Measurement of affective empathy with Pictorial Empathy Test (PET)

Iivo Koirikivi

Master's Thesis

Psychology

Department of Behavioral Sciences

April 2014

Supervisor: Marjaana Lindeman

Contents

1. Introduction

As human beings evolved to their present form, they developed a need to understand and share emotions and experiences: ultimately, there was a need to be altruistic (De Waal, 2008). It seems that for this purpose the phenomenon of empathy has developed. Empathy is the human ability to feel for the other person in distress, and thus direct altruistic helping behavior to other human beings (Batson et al., 1983). As a byproduct, it also allows a person to gain intuitive knowledge of the social situation, and gives him an ability to act successfully in that social situation (Vignemont & Singer, 2006).

Evolution of empathy can today be seen in the way children develop emotionally. The ability to feel empathic reactions is central to emotional development (Bryant, 1987), and experiencing an empathic feeling for another person ultimately leads a human child to develop morality (Hoffman, 2000). In the adult world, empathic reactions underlie everyday human behavior, and even in the age of internet and increasing disconnect, we have developed social media to communicate our thoughts and feelings with words and pictures.

However, good empathic ability is not a given in every human being. There is a variance in general population (Davis, 1983; Eisenberg & Strayer, 1987) and differences in empathic ability are found to have effect in person's everyday life, for example in his social relationships (Settoon & Mossholder, 2002). Although differences in affective empathy has been studied in controlled experimental settings for a long time, utilizing measures ranging from skin conductance to neurological measurement tools, short, reliable measurements available to administer for large populations at this time consist largely of text-based self-report measures. It should be possible to make a brief test of empathic ability whose stimuli directly arouse empathic feelings in the test subjects, thus giving us a more direct look of test subject's ability to feel empathy. The purpose of this study is to investigate and develop such a tool for the assessment of affective empathic arousal.

1.1 Definition of Empathy

The German word for empathy, *einfühlung*¸ was coined in the 18th century by Johann Gottfried Herder to describe a way to acquire scientific information; his aim was to "feel into" the object of scientific inquiry, to make himself intuitively familiar with that object, rather than acquiring distanced empirical evidence of it (Edwards, 2013). Herder's, and later Robert Vishcer's attempts to get empathic subjective observation

recognized as a valid scientific method were however unsuccessful. The term *einfühlung* was translated into English as "empathy" in 1910's when Edward Titchener tried, once again unsuccessfully, to argue its place as a method for psychological study. After the Second World War, empathy resurfaced, this time as an object for empirical scientific study (for example: Dymond, 1948).

In a landmark study, Stotland (1969, p. 232) defined empathy as "observer's reacting emotionally because he perceives that another is experiencing or is about to experience an emotion." Much of the research in the following decade focused on the emotion as a vicarious emotional experience as a central theme. At the same time, literature about empathy as a cognitive skill, or taking another's perspective, was published (for example, Hogan, 1969). Davis (1983) brought both perceiving distress of the other and having an emotional reaction to it together in his Interpersonal Reactivity Index.

Nowadays it is seen that empathy comprises at least two different and dissociable processes: cognitive empathy and affective or emotional empathy (Baron-Cohen & Wheelwright, 2004¸ Blair, 2005). These two forms of empathy are neurologically dissociable: affective empathy is impaired in inferior frontal gyrus lesions whereas cognitive empathy is impaired in ventromedial prefrontal lesions (Shamay-Tsoory, Aharon-Peretz, & Perry, 2009: see also Bernhardt & Singer, 2012). Nancy Eisenberg (2002, p. 135), a long-time empathy researcher, defined empathy as "an affective response that stems from the apprehension or comprehension of another's emotional state or condition, and that is similar to what the other person is feeling or would be expected to feel". This definition also includes some constraints on the features of an empathic reaction: it isn't enough for a person to experience an emotional reaction when observing the other in distress, the valence of emotional reaction has to be logical given the emotion of the observed other.

1.2 Affective and cognitive bases of empathy

Cognitive empathy is the ability to understand the other's emotions (Jolliffe & Farrington 2006), or the ability to take the other's perspective (Davis, 1983): hence, it is sometimes referred to in the literature as perspective-taking. Cognitive empathy is closely related to the concept of Theory of Mind, or the ability to infer another ones mental state. Sometimes researchers equate cognitive empathy and theory of mind with one and another (for example, Baron-Cohen and Wheelwright, 2004). However, it would be best to treat these concepts as separate for the following reasons: 1) empathy has an emotional component whereas the theory of mind does not (to illustrate, Ibanez et al., 2013), 2) theory of mind is only about understanding that others have a mind which can be different from one's own (Premack & Woodruff, 1978), whereas empathy is about understanding what goes inside that mind, and 3) the active inferential

component present in theory of mind may not be necessary for empathy: cognitive empathy is more concerned about here-and-now judgments, whereas theory of mind as a concept refers to a concretely developed psychological structure.

Emotional or affective empathy – the focus of this study – is an appropriate emotional response to the other's feelings or situation (Baron-Cohen & Wheelwright, 2004; Davis, 1983). Being able to have an emotional response to emotions displayed in a social situation is imperative to better social functioning and is impaired, for example, in many cases of traumatic brain injury (de Sousa et al., 2011). An inappropriate or nonexistent emotional response to other one's suffering has been linked to psychopathy and Machiavellianism (the tendency to be deceptive, manipulative and cynical) (Ali, Amorim, & Chamorro-Premuzic, 2009). In people who display psychopathic traits but are not offenders ("successful psychopaths") cognitive-empathy abilities are not impaired, though they fall short on self-report measures of affective empathy: this seems to be related to the disability to feel shame (Mullins-Nelson, Salekin & Leistico, 2006).

Affective empathic reactions have many socially stimulating consequences. Much has been written on helping behavior and its connection with affective empathic reactions, and their central role in giving humane care, especially in psychotherapeutic relationships (Rogers, 1957; Kohut, 1984). In children, experiencing vicarious affective empathy reactions is related to prosocial behavior, especially in boys (Roberts & Strayer, 1996). In adults, empathy is found to be moderately related to not only prosocial behavior, but also socially competent behavior (Eisenberg & Miller, 1987). Empathic, other-oriented emotion seems to be a source for altruistic motivation. If, instead of other-oriented empathy reaction, a more self-oriented feeling of personal distress emerges, more egoistical motives usually arise (Batson et al., 1983). Associations between empathy and altruistic motivation have also been found on a neurological level (Mathur, Harada, Lipke, & Chiao, 2010).

Affective empathic reactions seem to underlie everyday prosocial behavior, and when those emotional responses are not found in adults, as noted before, they may be linked to mental illness. However, it is important to note that there may be other reasons for empathic reactions not emerging rather than mental illness. It seems that one needs both to perceive other as in need and to value other's welfare so much as to adopt the other's perspective for empathic emotional reactions to happen (Batson et al., 2007).

Because affective empathic reactions are so fundamental to social behavior, and because deficiencies in those abilities are related to antisocial behavior and traits, affective empathic reactions have been measured in many ways. What follows is a look on the measurements currently available for the measurement of affective empathy.

1.3 Scales and tests of affective empathy

1.3.1 Questionnaires of affective empathy

As previously noted, empathy was for the first time considered as an object of scientific study in the late 1940's. Dymond's (1949) scale of empathic ability was one of the first rudimentary attempts at constructing a scale for measuring empathy. In her scale empathic ability was operationalized as agreement between the test subject's prediction of other people's ratings of both the first subject's and the other person's personality, and the actual ratings. This test could nowadays be considered measuring some aspects of cognitive empathy, because it is based on how the test subject perceives what others feel about him and themselves. It however has not been tested for reliability and validity.

Two more widely used self-report scales of empathy were published in the turn of the 1970's: Hogan's (1969) scale of empathy, and Mehrabian and Epstein's (1972) measure of emotional empathy. Both have been widely used in the scientific literature in the last decades.

Hogan (1969) placed emphasis on insight, social acuity and perceptiveness when describing a person with high empathy: sample items of high empathic ability include "Is socially perceptive of a wide range of interpersonal cues", "Evaluates the motivation of others in interpreting situations", and of low ability "Does not vary roles; relates to everyone in the same way" . These clearly relate to the cognitive aspects of empathy and researchers in the following decades regarded Hogan's scale as a measure of cognitive empathy (Davis, 1983; Jolliffe & Farrington, 2004). It is criticized for its lack of variation, because it uses a two-point true/false-scale (Jolliffe & Farrington, 2004). It was also based on defining first groups of people with high and low empathy and then examining the differences between these two groups, using the so-called "criterion method", which wasn't considered a valid way for scale-construction in the 1960's (Bogartz 1965), and even less so nowadays.

Mehrabian and Epstein (1970) developed a Questionary Measure of Emotional Empathy (QMEE), using items that are supposed to measure a tendency to have vicarious empathic reactions, with items such as "I tend to get emotionally involved with a friend's problems" and "I like to watch people open presents". It has been suggested that Mehrabian and Epstein seem to be rather measuring sympathetic reactions (when one is moved to act one another's behalf) than empathy (Jolliffe & Farrington, 2004). Authors themselves (Mehrabian, Young, & Sato, 1988) have suggested that the measure is more about being aroused emotionally by the environment in general rather than other people's distress in particular. Mehrabian has later updated his empathy scale to a version called Balanced Emotional Empathy Scale (BEES): however, this scale has not been published in scientific literature, and articles citing it report that scale would be available from the author himself. Mehrabian's personal website states that "professor Mehrabian no

longer distributes his tests and cannot respond to messages about them" (as of April 2014), so unfortunately BEES cannot be reviewed here.

First attempt to make a self-report measure of both affective and cognitive empathy was Davis's Interpersonal Reactivity Index (IRI: Davis, 1983). IRI features four subscales: Perspective-Taking (sample item: "I sometimes try to understand my friends better by imagining how things look from their perspective"), Fantasy ("I really get involved with feelings of the characters in a novel"), Empathic Concern ("I often have tender, concerned feelings for people less fortunate than me") and Personal Distress ("Being in a tense emotional situation scares me"). It is however debatable whether Perspective-Taking subscale measures ability to understand what the other person feels: like QMEE, it seems to equate empathic reactions with sympathetic ones (Jolliffe & Farrington, 2006). Fantasy and Personal Distress subscales seem to be rather distant from the concept of empathy (Baron-Cohen & Wheelwright, 2004), and in the case of Personal Distress, actively working against empathic ability (Batson et al. 1983).

Empathy Quotient (EQ: Baron-Cohen & Wheelwright, 2004) was developed as a unitary self-report measure to fit in with Baron-Cohen theory of male and female brain (Baron-Cohen, Knickmeyer & Belmonte, 2005). Muncer & Ling (2006) investigated EQ using confirmatory factor analysis, and found that EQ seems to consist of subscales of cognitive and affective empathy and social skills. Although there probably is some linkage between empathy and social skills, they are completely different concepts: for example, in psychopathic individuals, empathy is thought to be absent, but social skills (manipulating others) very much present. Therefore, it cannot be said that a measure taking account the social skills of the subject is providing a strict view of empathic tendencies of the subject: rather, EQ as a whole seems to be measuring something like being able to function socially.

Jolliffe and Farrington (2006) developed their own Basic Empathy Scale (BES) based on their criticism of aforementioned Hogan's scale, QMEE and IRI. Their items include assessment of general affective and cognitive empathy (sample item for affective empathy: "I get caught up in other people's feelings easily") and for five basic emotions specifically (sample item for affective empathy in the fear emotion: "I usually feel calm when other people are scared" (disagree)). They present their validation analysis concerning research on adolescents. It appears to work well as a method of assessing the empathic abilities of the adolescent, having French (D'Ambrosio, Oliver, Didon, & Besche, 2009) and Italian (Albiero, Matricardi, Speltri, & Toso, 2009) versions validated. However, it was developed with express intent to examine juvenile delinquents, and so much of the research done using BES examines adolescents with antisocial tendencies. Whether BES could be a tool to measure empathy of adults in the general population remains to be seen.

All of the above, whatever their strengths and weaknesses, are self-report measures that use text-based items. According to Batson (1987), all of the self-report measures of empathy face the problem that they don't account for difficulties in reading comprehension or interpretation of context in which the words appear. Moreover, being text-based makes the ecological validity of self-report measures questionable: although empathetic reactions to text, for example when one is reading a novel, plausibly exist, they don't characterize more commonplace processes of viewing the other in distress. It is obvious that a more direct approach to measuring empathy is needed, and although physiological (for example, skin concordance in Marci, Ham, Moran, & Orr, 2007) or neurological (for example, Singer et al., 2006) measures could be one avenue, their usefulness is hampered by their costliness and inaccurateness: how do we know the autonomous nervous system would be stimulated precisely because of empathic reactions? There is a need for a cost-effective, ecologically valid measure of affective empathy.

1.3.2 MET (Multifaceted Empathy Test)

There might be a middle ground between text-based self-report measures and direct biological measurements and this middle ground was employed by Dziobek et al. (2008). They used a photo-based measure to assess both cognitive and affective empathy. Picture stimuli were used for a better ecological validity than text-based self-report measures:

The stimulus in MET consists of 26 pictures of a person experiencing an emotion in a specific context. First subjects are presented with a context picture and asked to rate their emotional arousal to the context. Only then are they presented with a picture of a person placed into the context and asked to select from four different options the feeling-state that they think the person is in (a measure of cognitive empathy). Then the subject is presented with a feedback of their performance in the assessment of feeling-state of the person, i.e. they are told what the person is actually feeling. Then they are asked in an indirect way (How calm/aroused does the picture make you feel?) and then in direct way (How concerned are you for this person?) about their emotional arousal (two measures of affective empathy): the arousal to the context only, which was asked in the first step of the process, is also taken into account.

Indirect, implicit way of questioning was included so as to minimize the demands on self-reflecting at an abstract level and to reduce the likelihood of socially desirable answer patterns. I think that it is a good starting point that our wording of the question should be as direct and naturalistic as possible. This should reduce the chances that the respondent doesn't understand the question. It is however debatable whether social desirability is a source of distorted data (because of response style), or if it should be considered as substance (McCrae & Costa, 1983; Lönnqvist et al., 2007): after all, social desirability is correlated with

values of warmth and communality (for example, Schwartz, Verkasalo, Antonovsky, & Sagiv, 1997). All in all, the two different ways of questioning didn't cause a difference in Dziobek et al.'s (2008) research question about differences in empathy between autistic and control respondents. In a later, adapted version of MET, respondents are only asked what the person is feeling (Kirchner, Hatri, Heekeren, & Dziobek, 2011).

Overall, although MET has unique properties in the field of empathy measurement, it has a very complicated design. This may also hinder the validity of obtained assessments. Because in MET assessment of emotional empathy is done after test subjects have been told that the person they are viewing is in distress, it is debatable whether respondents are reporting affective arousal based on the picture stimuli or based on the concept of negative emotion they are told is associated with the picture stimuli. It would probably be for the best to assess cognitive and affective empathy with separate tests, other geared towards the appraisal of other people's emotional states and the other on the subjective emotional arousal when viewing the other in distress.

However, MET has some interesting ideas on how to go about examining subjective emotional arousal in search of quantitative scores for affective empathy. Using the picture stimuli of a distressed person is a lot more ecologically valid measure of affective empathy than asking the respondent whether they generally react emotionally to other people's distress. Met also used specific, universal negative emotions to elicit the affective response. It is probably better to use negative rather than positive emotions, because viewing other person experiencing happy emotions activating empathic reactions has been studied relatively little (see Jabbi, Swart & Keysers, 2007, for an exception), while empathic reactions to pain and suffering are more widely understood (Singer et al., 2004; Singer, et al., 2006; Saarela et al., 2007; Han et al., 2009) – indeed, most definitions of empathy include observing the other experiencing negative emotions, or "in distress"(Batson et al., 2007).

1.4 Aims of Pictorial Empathy Test (PET)

Previously discussed MET is made for use in an experimental laboratory setting, and thus is not a viable way to gather data from a large group of people. However, its use of pictures in trying to elicit emotional reaction of empathic concern in the test subject is an ecologically valid way of producing empathic reactions: seeing distress on the face of the other is a natural, neurological "resonance mechanism" for human beings (Balconi, Bortolotti & Gonzaga, 2011). Therefore there is a need for a pictorial measure of affective empathy. This was the starting point in the development of Pictorial Empathy Test (PET). PET includes free-to-use pictures with emotional content, and asks the subject to report his level of empathic affective arousal. This should be an easy way to gather data on the degree of affective empathy

experienced by the test subject. A simple image and a question about emotional arousal should be expected to be understood by all participants. This kind of measure can be deployed via web-based questionnaire, and thus should not be as cumbersome as some test measures of empathy, for example skin conductance or fMRI. Thus, this kind of test measure should be possible to use as a part of large data gathering and results should be readily usable.

The present study examines the reliability and validity of PET. Firstly, it should be demonstrated that the test in question has a good internal reliability. Moreover, modern computer modeling allows us to explore whether the hypothesized latent factor structure of the PET is to be found in the gathered data.

As for the validity, Cronbach and Meehl (1955) argued that for a test to be considered a valid measure of a given trait, it should be demonstrated that it relates to measures of other constructs in theoretically predictable ways. Campbell and Fiske (1959) proposed that, when measuring test validity, new test measure should be judged against other measures of the same and other traits. Using a correlation matrix, it should be determined that tests of constructs that theoretically should be correlated, are correlated with the test in question: this is referred to as convergent validity. It should also be determined whether tests of constructs that shouldn't correlate with the test under consideration actually do not: this is referred to as divergent validity.

Borsboom, Mellenbergh, & van Heerden (2004) put test validity slightly differently: for them, a test can be said to be a measure valid measure of a given trait if it a) the trait it is measuring is a truly existing attribute (based on previous research) and b) variations in attribute are linked to variations in measurement outcomes (previously mentioned convergent and divergent validity). They also noted that it is important to construct a clear theoretical network of what should happen in test scores when different people with differing abilities take the test. So when constructing hypotheses for PET's validity, we should be able to find, based on previous research, concepts that should correlate with affective empathy (convergent validity) and concepts that should not correlate with affective empathy (divergent validity).

1.5 Correlates of affective empathy

1.5.1 Sex and gender roles

Stereotypically it is seen that females are much better at inter-personal skills and socialization and thus better at empathizing with other people (for example, in Howe, 2013, p. 63), and many researchers take for granted sex differences in empathy (for example, Schulte-Rüther et al., 2008). There are also several scientific theories that suggest so. Baron-Cohen, Knickmeyer, and Belmonte (2005) proposed subsequently

influential theory on sex differences in the brain. They postulate a fundamental continuum with a neural base on individual differences on the capacity to empathize. Crespi and Badcock (2008) have also postulated a sex difference between males and females in mentalistic cognition which involves abilities of interacting with social environment, of which tasks empathic reactions have great bearing.

Based on the work done on these theories females report greater empathy than men in self-report measures of empathy (Baron-Cohen & Wheelwright 2004; Wheelwright et al. 2006), at least when the domain of emotional reactivity is tested (Muncer & Ling, 2006). This difference in self-report measures has been found earlier (for example, Davis, 1983): however, the difference is not found in tests relying on more direct measures than self-report (for reviews, see Hoffman, 1977, and Eisenberg & Lennon, 1983). This can also be found in a study by Baron-Cohen, Wheelwright, Hill, Raste, & Plumb (2001): when testing differences in ability to recognize emotions in the eyes on general population they found a minuscule sex difference that only approached statistical significance, which was likely due to the enormous size of the data set. Two recent neuroimaging studies seem to find no difference in immediate empathetic affective reactions between sexes (Singer et al., 2006; Han, Fan & Mao 2008). Empathic response may be more enduring and less fragile in females than in males (Singer et al., 2006), but this may be due to motivational issues related to perceived gender roles (Klein & Hodges, 2001).

Jolliffe and Farrington (2006) noted that self-report measures of empathy usually find a significant difference between male and female respondents but that it is unclear whether there is a direct causal link between sex and empathy or whether other factors, such as the gender role, mediate the relationship. In a world where socialization to gender roles is maybe as big a factor in the developing of gender differences as are biological differences in the brain, it is necessary to try to differentiate between the two. Ickes, Gesn, & Graham (2000) performed a quantitative meta-analysis, and found that sex differences only showed up when subjects were aware that they were tested for their empathic ability and when gender-role expectations were made salient. Karniol, Gabay, Ochion, & Harari (1998) found in their study examining the relationships between gender, gender-role and empathy that gender was found to be unrelated to empathy when gender role (masculinity/femininity) was taken into account. Femininity correlated with both affective and cognitive empathy, and masculinity was found to be uncorrelated with empathy. Clearly, associating gender role and empathy has stronger foundation than associating empathy and sex directly. It is therefore expected that empathic reactions should correlate with a measure of gender role. It is also expected that the relationship between sex and empathy is mediated by the gender role.

1.5.2 Social intelligence

Empathy and social intelligence have been seen as related, even so that ”empathy underlies social intelligence” (Hogan, 1969). Nowadays these two abilities are seen as separate, empathy being about understanding other’s thinking and feeling and reacting accordingly, whereas social intelligence has elements of being able to behave successfully in social situations in addition to being able to understand other people and their motivations (Silvera, Martinussen & Dahl, 2001). Although they are clearly distinct domains, they overlap: cognitive empathy may be seen as one aspect of social intelligence (Kaukiainen et al., 1999). At least, both empathy and social intelligence involve the ability to adequately assess other people’s internal states. There however doesn’t seem to be empirical research directly investigating the relationship between social intelligence and empathy. However, there seems to be correlation with the closely related, but a little bit narrower concept of emotional intelligence and empathy (Austin, Evans, Goldwater & Potter, 2005) and correlations between emotional intelligence and successful social behavior are well established (Brackett, Mayer & Warner, 2004; Schutte et al., 2001). Emotional intelligence also correlates with the development of theory of mind (Qualter, Barlow & Stylianou, 2011). It is therefore expected that measure of social intelligence should be somewhat related to a measure of emotional empathy.

1.5.3 Empathy and intuitive thinking style

Empathic reactions provide us with an immediate knowledge of another person’s situation (Vignemont & Signer, 2006). Therefore it should be expected, that tendency to greater empathic reactions could be related to the tendency to reason intuitively, as opposed to analytic and logical thinking. It has been seen that overall emotional arousal in human decision-making is connected with greater levels of intuitive thinking (Eliades, Mansell, Stewart, & Blanchette, 2012; Blanchette & Richards, 2004). It has also been found that physiological arousal produced by stimuli with a negative affect hinders logical thinking, which emotionally neutral stimuli do not (Blanchette & Leese, 2011). Also, thinking with fast and heuristic sensibilities has been found to be correlated with emotional expressivity (Pacini & Epstein, 1999). In neuroimaging studies, the so-called “X-system” of brain areas associated with fast and heuristic thinking is found to be associated with empathic affective reactions (for a review, see Lieberman, 2007). However, only one study (Norris & Epstein, 2011) has directly measured the relationship between a self-report of empathic reactivity and tendency to think intuitively or analytically. They found an association with intuitive thinking and affective empathic reactions, and no association between analytic thinking and affective

empathic reactions. Thus, it is also expected in this study that intuitive thinking style is related to emotional empathy. Moreover, it is expected that emotional empathy does not correlate with analytic thinking style.

1.5.4 Empathy deficits in autism spectrum disorders

Baron-Cohen, Knickmeyer and Belmonte (2005) postulate in their previously mentioned theory that people with autism spectrum disorders on the whole seem to be better at abilities involving the non-living world, for example at math calculation, rather than empathizing with human beings. Therefore people with symptoms of autism spectrum disorder (ASD) should score lower on a test of empathy. There seems to be different results on whether lack of emotional responses, including empathic emotional responses, correlates with ASD or not (Bölte, Feineis-Matthews, & Poustka, 2007; Fan et al., 2013). There are also different results on whether it would be just cognitive or both cognitive and emotional empathy that are related to ASD (Mathersul, McDonald, & Rushby, 2013; Dziobek et al., 2008). However, because cognitive empathy, the ability to know what the other person thinks and feels, is a necessary condition for an affective empathic response to arise, difference should be nevertheless be seen on affective empathy between people high or low on ASD symptoms. Gillespie, McCleery, & Oberman (2014) have recently suggested that there might be a fundamental difference for people with ASD in empathic abilities towards familiar versus unfamiliar faces, with unfamiliar faces provoking no empathic reactions in people with ASD. Thus, when the measurement of affective empathy involves people that are not familiar to the test subject, people who score high on a measure of ASD should score low on a measure of affective empathy.

1.6 Hypotheses

To test the reliability and the divergent and convergent validity of PET, the following hypotheses were set:

Firstly, latent one-factor structure of PET was hypothesized to be a good fit to the gathered data. It was also hypothesized that 2) females score higher in PET than men, 3) individuals with feminine gender role score higher in PET than others, 4) gender role mediates the relationship between PET scores and gender, and that 5) PET scores do not correlate with masculine gender role. It was also hypothesized that PET scores correlate positively with 6) self-reported empathy, 7) social intelligence, and 8) intuitive thinking style. Finally, it was expected that PET scores do not correlate with 9) analytic thinking style, and that 10) PET scores correlate negatively with symptoms of autism spectrum disorders.

2. Method

2.1. Pilot study

Participants in the pilot study included 91 respondents, of which men were 40 and women 48 (with missing sex data on three participants). Participants were between ages of 23 and 71, with mean age of 42.90 and standard deviation of 14.62. The mean age of men who participated was 37.89 and of the women 47.10. Participants for the pilot study were recruited using a convenience sample.

The pilot study of PET consisted of 22 pictures representing equally men, women, girls and boys. All of the pictures were licensed with a free-to-use license and were found in Wikimedia Commons. Scale of the pilot study was a five point scale answering the question "How touching do you find the photograph #?"; answer ranged from 1 = "Not at all", to 5 = "Very much". Pictures were selected to elicit an experience of viewing the other as a subject and in distress: they place a clear view of the person in a vulnerable state as the main subject of the photo. Based on previous research (Håkansson, 2006; Dziobek et al., 2008), these kinds of photos are ideal for eliciting empathic emotions. Purpose of the pilot study was to find most reliably touching set of pictures to use in the final version of PET.

Goal of the pilot study was to narrow down pictures used to a small amount of pictures with high reliability and variance to use in the final version of PET. One picture from the original 22 pictures was eliminated because its' availability for free use became questionable. After this, 7 pictures were left off for eliciting less emotion than the other pictures. After this the amount of pictures was reduced so, that final set of stimulus still had reliability of around .90 and that men, women, girls and boys were represented in the pictures. Final set of 8 pictures has two pictures of women, two pictures of men, two pictures of boys, one picture of a girl, and one picture of a baby (sex not identifiable in the picture).

Assessments of the pictures were found to have the following correlations: women found pictures more touching than men ($r = .24$, $p = .03$); older people found the pictures more touching than younger ones ($r = .25$, $p = .02$); and if the test subject had children of their own, they found pictures more touching than those who didn't ($r = .36$, $p < 0.001$). These correlations did not however influence the selection of pictures for Final PET, because then only neutral pictures could have been used if these correlations were to be left out.

2.2. Main study

2.2.1. Participants and procedure

Two thousand seven hundred and eighty-nine Finnish participants (65% females) took part in the study. Their mean age was 28 years (SD = 8.87, range 15–69). Of the participants, 27% were working, 64% were students, and 9% were otherwise occupied. Of the students, most were university students (85%) but polytechnic (7%), vocational school (4.5 %), upper secondary school (3%), and grammar school (0.5%) students were also included in the sample.

Of the 3086 people who originally took part in the study, two were excluded because their comments about the study revealed that they had not completed the survey seriously. In addition, sum variables were not calculated for participants who had 25% or more missing items in the scale to be computed. Because the survey was long (including also scales and tasks not reported here), many participants skipped one or more scales, resulting in a loss of 295 participants.

Data was collected via web-based questionnaire. The participants were recruited to the on-line study via several open internet discussion forums, several student mailing lists, and from a participant pool comprising individuals who had expressed an interest to participate in psychological studies. No exclusion criteria for participation were applied. The participants were told that the study concerned thinking and personality, and confidentiality and voluntary participation were emphasized. In the messages sent out to the internet forums and mailing lists, a hyperlink to the questionnaire was included. The respondents were given 3 weeks' time to participate in the study. As a compensation, all participants received a thinking style profile based on the Actively Open-Minded Thinking Scale (Stanovich & West, 1997) included in the survey, but not used in this study.

2.2.2. Material

2.2.2.1 Final PET

The participants were presented with a set of 7 pictures of people (women, men and children) in distress. All of the pictures were found on Wikimedia Commons, with searches including "emotions", "war", "death", "disabled persons", "scared", "fear" and "fright". All of the pictures are free to use (see Appendix), and licensed with Creative Commons license – this makes PET also free to use and freely distributable test. Originally, the participants rated 8 pictures. However, as we noted only later, one picture depicted a dead

woman, and because it is not reasonable to assess empathy towards the dead, the ratings of this picture were excluded from the analyses and from the final test.

After every picture subjects were asked to rate on a five-point Likert scale their answer to the question "How touching do you find the photograph #?" Possible answers include: 1 = "Not at all", 2 = "A little bit"; 3 ="It arouses some feelings", 4= "Quite a lot", 5 = "Very much". To obtain the PET score, a mean score of the 7 answers was calculated.

2.2.2.2 Convergent and Divergent validity measures

To assess PET's relation to self-reported empathy, Empathy Quotient (EQ) was used. EQ is a self-report measure of empathy (Muncer & Ling, 2006). It contains 15 to which test subjects respond with a four-point scale (1 = "strongly disagree", 2 = "slightly disagree", 3 = "slightly agree", 4 = "strongly agree"). In this study included were items that have been demonstrated to fall into three different factors: emotional reactivity, cognitive empathy and social skills (Muncer & Ling 2006). Items on emotional reactivity include questions on how greatly the respondent reacts to social or emotional stimulus (friend's problems, movies). Items on cognitive empathy reflect respondent's ability to intuitively understand and predict other people's emotions. Items on social skills reflect the respondent's ability to function in social situations. Reliability for the scale, measured by Cronbach's alpha, for full Empathy Quotient sum-measure was .81, and for subscales of emotional reactivity .67, for cognitive empathy .79 and for social skills .71. On emotional reactivity and cognitive empathy these are slightly lower than in Muncer & Ling's study (2006). Cronbach's alpha for social skills subscale is better than in Muncer & Ling's study.

Social intelligence was assessed with Tromsø Social Intelligence Scale, which is a self-report measure of social intelligence (Silvera, Martinussen & Dahl, 2001). It contains 21 items on social information processing, social skills and social awareness (sample items: "I fit in easily in social situations", "I know how my actions will make others feel", "I find people unpredictable"(disagree)). Respondents were asked to rate each item from 1 ("Describes me extremely poorly") to 7 ("Describes me extremely well). Cronbach's alpha for the scale in this study was good, .90, as it was found to be in Grieve & Mahar (2013).

Gender role was assessed with Bem Sex Role Inventory (SRI). SRI is a self-report inventory which is designed to reflect test subjects masculinity, femininity and androgyny (Bem, 1981). It contains 20 items on typical masculine or feminine characteristics. Cronbach's alpha for the inventory in this study is for the femininity scale .89 and for the masculinity scale .84: these are comparable to previous found reliability estimates for the scales (Holt & Ellis, 1998). In SRI, subjects were asked if certain descriptions fit them (sample item for

femininity scale: “compassionate”, for masculinity scale: “independent”), and respondents answered on a seven-point scale, from 1 = “never”, to 7 = “always”.

Intuitive and analytic thinking was assessed with Rational/Experiential Inventory (REI). REI is a scale measuring with self-report how typical for the respondent is to think either rationally or experientially (Norris & Epstein, 2011). Part of Experiential thinking is Faith in Intuition-subscale, consisting of questions on how respondent uses and trusts initial, intuitive feelings and impressions: sample items include “I often go by my instincts when deciding on a course of action”, and “I trust my initial feelings about people”. Other subscale used in this study is Need for Cognition-subscale that consists on questions of a preference for analysis and thinking through problems: sample items include “I enjoy intellectual challenges”, and “Using logic usually works well for me in figuring out problems in my life”. Items were rated by the respondents using a four-point scale, from “strongly agree” to “strongly disagree”. Cronbach’s alpha for the Faith in Intuition subscale in this study was .79, and for Need for Cognition subscale .86, similar to what Norris & Epstein (2011) obtained.

Asperger symptoms were assessed using Autism Spectrum Quotient (ASQ). ASQ is a self-report measure of autistic traits, originally developed by Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley (2001). This study uses the 28-item abridged version (Hoekstra et al., 2011). It includes items on social skills (sample item: “I would rather go to a library than to a party”), preference for routines (“I prefer to do things the same way over and over again”), difficulties in switching (“I frequently get strongly absorbed in one thing”), and difficulties in imagination (“I find it difficult to imagine what it would be like to be someone else”). Items are rated on a four-point scale, from “strongly agree” to “strongly disagree”. Reliability for the scale in this study was good (Cronbach’s Alpha=.81), as in Hoekstra et al. (2011).

3. Results

3.1 Reliability of PET's one-factor structure

The Internal reliability of the PET scale was measured by Cronbach's Alpha, which was .90. This is a strong internal reliability for a seven-item scale.

The factor structure of PET was investigated further using confirmatory factor analysis (CFA). Structural Equation Modeling (SEM) analyses were run on AMOS Graphics 21.0 software. SEM models are thought to have good fit, if it has absolute fit, evaluated by Chi-Square test where the null hypothesis is that model fits the data, and is rejected if $p < .05$. Another absolute statistic usually calculated is standardized root mean square residual (SRMR), which is the discrepancy between the correlations in the observed matrix and the correlations between the model: SRMR value below .05 would indicate good fit. Moreover, usually comparative fit is investigated, using comparative fit index (CFI), which should be higher than .95, and root mean error square statistic (RMSEA), which should be lower than .08 for an acceptable fit, and lower than .05 for a good fit.

Specified model included one latent factor (affective empathy), of which all seven items of PET were considered equal predictors, and for all of which a latent error variance factor was included. This original model showed marginal comparative fit and SRMR, but poor absolute fit as investigated by Chi-Square and also poor root mean error square statistic (X^2 (6) = 574.39, $p < .001$; SRMR = .04; CFI = .95; RMSEA = .12). Using modification indexes, model was corrected to include covariance in the error terms of PET items: this was only done in cases where there was a shared subject matter between the used photos (PET photographs one, two and five display profound sadness in subject's face; pictures three and four include very young children; pictures one, five and seven all feature multiple people, one grieving or helping the other who is injured or dead; pictures five and six elicit empathy for a young victim; six and seven feature mild levels of grisly imagery). Correlation residuals between the photographs were low, which is further proof that the poor original fit was caused by method variance. This reflects the fact that PET is designed to be an ecologically valid measure: this means that it is not possible to find pictures with orthogonal levels of measurement error covariance. The corrected model showed excellent comparative fit and SRMR, with acceptable root mean error square, but still poor absolute fit. (X^2 (6) = 79.65, $p < .001$; SRMR = .01; CFI = .99; RMSEA = .066) However, it is usually difficult to get good absolute fit when dealing with large samples (Bentler & Bonett, 1980).

3.2 Convergent and divergent validity of PET

In order to evaluate the convergent and divergent validity of PET, Pearson Correlations were calculated between PET scores and other measures representing theoretical constructs that theoretically should or should not correlate with a measure of affective empathy. Results are presented in Table 1.

Table 1

Pearson correlations between PET scores and the scores of other measures included in the study

Measure	Pearson Correlation with PET Scores
Full EQ Scale	.48***
EQ Subscale Emotional Reactivity	.53***
EQ Subscale Cognitive Empathy	.30***
EQ Subscale Social Skills	.26***
Faith in Intuition	.26***
Need for Cognition	-.08***
Tromsø Social Intelligence Scale	.29***
Autism Quotient	-.32***
Feminine sex role identity	.51***
Masculine sex role identity	0.02

*** p < .001

When compared with a self-report-measure of empathy (EQ), PET was found to have strong correlation. Assesment of EQ subscales found that this correlation was mostly due to EQ's emotional reactivity subscale, which assesses affective empathy, as PET does. Other EQ subscales were found to have weak, but still significant correlations. These findings supported the Hypothesis 6.

PET scores only weakly correlated with intuitive thinking. This reflects the two concept's weak, but significant relationship, thus supporting the Hypothesis 8. Assessment of REI's Need for Cognition-subscale found a statistically significant correlation ($p < .001$) contrary to the stated Hypothesis 9 : however, the absolute value of correlation is very small (-.08), so the statistical significance can probably be attributed to the size of the data set. Relationship between PET scores and social intelligence was found to be weak, but statistically significant, thus supporting the Hypothesis 7. There was also a weak negative correlation between PET scores and autism spectrum symptoms that supports the Hypothesis 10. All in all, correlations between PET scores and scores of other measures were in the same direction that they were hypothesized to be.

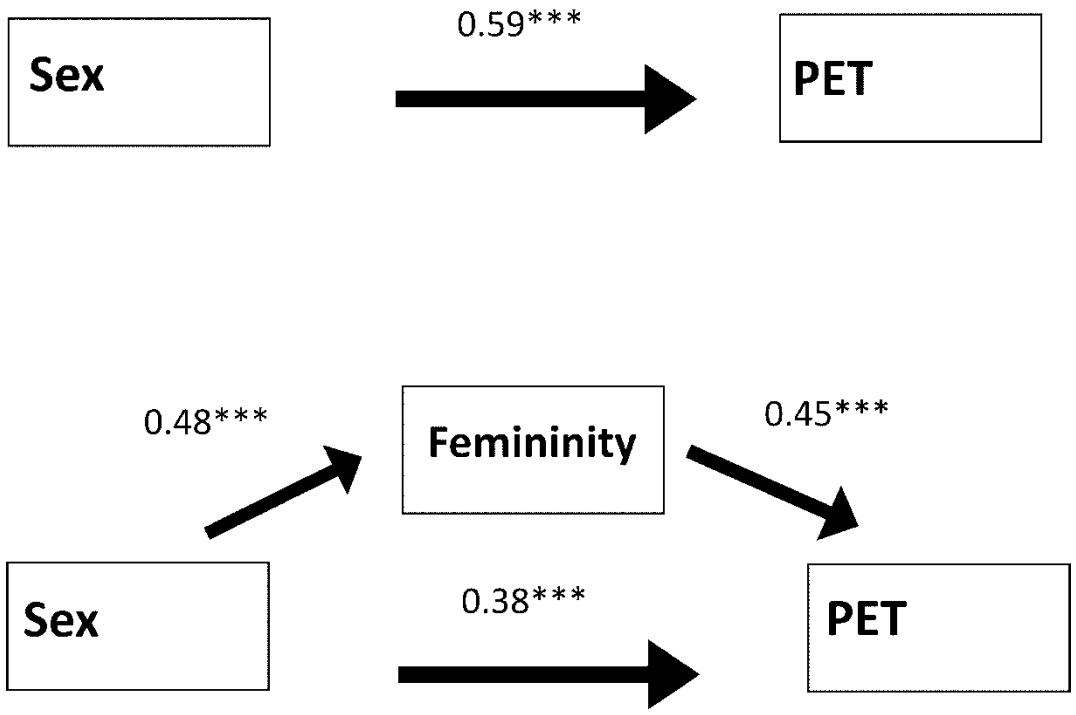
3.2.1 PET scores, gender roles, and sex differences

Feminine gender role was strongly related to PET scores (.51, $p < .001$). There was no correlation between PET scores and masculine gender identity (.02, $p = .40$). These results support hypotheses 3 and 5. Using an independent samples t-test, it was determined that men scored significantly lower than women in PET, (2790) = -16.40, df(2790), $p < .001$, with men's mean score being 3.14 and women's 3.73. This result supported hypothesis number 2. To analyze whether female sex role mediates the relationship between sex and PET scores, a mediation analysis was carried out using Preacher & Hayes's (2004) SPSS Macro. This analysis enables us to investigate all of the steps for testing mediation (Baron & Kenny, 1986) and to obtain the Sobel test (Sobel, 1982). According to Baron & Kenny (1986), to prove a mediating effect, one must 1) show that independent variable (here participant' sex) significantly influences the mediating variable (here feminine gender role, as measured by SRI) 2) the independent variable significantly influences the dependent variable (here PET scores), 3) mediating variable has a unique effect on dependent variable and 4) the effect of the independent variable on the dependent variable shrinks when the effect of the mediating variable is controlled. If the effect of the independent variable in the fourth step isn't significant after the effect of the mediating variable is controlled, a full mediation is said to have occurred. If the effect shrinks but remains significant, partial mediation is said to have occurred. In the latter case, Sobel test (Sobel, 1982) has to be performed, to determine whether the effect of the independent variable via the mediating variable on the dependent variable significantly differs from zero.

Unstandardized regression coefficients are reported in Figure 1.

Figure 1

Mediating effect of feminine gender role on the relationship between participant's sex and PET scores



Participant's sex had an effect on the PET scores (B = 0.59, t = 16.47, p < .001), and on the feminine gender role in the SRI Feminine sex role scale (B = 0.48, t = 12.64, p < .001). Feminine gender role had a unique effect on the PET scores (B = 0.45, t = 28.23, p < .001). However, when the effect of feminine gender role on PET scores was controlled, the effect of participant's sex on PET scores was still significantly different from zero (B = 0.38, t = 11.58, p < .001). Sobel test indicated that the indirect effect from participant's sex through feminine gender role to PET scores was significantly different from zero (z = 11.53, p < .001). This confirms a partial mediation on the effect of the participant's sex on PET scores via feminine gender role. However, participant's sex had still a significant direct effect on PET scores even after the mediation was taken into account.

4. Discussion

The purpose of this study was to examine the reliability and validity of Pictorial Empathy Test (PET). The hypotheses examined were largely upheld, thus indicating that PET is a reliable and valid measure of affective empathy.

The internal reliability of PET was found to be good. Moreover, the model with a one-latent-factor structure behind PET scores was found to have a moderate fit to the data gathered, implying that the variance in all of the pictures in PET was because of differences in one latent factor, hypothesized as affective empathic arousal. However, we had to assume small error covariance behind some of the pictures, indicating that the level of arousal between the subsets of seven pictures wasn't completely unrelated. When the purpose is to develop an ecologically valid measure of a given trait, this is acceptable: it would be impossible to find a diverse set of emotionally stimulating pictures whose effects are completely unrelated from one another except for the one factor that we are studying.

The relationships between the PET scores and the other measurements of empathy and the related concepts were found to be in the hypothesized direction, supporting thus divergent and convergent validity of PET. The PET scores were related to self-reported empathy, and especially to the emotional reactivity component of self-reported empathy. The association between the PET scores and cognitive empathy component was weaker, lending credibility to the already prevailing notion that cognitive and emotional empathy are dissociable processes (Blair, 2005; Shamay-Tsoory, Aharon-Peretz, & Perry, 2009), and consequently to PET's ability to measure especially affective empathy.

The relationship between the PET scores and social intelligence was positive and significant, but weak. The relationship was as hypothesized, and the fact that PET is positively related to social intelligence lends credibility to PET's validity. There is currently little research literature between affective empathy and social intelligence, and this relationship should be further investigated: could one develop and enjoy being socially intelligent without feeling for the other person?

Intuitive thinking and the PET scores were positively related whereas a preference for an analytic thinking style and the PET scores were unrelated. These findings offer further support for PET's validity as a test of affective empathy, because it has been shown that intuitive thinking and emotionality are related (Eliades, Mansell, Stewart, & Blanchette, 2012; Blanchette & Richards, 2004). However, intuition is a very large and multifaceted concept, and though some of its facets are related to emotional arousal (and the PET scores), it has been argued that some of its facets are not (Glöckner & Witteman, 2010). It might be that there is some facet of intuition that is specifically related to emotional empathy, or that the capacity to emotional

arousal in general is reflected on PET scores, and therefore in the relationship between PET scores and intuitive thinking. Analytical thinking however seems to be unrelated with empathic reactions.

PET scores were strongly and positively associated with feminine gender role, but not with masculine gender role, as hypothesized. Feminine gender role identity was responsible for a significant part of the relationship between participants' sex and their PET scores. However, it should be noted that gender role did not explain all of the variance between sex and the PET scores opposed to what Karniol et al. (1998) found in their study. In contrast to Karniol et al. (1998), this study used the mediation analysis principles set by Baron & Kenny (1986), so the conclusion in this study is made with more widely adopted practices in mediation testing.

The fact that a significant amount of the difference in PET scores wasn't explained by female gender role suggests that there is a direct relationship between one's sex and the nature of affective empathic reactions. Thus, this study lends some credibility to Crespi & Badcock's (2008) theory of genuine sex differences in emotional abilities. However, Eisenberg & Lennon (1983) found in their meta-analysis that sex differences in empathy were a function of the method used to assess empathy, and sex differences found in this study may be a function of PET's nature as something between a direct self-report and a neurological measure. In Singer et al. (2006), initial neural responses between the sexes observing the other in distress seemed to be similar, but when given a reason not to feel empathy for the other, males were seen to suppress their empathic emotions, whereas females did not. Maybe male subjects do the same in PET, not reporting as much of the initial empathic arousal. However, it should be noted that during the administering of PET, participants were not explicitly told that empathic reactions were being investigated, and they were not made aware of gender-stereotypical expectations: these were, after all, suggested to be the causes of sex differences in self-reporting empathic tendencies by Ickes, Gesn, & Graham (2000).

As a whole, the results discussed above support the reliability and validity of PET. It seems that PET may be a good alternative to text-based questionnaires: using a picture stimulus creates an opportunity for empathic emotional arousal in the moment the test is administered, whereas text-based measures only ask test subject to think back and figure, whether he might have had empathic emotions previously. Problems with text-based measures in the assessment of empathy were noticed by Batson (1987) almost thirty years ago, yet other kinds of methods are not available in the present day to use alongside text-based self-report measures. Dziobek et al.'s (2008) development of MET showed that using facial expressions of distress in human beings is a potent way to acquire information on affective empathic reactions. Seeing distress in the face of the other is a natural, neurological "resonance mechanism" for human beings (Balconi, Bortolotti & Gonzaga, 2011), and PET takes this natural resonance mechanism to the task.

In contrast to neurological measures, in administering PET we can only see the final answer by the participant. Thus PET can't be considered replacing neurological or skin-conductance measures, and is akin to self-report measures in that participant may report all of the emotional arousal he experiences. However, the goal of PET was not to replace these direct laboratory measurements of arousal, but to develop an ecologically sound measure for the testing of large masses of participants. PET is short enough and easily answerable, so it is not a gargantuan task to ask the respondent to complete.

In this study convergent and divergent validity measures have been investigated, and the results support the hypothesis that PET is measuring affective empathic reactions. It could, however, be beneficial to use PET as a part of larger confirmatory analysis on the construct validity of affective and cognitive empathy in the future: in a specifically gathered data, a multitrait-multimethod analysis (Fiske & Campbell, 1959) could be carried out. Multitrait-multimethod analysis would be a robust way to give more evidence of the validity of several tests and self-reports of empathy and especially of the latent factors that are theorized to be measured by these tests and self-reports. This kind of analysis however requires multiple, traditionally at least three, different measures of a given "trait", and in at least two traits. Thus, data to run multitrait-multimethod analysis needs usually to be acquired with the specific goal of running multitrait-multimethod analysis in mind, making this kind of analysis very costly.

However, the idea of affective empathy as a latent construct has not been questioned in research literature. Admittedly, emotional empathic reactions are not currently investigated as vigorously as many more recent concepts, such as social intelligence. Relations of these new concepts to the well-being of individuals and society may not be well understood, whereas the value and place of affective empathy, in contrast, are clearly evident. The pull to feel for a person in distress is both intuitively understood by many humans, and is backed by a cumulative body of evidence from half a century of research. This empathic feeling for the other is important, because it seems to explain why human beings would be capable of putting other person and his distress above egoistical self-interest (this being the case made by Batson, 1991).

In sum, PET is a free-to-use, easy way to assess affective empathy. It is a valid and reliable measure of emotional empathic reactions. It can be deployed via a web-based questionnaire and it is thus not cumbersome like some test measurements of empathy, for example skin conductance or fMRI. It has ecological validity and easily understood interface that makes it a good alternative to text-based self-report measurements. It is possible to use PET as a part of large data gathering and the results are readily usable. PET could be a significant tool in the assessment of affective empathic reactions during forthcoming research.

References

Albiero, P., Matricardi, G., Speltri, D., & Toso, D. (2009). The assessment of empathy in adolescence: A contribution to the italian validation of the “Basic empathy scale”. *Journal of Adolescence, 32*(2), 393-408.

Ali, F., Amorim, I. S., & Chamorro-Premuzic, T. (2009). Empathy deficits and trait emotional intelligence in psychopathy and machiavellianism. *Personality and Individual Differences, 47*(7), 758-762.

Austin, E. J., Evans, P., Goldwater, R., & Potter, V. (2005). A preliminary study of emotional intelligence, empathy and exam performance in first year medical students. *Personality and Individual Differences, 39*(8), 1395-1405.

Balconi, M., Bortolotti, A., & Gonzaga, L. (2011). Emotional face recognition, EMG response, and medial prefrontal activity in empathic behaviour. *Neuroscience Research, 71*(3), 251-259.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173.

Baron-Cohen, S., Knickmeyer, R. C., & Belmonte, M. K. (2005). Sex differences in the brain: Implications for explaining autism. *Science, 310*(5749), 819-823.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders, 34*(2), 163-175.

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.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. *Journal of Autism and Developmental Disorders, 31*(1), 5-17.

Batson, C. D. (1991). *The altruism question*. Hillsdale, NJ: Lawrence Erlbaum associates.

Batson, C. D., Eklund, J. H., Chermok, V. L., Hoyt, J. L., & Ortiz, B. G. (2007). An additional antecedent of empathic concern: Valuing the welfare of the person in need. *Journal of Personality and Social Psychology, 93*(1), 65.

Batson, C. D., O'Quin, K., Fultz, J., Vanderplas, M., & Isen, A. M. (1983). Influence of self-reported distress and empathy on egoistic versus altruistic motivation to help. *Journal of Personality and Social Psychology, 45*(3), 706.

Bem, S. L. (1981). *Bem sex-role inventory: Professional manual* Palo Alto, California: Consulting Psychologists Press. Palo Alto, CA.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588.

Bernhardt, B. C., & Singer, T. (2012). The neural basis of empathy. *Annual Review of Neuroscience, 35*, 1-23.

Blair, R. J. R. (2005). Responding to the emotions of others: Dissociating forms of empathy through the study of typical and psychiatric populations. *Consciousness and Cognition, 14*(4), 698-718.

Blanchette, I., & Richards, A. (2004). Reasoning about emotional and neutral materials is logic affected by emotion? *Psychological Science*, *15*(11), 745-752.

Blanchette, I., & Leese, J. (2011). The effect of negative emotion on deductive reasoning: Examining the contribution of physiological arousal. *Experimental Psychology, 58*(3), 235.

Bogartz, R. S. (1965). The criterion method: Some analyses and remarks. *Psychological Bulletin, 64*(1), 1.

Bölte, S., Feineis-Matthews, S., & Poustka, F. (2008). Brief report: Emotional processing in high-functioning autism—physiological reactivity and affective report. *Journal of Autism and Developmental Disorders, 38*(4), 776-781.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111*(4), 1061.

Brackett, M. A., Mayer, J. D., & Warner, R. M. (2004). Emotional intelligence and its relation to everyday behaviour. *Personality and Individual Differences, 36*(6), 1387-1402.

Bryant, B. K. (1987). Mental health, temperament, family, and friends: Perspectives on children's empathy and social perspective taking. In N. Eisenberg, & J. Strayer (Eds.), *Empathy and its development* (pp. 245-270). Cambridge: Cambridge University Press.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81.

Crespi, B., & Badcock, C. (2008). Psychosis and autism as diametrical disorders of the social brain. *Behavioral and Brain Sciences, 31*(03), 241-261.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281.

D'Ambrosio, F., Olivier, M., Didon, D., & Besche, C. (2009). The basic empathy scale: A french validation of a measure of empathy in youth. *Personality and Individual Differences, 46*(2), 160-165.

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

de Sousa, A., McDonald, S., Rushby, J., Li, S., Dimoska, A., & James, C. (2011). Understanding deficits in empathy after traumatic brain injury: The role of affective responsivity. *Cortex, 47*(5), 526-535.

De Vignemont, F., & Singer, T. (2006). The empathic brain: How, when and why? *Trends in Cognitive Sciences, 10*(10), 435-441.

De Waal, F. B. (2008). Putting the altruism back into altruism: The evolution of empathy. *Annual Review of Psychology, 59*, 279-300.

Dymond, R. F. (1948). A preliminary investigation of the relation of insight and empathy. *Journal of Consulting Psychology, 12*(4), 228.

Dymond, R. F. (1949). A scale for the measurement of empathic ability. *Journal of Consulting Psychology, 13*(2), 127.

Dziobek, I., Rogers, K., Fleck, S., Bahnemann, M., Heekeren, H. R., Wolf, O. T., & Convit, A. (2008). Dissociation of cognitive and emotional empathy in adults with asperger syndrome using the multifaceted empathy test (MET). *Journal of Autism and Developmental Disorders, 38*(3), 464-473.

Edwards, L. H. (2013). A brief conceptual history of einfühlung: 18th-century germany to post-world war II US psychology. *History of Psychology, 16*(4), 269.

Eisenberg, N. (2002). Empathy-related emotional responses, altruism, and their socialization. In R. J. Davidson, & A. Harrington (Eds.), *Visions of compassion: Western scientists and tibethan buddhists examine human nature* (pp. 131-164). Oxford: Oxford University Press.

Eisenberg, N., & Strayer, J. (1987). Critical issues in the study of empathy. In N. Eisenberg, & J. Strayer (Eds.), *Empathy and its development* (pp. 3-16). Cambridge, UK: Cambridge University Press.

Eisenberg, N., & Lennon, R. (1983). Sex differences in empathy and related capacities. *Psychological Bulletin, 94*(1), 100.

Eisenberg, N., & Miller, P. A. (1987). The relation of empathy to prosocial and related behaviors. *Psychological Bulletin, 101*(1), 91.

Eliades, M., Mansell, W., Stewart, A. J., & Blanchette, I. (2012). An investigation of belief-bias and logicality in reasoning with emotional contents. *Thinking & Reasoning, 18*(4), 461-479.

Fan, Y., Chen, C., Chen, S., Decety, J., & Cheng, Y. (2013). Empathic arousal and social understanding in individuals with autism: Evidence from fMRI and ERP measurements. *Social Cognitive and Affective Neuroscience,* , nst101.

Gillespie, S. M., McCleery, J. P., & Oberman, L. M. (2014). Spontaneous versus deliberate vicarious representations: Different routes to empathy in psychopathy and autism. *Brain*. doi:10.1093/brain/awt364

Glöckner, A., & Witteman, C. (2010). Beyond dual-process models: A categorisation of processes underlying intuitive judgement and decision making. *Thinking & Reasoning, 16*(1), 1-25.

Grieve, R., & Mahar, D. (2013). Can social intelligence be measured? Psychometric properties of the Tromsø Social intelligence Scale. *The Irish Journal of Psychology, 34*(1), 1-12.

Han, S., Fan, Y., & Mao, L. (2008). Gender difference in empathy for pain: An electrophysiological investigation. *Brain Research, 1196*, 85-93.

Han, S., Fan, Y., Xu, X., Qin, J., Wu, B., Wang, X., . . . Mao, L. (2009). Empathic neural responses to others' pain are modulated by emotional contexts. *Human Brain Mapping, 30*(10), 3227-3237.

Hoekstra, R. A., Vinkhuyzen, A. A., Wheelwright, S., Bartels, M., Boomsma, D. I., Baron-Cohen, S., . . . van der Sluis, S. (2011). The construction and validation of an abridged version of the autism-spectrum quotient (AQ-short). *Journal of Autism and Developmental Disorders, 41*(5), 589-596.

Hoffman, M. L. (2000). *Empathy and moral development: Implications for caring and justice*. Cambridge: Cambridge University Press.

Hoffman, M. L. (1977). Sex differences in empathy and related behaviors. *Psychological Bulletin, 84*(4), 712.

Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology, 33*(3), 307.

Holt, C. L., & Ellis, J. B. (1998). Assessing the current validity of the Bem Sex-Role inventory. *Sex Roles, 39*(11-12), 929-941.

Howe, D. (2013). *Empathy: What it is and why it matters*. Hampshire: Palgrave Macmillan.

Ibanez, A., Huepe, D., Gempp, R., Gutiérrez, V., Rivera-Rei, A., & Toledo, M. I. (2013). Empathy, sex and fluid intelligence as predictors of theory of mind. *Personality and Individual Differences, 54*(5), 616-621.

Ickes, W., Gesn, P. R., & Graham, T. (2000). Gender differences in empathic accuracy: Differential ability or differential motivation? *Personal Relationships, 7*(1), 95-109.

Jabbi, M., Swart, M., & Keysers, C. (2007). Empathy for positive and negative emotions in the gustatory cortex. *Neuroimage, 34*(4), 1744-1753.

Jolliffe, D., & Farrington, D. P. (2004). Empathy and offending: A systematic review and meta-analysis. *Aggression and Violent Behavior, 9*(5), 441-476.

Jolliffe, D., & Farrington, D. P. (2006). Development and validation of the basic empathy scale. *Journal of Adolescence, 29*(4), 589-611.

Karniol, R., Gabay, R., Ochion, Y., & Harari, Y. (1998). Is gender or gender-role orientation a better predictor of empathy in adolescence? *Sex Roles, 39*(1-2), 45-59.

Kaukiainen, A., Björkqvist, K., Lagerspetz, K., Österman, K., Salmivalli, C., Rothberg, S., & Ahlbom, A. (1999). The relationships between social intelligence, empathy, and three types of aggression. *Aggressive Behavior, 25*(2), 81-89.

Kirchner, J. C., Hatri, A., Heekeren, H. R., & Dziobek, I. (2011). Autistic symptomatology, face processing abilities, and eye fixation patterns. *Journal of Autism and Developmental Disorders, 41*(2), 158-167.

Klein, K. J., & Hodges, S. D. (2001). Gender differences, motivation, and empathic accuracy: When it pays to understand. *Personality and Social Psychology Bulletin, 27*(6), 720-730.

Kohut, H. (1984). The role of empathy in psychoanalytic cure. *How does analysis cure?* (pp. 172-191). London: University of Chicago Press.

Lieberman, M. D. (2007). Social cognitive neuroscience: A review of core processes. *Annu.Rev.Psychol., 58*, 259-289.

Lönnqvist, J., Paunonen, S., Tuulio-Henriksson, A., Lönnqvist, J., & Verkasalo, M. (2007). Substance and style in socially desirable responding. *Journal of Personality, 75*(2), 291-322.

Marci, C. D., Ham, J., Moran, E., & Orr, S. P. (2007). Physiologic correlates of perceived therapist empathy and social-emotional process during psychotherapy. *The Journal of Nervous and Mental Disease, 195*(2), 103-111.

Mathersul, D., McDonald, S., & Rushby, J. A. (2013). Understanding advanced theory of mind and empathy in high-functioning adults with autism spectrum disorder. *Journal of Clinical and Experimental Neuropsychology, 35*(6), 655-668.

Mathur, V. A., Harada, T., Lipke, T., & Chiao, J. Y. (2010). Neural basis of extraordinary empathy and altruistic motivation. *Neuroimage, 51*(4), 1468-1475.

McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51*(6), 882.

Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy. *Journal of Personality, 40*(4), 525-543.

Mehrabian, A., Young, A. L., & Sato, S. (1988). Emotional empathy and associated individual differences. *Current Psychology, 7*(3), 221-240.

Mullins-Nelson, J. L., Salekin, R. T., & Leistico, A. R. (2006). Psychopathy, empathy, and perspective-taking ability in a community sample: Implications for the successful psychopathy concept. *International Journal of Forensic Mental Health, 5*(2), 133-149.

Muncer, S. J., & Ling, J. (2006). Psychometric analysis of the empathy quotient (EQ) scale. *Personality and Individual Differences, 40*(6), 1111-1119.

Norris, P., & Epstein, S. (2011). An experiential thinking style: Its facets and relations with objective and subjective criterion measures. *Journal of Personality, 79*(5), 1043-1080.

Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology, 76*(6), 972.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers, 36*(4), 717-731.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(04), 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.

Roberts, W., & Strayer, J. (1996). Empathy, emotional expressiveness, and prosocial behavior. *Child Development, 67*(2), 449-470.

Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology, 21*(2), 95.

Saarela, M. V., Hlushchuk, Y., Williams, Amanda C de C, Schürmann, M., Kalso, E., & Hari, R. (2007). The compassionate brain: Humans detect intensity of pain from another's face. *Cerebral Cortex, 17*(1), 230-237.

Schulte-Rüther, M., Markowitsch, H. J., Shah, N. J., Fink, G. R., & Piefke, M. (2008). Gender differences in brain networks supporting empathy. *Neuroimage, 42*(1), 393-403.

Schutte, N. S., Malouff, J. M., Bobik, C., Coston, T. D., Greeson, C., Jedlicka, C., . . . Wendorf, G. (2001). Emotional intelligence and interpersonal relations. *The Journal of Social Psychology, 141*(4), 523-536.

Schwartz, S. H., Verkasalo, M., Antonovsky, A., & Sagiv, L. (1997). Value priorities and social desirability: Much substance, some style. *British Journal of Social Psychology, 36*(1), 3-18.

Settoon, R. P., & Mossholder, K. W. (2002). Relationship quality and relationship context as antecedents of person-and task-focused interpersonal citizenship behavior. *Journal of Applied Psychology, 87*(2), 255.

Shamay-Tsoory, S. G., Aharon-Peretz, J., & Perry, D. (2009). Two systems for empathy: A double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain, 132*(3), 617-627.

Silvera, D., Martinussen, M., & Dahl, T. I. (2001). The Tromsø social intelligence scale, a self-report measure of social intelligence. *Scandinavian Journal of Psychology, 42*(4), 313-319.

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature, 439*(7075), 466-469.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science, 303*(5661), 1157-1162.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology, 13*(1982), 290-312.

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology, 89*(2), 342.

Stotland, E. (1969). Exploratory investigations of empathy. *Advances in Experimental Social Psychology, 4*, 271-314.

Wheelwright, S., Baron-Cohen, S., Goldenfeld, N., Delaney, J., Fine, D., Smith, R., . . . Wakabayashi, A. (2006). Predicting autism spectrum quotient (AQ) from the systemizing quotient-revised (SQ-R) and empathy quotient (EQ). *Brain Research, 1079*(1), 47-56.

Appendix

Subject matter and source of pictures used in final version of PET:

PICTURE 1

Ten-year old girl crying over her dead sister

http://commons.wikimedia.org/wiki/File:Julien_Bryan_-_Life_-_50893.jpg

This picture is in the public domain, so it is possible to present it here as a sample picture. (Other pictures are free-to-use, but place restrictions in the licensing of the final product: thus, only their origin is reported here.) This picture was taken by Julien Bryan in 1939 during the Siege of Warsaw.

PICTURE 2

Young boy crying

http://commons.wikimedia.org/wiki/File:2012_East_Azerbaijan_earthquakes._by_Mardetanha_1527.JPG

PICTURE 3

About a year old boy, with skin suggesting a deficiency disease, crying

http://commons.wikimedia.org/wiki/File:Rahima_Banu.jpg

PICTURE 4

Baby experiencing physical distress

http://commons.wikimedia.org/wiki/File:Tratamiento_epidermolisis_bullosa.jpg

PICTURE 5

Man, holding either wounded or dead child, crying

http://commons.wikimedia.org/wiki/File:V_rekonstrukcja_Bitwy_o_M%C5%82aw%C4%99,_miasto_0992.jpg Adam Kliczek / Wikipedia, licence: CC-BY-SA-3.0

PICTURE 6

Young girl with scars and other deformities on her face on a hospital bed

http://commons.wikimedia.org/wiki/File:Bala_Baluk_massacre_by_US_troops.jpg

PICTURE 7

One man with blood on his face, and several other men helping him

http://commons.wikimedia.org/wiki/File:Wounded_Minsk_blast_2.jpg