	0.68	15.01	.51	.17
Flanker (IQ)	-3.44	31.36	**.00**	.52	0.77	16.67	.46	.18
ASEBA self (r)	-1.07	5.82	.33	.40	-0.75	13.52	.47	.20
ASEBA self (T)	-1.02	5.75	.35	.39	-1.18	12.78	.26	.31
ASEBA other (r)	-2.98	9.01	**.02**	.70	0.37	8.50	.72	.12
ASEBA other (T)	-3.65	18.09	.00	.65	0.45	8.30	.67	.15
MSJ	-0.99	14.06	.34	.25	1.66	18.10	.11	.36
MSJ v-ua	-1.29	10.34	.23	.37	0.88	14.82	.39	.22
MSJ v-a	-0.64	13.83	.53	.17	1.31	17.23	.21	.30
MSJ nv-ua	-0.58	11.66	.58	.17	2.00	23.70	.06	.38
MSJ nv-a	-0.15	26.27	.88	.03	1.85	18.05	.08	.40
IST	-1.14	15.59	.27	.28	0.81	13.40	.43	.22
uIST v-ua	-1.53	15.68	.15	.36	0.11	12.26	.91	.03
uIST v-a	-1.14	13.39	.27	.30	1.53	16.76	.15	.35
uIST nv-ua	-0.42	19.15	.68	.10	0.67	12.62	.52	.18
uIST nv-a	-1.10	16.34	.29	.26	1.20	14.71	.25	.30
IRT-log	0.86	10.90	.41	.25	-1.60	15.90	.13	.37
IRT-log v-ua	0.75	10.89	.47	.22	-0.18	14.93	.86	.05
IRT-log v-a	0.28	12.08	.79	.08	-0.94	17.27	.36	.22
IRT-log nv-ua	1.21	11.34	.25	.34	-2.68	14.34	**.02**	.58
IRT-log nv-a	0.72	11.70	.49	.21	-1.71	22.72	.10	.34
IST-ratio	-0.10	14.37	.92	.03	-2.61	16.64	**.02**	.54
IST-ratio v-ua	0.27	12.09	.79	.08	-0.60	16.30	.56	.15
IST-ratio v-a	0.31	16.77	.76	.07	-1.45	12.78	.17	.38
IST-ratio nv-ua	-0.77	12.76	.45	.21	-2.91	13.60	**.01**	.62
IST-ratio nv-a	-0.15	13.37	.88	.04	-2.06	11.44	.06	.52

Note. MSJ = mean mental state justifications, uIST = mean unique internal state terms, IRT = mean initial response time (sec.), IST-ratio = ratio of sum of overall response time (sec.) to sum of IST, v-u = verbal unambiguous, v-a = verbal ambiguous, nv-u = non-verbal unambiguous, nv-a = non-verbal ambiguous

r = raw score, S = Wechsler scaled score (M=10, SD=6), IQ = IQ-normed score (M=100, SD=15), T = T-score (M=50, SD=10)

Table C 7

Rank-correlations (Spearman) between FASC-variables and constructs of interest with gender-weighted scores for adolescents

Measures	n	MSJ	IST	IRT	ISTr
Vocabulary (r)	157	.61**	.60**	.29**	−.25**
Vocabulary (S)	157	.35**	.32**	.29**	−.00
Digit Span (r)	157	.41**	.41**	.10	−.25**
Digit Span (S)	157	.08	.06	.09	.03
DCCS	157	.39**	.44**	−.05	−.44**
DCCS (IQ)	157	−.05	−.03	−.30**	−.34**
Flanker	156	.31**	.35**	−.04	−.33**
Flanker (IQ)	156	−.08	−.03	−.20*	−.15
ASEBA self (T)	149	.29**	.32**	.17*	−.01
ASEBA other (T)	133	.09	.06	.09	−.08

Note. r = raw score, S = Wechsler scaled score (*M*=10, *SD*=6), IQ = IQ-normed score (*M*=100, *SD*=15), T = T-score (*M*=50, *SD*=10)

Table C 8

Correlations between FASC-variables and constructs of interest, z-standardized within each age group, pooled across age groups with gender-weighted scores for adolescents

Measures	n	MSJ (z)	uIST (z)	IRT (z)	ISTr (z)
Vocabulary (z)	157	.28**	.22**	.13	.01
Digit Span (z)	157	.12	.13	−.05	−.05
DCCS (z)	157	.16*	.23**	−.16*	−.36**
Flanker (z)	156	.16*	.15	−.05	−.05
ASEBA self (z)	149	.23**	.25**	−.02	−.02
ASEBA other (z)	133	−.03	−.07	−.14	−.07

Note. MSJ = mean mental state justifications, uIST = mean unique internal state terms, IRT = mean initial response time, ISTr = ratio of sum of overall response time to sum of IST

Table C 9

Correlations between FASC-variables and constructs of interest, z-standardized and gender-weighted scores for adolescents

Measures	n	MSJ (z)	uIST (z)	IRTm (z)	ISTr (z)
Vocabulary (z)	40	.23	.04	.15	.25
DCCS (z)	40	−.06	.31	−.17	−.62**
Flanker (z)	40	.10	.12	.07	−.29
ASEBA self (z)	33[a]	.30	.25	.25	−.10

Note. MSJ = mean mental state justifications, IST = mean unique internal state terms, IRT = mean initial response time, ISTr = ratio of sum of overall response time to sum of IST

[a] because of more missing data in male subjects, these correlations are based on only 21 female subjects weighted 1 and 6 male subjects weighted 2

Appendix D: cTRUE item examples and overview of emotions covered in TEC and cTRUE items

Table D 1

Overview of the number of items per component in the TEC and CTEU and covered emotions

	TEC		CTEU	
component	**#**	**emotive states covered**	**#**	**emotive states covered**
Recognition	5	happy, neutral, sad, scared, angry	15	happy (3), neutral (3), sad (3), angry (3)
External Causes	5	happy, neutral, sad, scared, angry	15	happy (3), neutral (3), sad (3), scared (3), angry (3)
Desires	1	happy, sad	2	happy (2), sad (2)
Beliefs	1	neutral	5	neutral, sad (2), scared, angry
Reminder	1	sad	3	sad, scared, angry
Regulation	1	cognitive emotion regulation (5)	5	cognitive emotion regulation (5)
Hiding	1	angry	5	neutral, sad, scared, angry (2)
Mixed emotions	1	happy/scared	5	happy/scared (3), happy/sad, sad/angry
Morality	1	sad	3	sad (3)

Note: All cTRUE items were created by Hans Haslinger with the software bitstrips. Figures are not printed in the original size. In the cTRUE procedure, pictures are presented consecutively. Response pictures are only shown for components which do not use the emotion faces used in the component of emotion recognition. The test procedure printed here is translated from german.

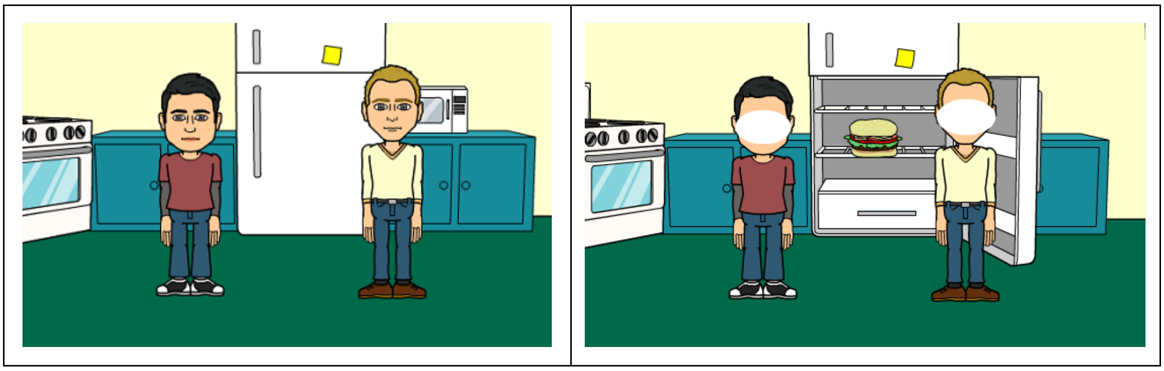

Figure D 1. CTRUE example item for the desires component

This is Tom and Peter. They are both very hungry and want something to eat. Tom likes burgers and would like to eat burgers every day. Peter does not like burgers at all. He would like to eat something else.

Control question 1: Does Tom like burgers? Yes or no?

Right! / Wrong! Tom likes burgers. He would like to eat burgers every day.

Control question 2: Does Peter like burgers? Yes or no?

Right! / Wrong! Peter does not like burgers. He would like to eat something else.

Let’s see what’s inside the fridge. Please click with the mouse on the fride to open it.

There's a burger in the fridge!

Test question 1: How does Tom feel now? Does he feel sad, angry, alright or happy?

Test question 2: How does Peter feel now? Does he feel angry, alright, sad or happy?

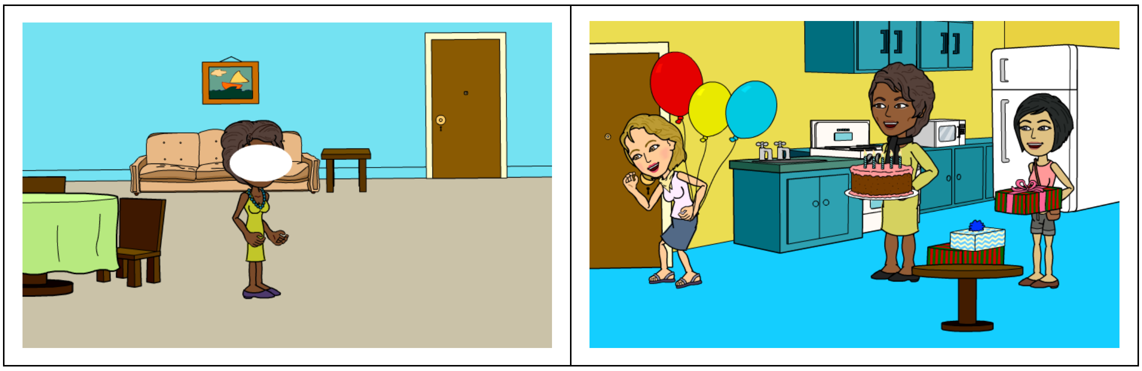

Figure D 2. CTRUE example item for the belief-based emotions component

This is Marie. Today is her birthday but when she comes home nobody is there to congratulate her. Please click on the door to see what's behind.

In the other room Marie's mother and her friends are hiding to surprise her. Please click on the door again to return to Marie.

Control question: Does Marie know that there is a surprise party for her?

Right! Wrong! Marie does (not) know, that there is a surprise party for her. She thinks everyone forgot about her birthday.

Test question: How does Marie feel right now? Does she feel alright, happy, sad or scared?

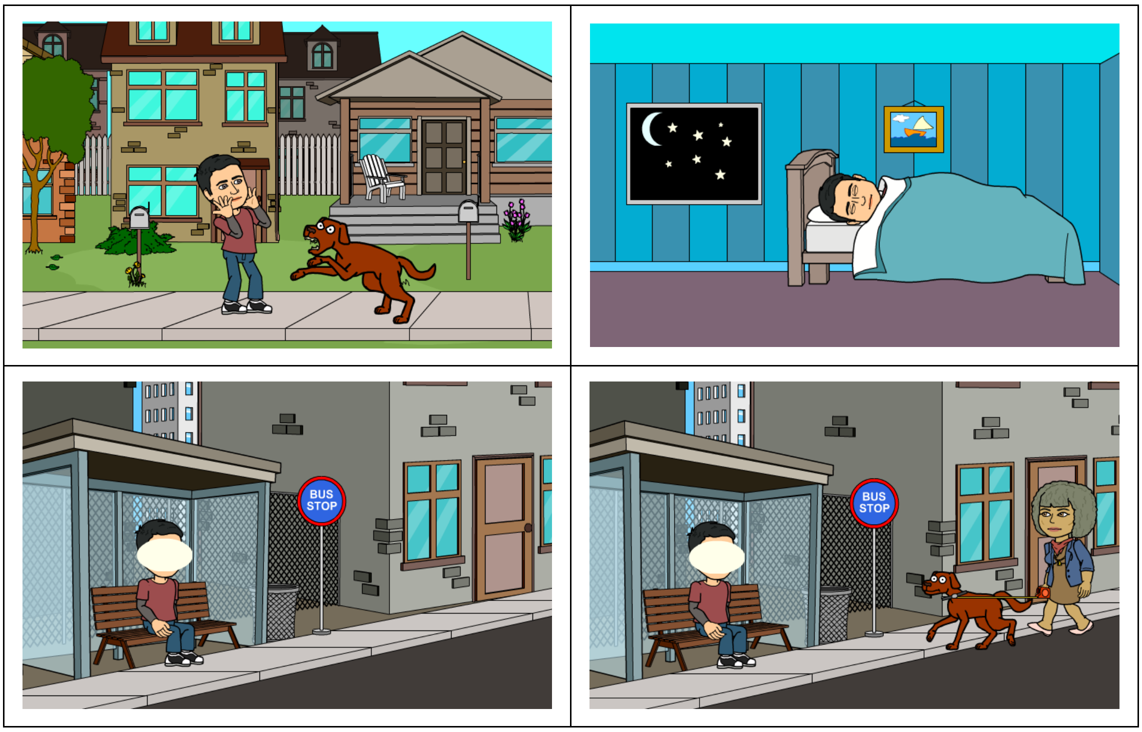

Figure D 3. CTRUE example item for the reminder component

This is Tom. A large brown dog jumps up at Tom and he is very scared.

Later that evening he goes to bed and sleeps

The next day Tom sits at the bus station and waits for the bus.

Control question: How does Tom feel while he is sitting at the station and waiting for the bus? Does he feel scared, angry, happy or alright?

Right! / Wrong! Tom feels alright while he is waiting for the bus.

A bit later, a woman with the large brown dog of yesterday comes along.

Test question: How does Tom feel when the woman with the dog comes along? Does he feel scared, sad, alright or happy.

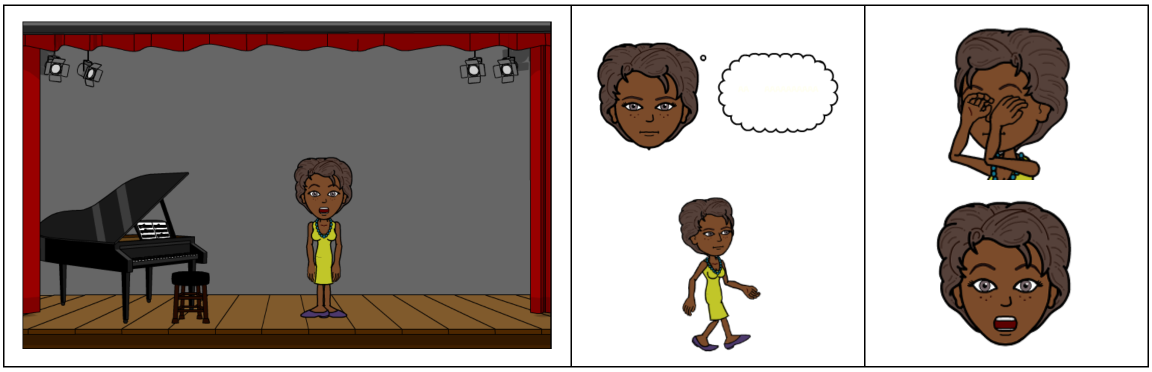

Figure D 4. CTRUE example item and response options for the emotion regulation component.

Marie is performing in front of a large audient. She is afrad that she might play the wrong note and people might laugh at her.

Test question: What can Marie do to stop herself from being afraid?

Can she cover her eyes to stop being afraid?

Can she walk out of the hall to stop being afraid?

Can she think that she has practiced a lot to stop being afraid?

Can Marie do nothing to stop being afraid?

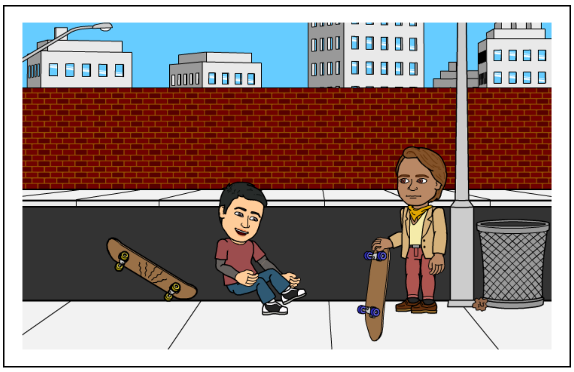

Figure D 5. CTRUE example item for the hiding emotions component (appearance-reality).

This is Tom and Alex. Tom just fell down with his skateboard. He laughs to not show his real feelings.

Test question: How does Tom really feel? Does he feel sad, angry, scared or happy?

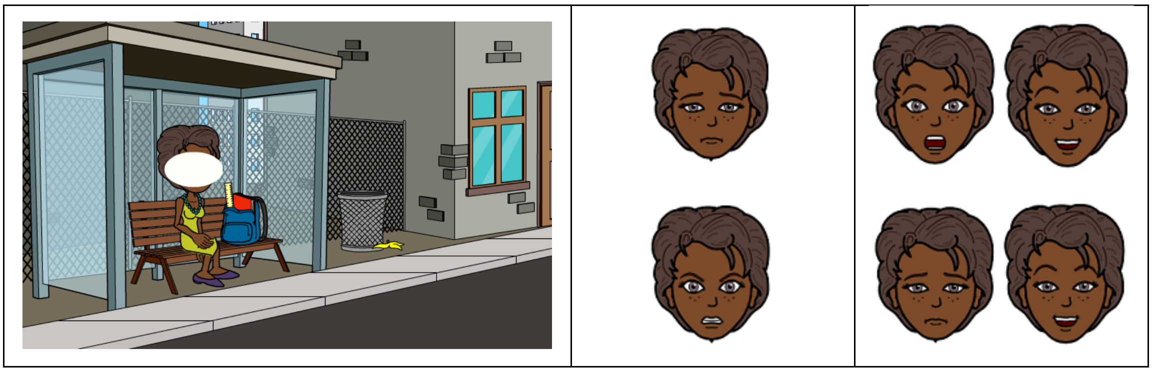

Figure D 6. CTRUE example item and response options for the mixed-emotions component.

Marie is sitting at the bust station and waiting for the schoolbus. Today is the first day after school holidays and Marie thinks that she has to get up early each day again. At the same time she thinks that she will meet all her friends at school again.

Test question: How does Marie feel? Does she feel sad, scared and happy, angry or sad and happy?

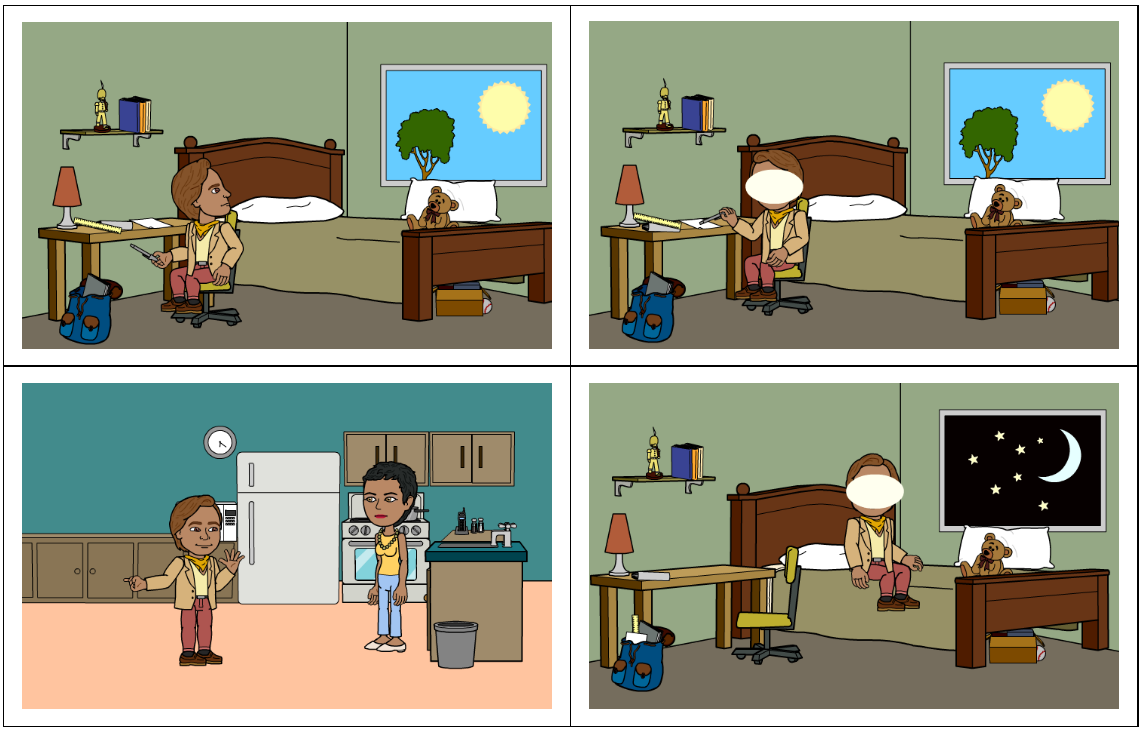

Figure D 7. CTRUE example item for the morality component.

Alex is sitting in his room and doing homework. Outside the weather is very nice and he would rather go out and play than doing his homework. He thinks about just telling his mother that he’s done his homework already to be allowed to go outside.

Control question: Is it ok for Alex to lie to his mother to be allowed to go outside?

Right! / Wrong! It is wrong for Alex to lie to his mother only to be allowed to go outside. You shouldn’t lie.

Alex decides to not lie to his mother and to finish his homework first.

Question: How does Alex feel because he did not lie to his mother? Does he feel sad,, angry, alright or happy?

Shortly after, Alex can’t stop himself and tells his mother that he has finished his homework although that’s not true. In the evening before going to bed, Alex thinks about all the things he has done today. He remembers lying to his mother only to be allowed to go outside.

Test question: How does Alex feel? Does he feel happy, alright, sad or scared?

Appendix E: Additional results for study 2

Table E 1

Distribution statistics of variables used in the analyses of study 2

Variable	Skew	*SE*-Skew	Kurt	*SE*-Kurt	*Z*-Skew	*Z*-Kurt	Shapiro-Wilk *W*	*df*	*p*
Age (years)	0.60	0.32	-0.61	0.63	1.87	-0.96	0.94	55	0.01
Vocabulary (standard score)	0.10	0.32	-0.69	0.63	0.30	-1.08	0.98	55	0.33
cTRUE	-0.46	0.32	-0.14	0.63	-1.43	-0.22	0.97	55	0.13
cTRUE: recognition	-2.60	0.32	8.55	0.63	-8.07	13.49	0.66	55	0.00
cTRUE: ext. causes	-1.14	0.32	2.54	0.63	-3.52	4.00	0.89	55	0.00
cTRUE: reminder	-1.02	0.32	-0.07	0.63	-3.18	-0.11	0.70	55	0.00
cTRUE: desires	-0.97	0.32	-0.47	0.63	-3.02	-0.74	0.70	55	0.00
cTRUE: beliefs	-0.62	0.32	-1.12	0.63	-1.92	-1.76	0.75	55	0.00
cTRUE: hiding	-1.04	0.32	0.78	0.63	-3.24	1.23	0.86	55	0.00
cTRUE: regulation	-1.51	0.32	1.45	0.63	-4.69	2.29	0.73	55	0.00
cTRUE: mixed	-0.62	0.32	-0.67	0.63	-1.92	-1.05	0.88	55	0.00
cTRUE: morality	-0.80	0.32	-0.47	0.63	-2.48	-0.74	0.77	55	0.00
cTRUE: RT (ms)	0.55	0.32	-0.36	0.63	1.71	-0.57	0.96	55	0.06
cTRUE: recognition RT	1.05	0.32	0.58	0.63	3.27	0.92	0.91	55	0.00
cTRUE: external causes RT	0.77	0.32	-0.04	0.63	2.40	-0.06	0.94	55	0.01
cTRUE: reminder RT	0.51	0.32	-0.32	0.63	1.58	-0.50	0.97	55	0.11
cTRUE: desires RT	1.19	0.34	1.44	0.67	3.51	2.15	0.89	49	0.00
cTRUE: beliefs RT	1.10	0.32	1.04	0.63	3.41	1.63	0.92	55	0.00
cTRUE: hiding RT	0.67	0.33	-0.84	0.65	2.02	-1.29	0.90	52	0.00
cTRUE: regulation RT	0.91	0.33	0.00	0.64	2.78	0.00	0.90	54	0.00
cTRUE: mixed RT	1.19	0.34	1.60	0.66	3.52	2.41	0.90	50	0.00
cTRUE: morality RT	0.86	0.33	0.42	0.64	2.65	0.66	0.94	54	0.01
TEC	-0.26	0.32	-0.58	0.63	-0.81	-0.91	0.93	55	0.00
EK: total	-0.53	0.32	0.09	0.63	-1.66	0.15	0.96	55	0.09
Multiple Emotions Task	1.89	0.32	2.63	0.63	5.86	4.15	0.59	55	0.00
SSIS: Social Skills	-0.23	0.37	-0.34	0.72	-0.63	-0.46	0.98	41	0.72
SSIS: Problem Behaviors	0.30	0.37	-0.68	0.72	0.83	-0.94	0.97	42	0.40
SSIS: Academic Competence	-0.48	0.37	-0.23	0.72	-1.30	-0.31	0.97	42	0.22
YSR: Social Competence	0.11	0.32	-0.34	0.63	0.34	-0.54	0.98	55	0.59
PRQ: Prosocial	-0.45	0.37	-0.44	0.72	-1.24	-0.61	0.97	42	0.25
PRQ: Defender	-0.24	0.37	-0.26	0.72	-0.66	-0.36	0.96	42	0.20
PRQ: Consoler	-0.43	0.37	-0.22	0.72	-1.18	-0.31	0.94	42	0.04
PRQ: Mediator	-0.34	0.37	-0.60	0.72	-0.92	-0.83	0.96	42	0.10
PRQ: Probully	0.72	0.37	-0.04	0.72	1.98	-0.06	0.94	42	0.03
PRQ: Bully	1.02	0.37	0.09	0.72	2.78	0.12	0.81	42	0.00
PRQ: Reinforcer	-0.11	0.37	-0.62	0.72	-0.30	-0.87	0.96	42	0.10
PRQ: Assistant	1.61	0.37	3.55	0.72	4.41	4.95	0.82	42	0.00
PRQ: Outsider	1.48	0.37	1.91	0.72	4.05	2.67	0.78	42	0.00
PRQ: Victim	0.08	0.37	0.41	0.72	0.22	0.58	0.96	42	0.20
FASC: MSJ	0.44	0.35	1.01	0.69	1.25	1.47	0.89	46	0.00
FASC: MJTRr	-0.46	0.35	-0.61	0.69	-1.30	-0.89	0.89	46	0.00

Variable	Skew	*SE*-Skew	Kurt	*SE*-Kurt	*Z*-Skew	*Z*-Kurt	Shapiro-Wilk *W*	*df*	*p*
FASC: IST	1.19	0.35	0.57	0.69	3.39	0.83	0.85	46	0.00
FASC: ISTr	0.54	0.35	-0.26	0.69	1.53	-0.38	0.94	46	0.03
FASC: IRT (s)	1.05	0.35	1.10	0.69	2.99	1.60	0.93	46	0.01

Note. cTRUE = Computerized Task of Recognizing and Understanding Emotions, EK = emotion knowledge, FASC = Flexibility and Automaticity of Social Cognition, IRT = initial response time, IST = sum of internal state terms, ISTr = ratio of IST to total number of words, MET = Multiple Emotions Task, MSJ = sum of mental justifications, MSJTRr = ratio of MSJ to sum of total responses, PRQ = Participant Roles Questionnaire, RT = response time (in ms), SSIS = Social Skills Improvement System, TEC = Test of Emotion Comprehension

Table E 2

Descriptive statistics for sub-scores of emotion understanding measures

Measure	*M*	*SD*	min	max
cTRUE: total	7.18	1.09	3.33	8.93
cTRUE: recognition	.95	.08	.60	1.00
cTRUE: ext. causes	.87	.10	.47	1.00
cTRUE: reminder	.84	.22	.33	1.00
cTRUE: desires	.73	.37	.00	1.00
cTRUE: beliefs	.86	.16	.60	1.00
cTRUE: hiding	.71	.27	.00	1.00
cTRUE: regulation	.82	.26	.00	1.00
cTRUE: mixed	.65	.32	.00	1.00
cTRUE: morality	.76	.28	.00	1.00
cTRUE: mean RT	1683	457	979	2826
cTRUE: recognition RT	3102	839	1971	5512
cTRUE: ext. causes RT	1506	628	575	3190
cTRUE: reminder RT	1500	692	286	3200
cTRUE: desires RT [1]	1343	784	333	3969
cTRUE: beliefs RT	1430	754	431	3695
cTRUE: hiding RT [2]	1621	948	345	3564
cTRUE: regulation RT [3]	1305	751	362	3323
cTRUE: mixed RT [4]	1734	1152	283	5500
cTRUE: morality RT [3]	1492	917	71	4237
TEC: total	6.91	1.39	4.00	9.00
TEC: recognition	.96	.19	.00	1.00
TEC: ext. causes	.95	.23	.00	1.00
TEC: reminder	.85	.36	.00	1.00
TEC: desires [1]	.82	.39	.00	1.00
TEC: beliefs	.58	.50	.00	1.00
TEC: hiding	.73	.45	.00	1.00
TEC: regulation	.84	.37	.00	1.00
TEC: mixed	.60	.49	.00	1.00
TEC: morality	.58	.50	.00	1.00
EK: total	29.78	4.64	16.00	37.00
EK: facial expressions	13.80	1.59	8.00	16.00
EK: social behaviours	6.93	2.26	1.00	11.00
EK: social situations	9.05	1.91	4.00	12.00
MET (any 2 emotions)	.51	.90	.00	3.00
MET (normative)	.45	.86	.00	3.00

Measure	*M*	*SD*	min	max

Note. all n=55, except [1] n=49, [2] n=52, [3] n=54, [4] n=50, cTRUE = Computerized Task of Recognizing and Understanding Emotions, RT = response time in ms, TEC = Test of Emotion Comprehension, EK = emotion knowledge, MET = multiple emotions task

Table E 3
cTRUE item characteristics

Item	Item difficulty (*p*)	Item-scale-correlation	Cronbach's Alpha excl. item
Recognition 1	1.00	-	-
Recognition 2	.98	-.05	.81
Recognition 3	1.00	-	-
Recognition 4	.96	-.11	.81
Recognition 5	1.00	-	-
Recognition 6	.96	.51	.80
Recognition 7	.98	.52	.81
Recognition 8	.75	.33	.80
Recognition 9	.96	-.03	.81
Recognition 10	.85	.39	.80
Recognition 11	.98	.18	.81
Recognition 12	.98	.04	.81
Recognition 13	1.00	-	-
Recognition 14	.95	.12	.81
Recognition 15	.84	.42	.80
External Causes 1	.96	.37	.81
External Causes 2	.98	.52	.81
External Causes 3	.42	.17	.81
External Causes 4	.89	.14	.81
External Causes 5	.91	.09	.81
External Causes 6	.89	.26	.81
External Causes 7	.93	.34	.81
External Causes 8	.91	.17	.81
External Causes 9	.95	.43	.80
External Causes 10	.98	-.05	.81
External Causes 11	.73	.42	.80
External Causes 12	.84	.18	.81
External Causes 13	.75	.22	.81
External Causes 14	.98	.52	.81
External Causes 15	.91	.10	.81
Reminder 1	.84	.06	.81
Reminder 2	.76	.32	.81
Reminder 3	.91	.08	.80
Desire-based-emotions 1	.69	.27	.81
Desire-based-emotions 2	.76	.28	.80
Belief-based-emotions 1	.65	.13	.81
Belief-based-emotions 2	.98	-.20	.80
Belief-based-emotions 3	.87	.38	.81
Belief-based-emotions 4	.87	.28	.81
Belief-based-emotions 5	.93	.46	.81

Item	Item difficulty (p)	Item-scale-correlation	Cronbach's Alpha excl. item
Hiding Emotions 1	.67	.50	.80
Hiding Emotions 2	.75	.42	.81
Hiding Emotions 3	.76	.13	.81
Hiding Emotions 4	.91	.59	.81
Hiding Emotions 5	.45	.25	.81
Emotion Regulation 1	.84	.28	.81
Emotion Regulation 2	.76	.17	.81
Emotion Regulation 3	.87	.10	.81
Emotion Regulation 4	.80	.33	.81
Emotion Regulation 5	.82	.46	.80
Mixed Emotions 1	.60	.52	.81
Mixed Emotions 2	.56	.09	.80
Mixed Emotions 3	.71	.39	.81
Mixed Emotions 4	.62	.34	.81
Mixed Emotions 5	.76	.35	.81
Morality 1	.75	-.08	.81
Morality 2	.78	.30	.81
Morality 3	.76	.09	.80

Note. N=55. Empty cells represent items with zero variance.

Table E 4

Correlations between TEC components and social skills

Measure	SSIS:SS	SSIS:PB	SSIS:AC	YSR
TEC[1]	.17	.10	.08	-.13
TEC: recognition	.24	-.07	-.21	.09
TEC: external causes	.13	.08	-.04	-.09
TEC: reminder	-.10	.10	.08	.07
TEC: desires	-.10	.02	-.07	.04
TEC: beliefs	.14	.07	.14	-.07
TEC: hiding	.13	.19	.08	.01
TEC: regulation	.06	.02	.10	-.12
TEC: mixed emotions[1]	.38*	-.01	-.08	-.11
TEC: morality	-.22	.07	.05	-.20

Note. N=42, except n=41 for SSIS-SS and n=55 for YSR. TEC = Test of Emotion Comprehension, SSIS = Social Skills Improvement System, SS = social skills, PB = problem behaviour, AC = academic competence, YSR = Youth Self Report

[1]rank correlation (Spearman)

$^{+}$ p<.10, * p<.05, ** p<.01

Table E 5
Rank correlations (Spearman) between desire-based emotion items and measures of social competence and intercorrelations between items

Measure	cTRUE desires	cTRUE football	cTRUE salad	TEC desires	TEC coke	TEC salad
SSIS-ss	-.26	-.29⁺	-.11	-.14	-.16	-.10
PRQ prosocial	-.26	-.12	-.30⁺	-.20	-.19	-.18
PRQ defender	-.01	.05	-.07	-.01	-.05	.04
PRQ consoler	-.35*	-.16	-.42**	-.25	-.19	-.28⁺
PRQ mediator	-.33*	-.23	-.30⁺	-.28⁺	-.27⁺	-.25
SSIS-pb	.21	.30	.01	.11	.18	.02
PRQ probully	.30⁺	.40**	.06	.17	.26	.05
PRQ bully	.24	.35*	.01	.15	.20	.07
PRQ assistant	.36*	.42**	.15	.19	.28*	.05
PRQ reinforcer	.20	.31*	.00	.13	.23	.01
cTRUE desires	1.00					
cTRUE football	.84**	1.00				
cTRUE salad	.81**	.37**	1.00			
TEC desires	.69**	.64**	.50**	1.00		
TEC coke	.69**	.69**	.43**	.95**	1.00	
TEC salad	.61**	.50**	.51**	.94**	.78**	1.00

Note. N=42 except for SSIS-ss. cTRUE = Computerized Task of Recognizing and Understanding Emotions, PRQ = Participants Roles Questionnaire, TEC = Test of Emotion Comprehension, SSIS = Social Skills Improvement System, ss = social skills, pb = problem behaviours
⁺ p<.10, * p<.05, ** p<.01

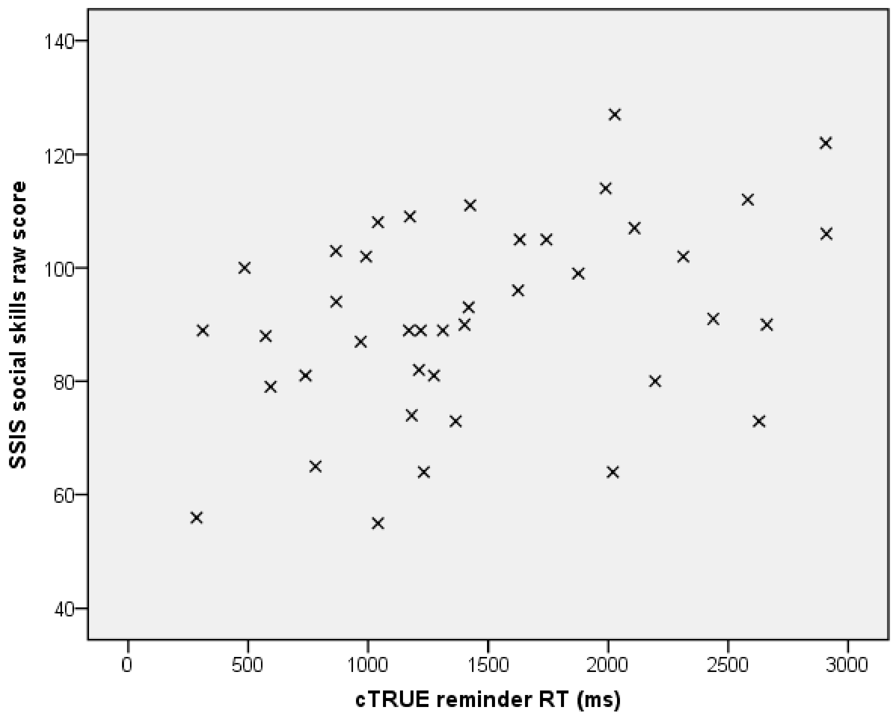

Figure E 1. Scatter plot of mean cTRUE reminder component response times and SSIS social skills score

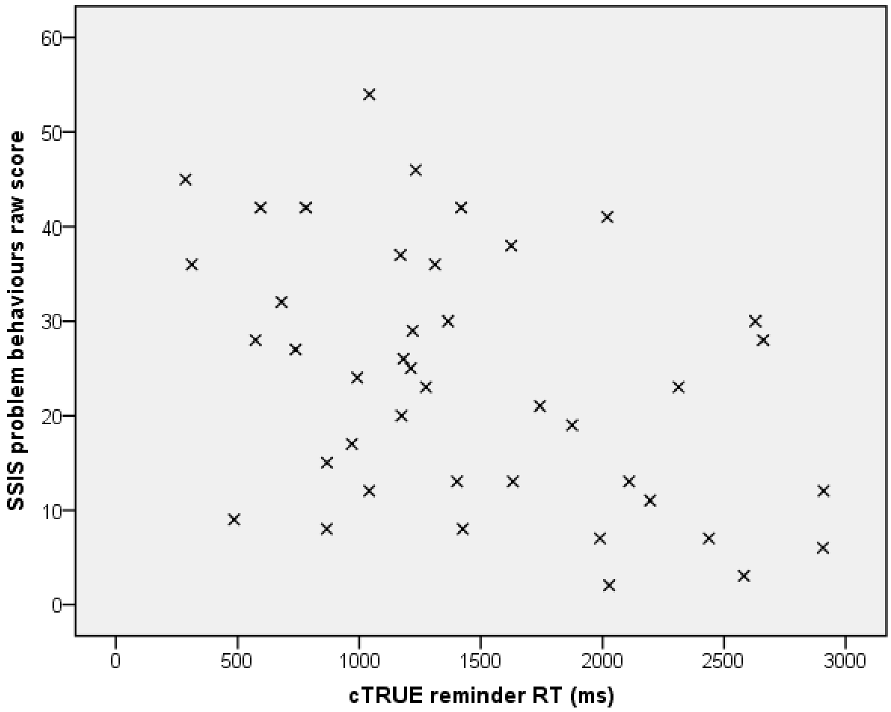

Figure E 2. Scatter plot of mean cTRUE reminder component response times and SSIS problem behaviours score

Appendix F: EmoJump Credits

EmoJump Credits

The development of EmoJump was a joint effort between the Clinical Child and Adolescent Psychology at the Faculty of Psychology (University of Vienna) under Univ.-Prof. Mag. Dr. Manuel Sprung and the Entertainment Computing Group of the Faculty of Computer Science (University of Vienna) under Univ.-Prof. Dipl.-Ing. Dr. Helmut Hlavacs. Manuel Sprung and Jakob Leyrer developed the idea. Several master-students of psychology and two master students of informatics were involved in the stimulus creation and development of the detailed gameplay. These were in alphabetic order: Sandra Anderl, Krisztina Halasz, Hans Haslinger, Katharina Meusburger, Laura Neumann, Judith Reiss, Theresa Resch, Tanja Rüscher, Natascha Schweiger, Christina Zauner, Vanessa Zechner. Voice acting was done by Tanja Rüscher and Vanessa Zechner. The two informatics master students also coded the program and were supervised by Univ.-Prof. Dipl.-Ing. Dr. Helmut Hlavacs. Jakob Leyrer contributed to the gameplay and creative design of the program and was head of the project in terms of decisions regarding gameplay and design.

Appendix G: Additional results for study 3

Table G 1

Distribution statistics of residuals of variables used in the analyses of study 3

							Shapiro-Wilk		
Residuals of variables	Skew	*SE*-Skew	Kurt	*SE*-Kurt	*Z*-Skew	*Z*-Kurt	*W*	*df*	*p*
cTRUE pre	-0.55	0.46	-0.32	0.89	-1.20	-0.36	0.95	26	0.27
cTRUE post	-0.67	0.46	0.33	0.89	-1.46	0.37	0.96	26	0.31
cTRUE external causes pre	-1.36	0.46	2.36	0.89	-2.99	2.66	0.86	26	0.00
cTRUE external causes post	-0.83	0.46	-0.13	0.89	-1.83	-0.15	0.89	26	0.01
cTRUE beliefs pre	-0.47	0.46	-0.67	0.89	-1.03	-0.76	0.78	26	0.00
cTRUE beliefs post	-0.91	0.46	0.40	0.89	-1.99	0.45	0.86	26	0.00
cTRUE mixed emotions pre	-1.25	0.46	0.44	0.89	-2.73	0.50	0.78	26	0.00
cTRUE mixed emotions post	-1.37	0.46	2.15	0.89	-2.99	2.42	0.84	26	0.00
FASC MSJ pre	0.46	0.46	0.71	0.89	1.01	0.79	0.95	26	0.21
FASC MSJ post	0.61	0.46	0.76	0.89	1.33	0.85	0.87	26	0.00
FASC unique IST pre	0.89	0.46	1.85	0.89	1.95	2.08	0.94	26	0.15
FASC unique IST post	0.66	0.46	0.54	0.89	1.45	0.61	0.96	26	0.43
FASC IST ratio pre	0.91	0.46	0.96	0.90	1.95	1.07	0.93	25	0.08
FASC IST ratio post	-0.02	0.46	-0.79	0.90	-0.05	-0.88	0.96	25	0.45
FASC IRT pre	1.97	0.47	2.90	0.92	4.18	3.15	0.69	24	0.00
FASC IRT post	1.16	0.47	2.89	0.92	2.47	3.15	0.91	24	0.04
FASC IRT reciproke pre	0.46	0.47	-0.34	0.92	0.97	-0.37	0.97	24	0.55
FASC IRT reciproke post	0.74	0.47	0.09	0.92	1.57	0.09	0.92	24	0.07
DCCS pre	-1.64	0.46	1.80	0.89	-3.59	2.02	0.77	26	0.00
DCCS post	-1.13	0.46	0.49	0.89	-2.47	0.56	0.86	26	0.00
Flanker pre	-1.25	0.46	1.26	0.89	-2.73	1.42	0.88	26	0.01
Flanker post	-0.54	0.46	0.12	0.89	-1.18	0.14	0.97	26	0.62
PANAS positive affect pre	-0.12	0.46	-0.78	0.89	-0.27	-0.88	0.97	26	0.53
PANAS positive affect post	-0.47	0.46	-0.14	0.89	-1.02	-0.16	0.94	26	0.14
PANAS negative affect pre	0.64	0.46	-0.24	0.89	1.40	-0.27	0.93	26	0.09
PANAS negative affect post	-0.27	0.46	-0.37	0.89	-0.59	-0.41	0.98	26	0.76

Note. cTRUE = Computerized Task of Recognizing and Understanding Emotions, DDCS = Dimensional Change Card Sort, FASC = Flexibility and Automaticity of Social Cognition, IRT = initial response time, IST = sum of internal state terms, IST ratio = ratio of IST to total number of words, MSJ = sum of mental justifications, PANAS = Positive and Negative Affective Schedule

Appendix H: Abstract (German)

Sozial-emotionale Kompetenzen sind wichtige Einflussfaktoren für emotionale Resilienz und psychische Gesundheit. Die Möglichkeit Personen mit Entwicklungsbedarf in diesem Bereich frühzeitig zu identifizieren und Trainingsprogramme praktisch und ökonomisch zu verbreiten, könnte helfen, die weltweite Belastung durch psychische Gesundheitsprobleme zu reduzieren. Vor diesem Hintergrund und angesichts des Umstandes, dass Computerspiele fixer Bestandteil der Lebenswelt von Kindern sind, können computer-basierte Interventionen als Erweiterung des sogenannten *task-shifting* Ansatzes der Weltgesundheitsorganisation gesehen werden. Studie eins untersucht sozial-kognitives Schlussfolgern und das Sprechen über mentale Zustände (Englisch: mental state talk) quer über die Lebensspanne mithilfe des neuen Verfahrens *Flexibility and Automaticity of Social Cognition* (FASC), das auf Cartoon-Vignetten basiert und auch in Studie zwei und drei zum Einsatz kommt. Die Flexibilität der mentalen Zustandserklärungen des Verhaltens der Charaktere in den Geschichten stieg von den Kindern über die Jugendlichen bis zu den Erwachsenen an und zeigte bei den älteren Erwachsenen wieder einen Abfall. Außerdem modulierte das Vorhandensein von verbalen Hinweisen, sowie das Ausmaß an Ambiguität, Faktoren, die bei anderen Verfahren der *advanced theory of mind* üblicherweise vernachlässigt werden, die Ergebnisvariablen. FASC ist ein vielversprechendes neues Verfahren um die Mentalisierungsfähigkeit über die Lebensspanne zu erforschen. Studie zwei präsentiert das Verfahren *computerized Task of Recognizing and Understanding Emotions* (cTRUE) das auf auf dem *Test of Emotion Comprehension* (TEC) basiert. Der cTRUE Gesamtscore zeigte gute interne Konsistenz und konvergente Validität mit etablierten Verfahren zur Messung von Emotionsverständnis. Für das Verständnis externaler Ursachen von Emotionen bzw. Emotionsregulation wurden Hinweise gefunden, dass Reaktionszeiten im cTRUE akademische Kompetenzen bzw. pro-soziales Rollenverhalten über die Genauigkeitsvariablen von TEC und cTRUE hinaus vorhersagen können. Es besteht jedoch noch weiterer Entwicklungsbedarf hinsichtlich einzelner Komponenten des Emotionsverständnisses, die bezüglich Reliabilität und Validität eine große Heterogenität aufweisen. Studie drei beschreibt die Entwicklung und Evaluation von EmoJump, einem Computerspiel zur Förderung des externalen, mentalen und reflexiven Emotionsverständnisses im Grundschulalter. In einer randomisierten kontrollierten Studie zeigte sich nach 12 Trainingseinheiten zu je 20-30 Minuten ein verbessertes Verständnis von gemischten Emotionen. Implikationen für die weitere Entwicklung von Spielen zur Förderung des Emotionsverständnisses werden diskutiert.