Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Brett Adam Murphy Date

The Varieties of Self-Reported Empathic Tendencies

By

Brett A. Murphy
Doctor of Philosophy

Psychology

_____

Scott O. Lilienfeld, Ph.D.
Advisor

_____

Philippe Rochat, Ph.D.
Committee Member

_____

Elaine F. Walker, Ph.D.
Committee Member

_____

Yuk Fai Cheong, Ph.D.
Committee Member

_____

Jessica Barber, Ph.D.
Committee Member

Accepted:

_____

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____

Date

The Varieties of Self-Reported Empathic Tendencies

By

Brett A. Murphy
J.D., Harvard Law School, 2011
B.A., Rice University, 2004

Advisor: Scott O. Lilienfeld, Ph.D.

An abstract of

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Psychology
2019

Abstract

The Varieties of Self-Reported Empathic Tendencies
By Brett A. Murphy

Emotionally empathic tendencies can manifest in a wide range of forms. Existing measures of such tendencies often differ widely in item content, but no individual questionnaire distinguishably measures more than one or two narrow aspects of the broader construct. This dissertation project aimed to develop new multi-dimensional emotional empathy scales that effectively measure a broader range of empathy domains.

In Study 1, I administered a comprehensive collection of more than 200 potential empathy items to an online community sample (n = 712). Using a wide range of techniques, including bifactor modeling and item response theory (IRT) methods, I constructed and preliminarily evaluated nine different scales of empathic tendencies. Based upon potential limitations revealed in analyses, I subsequently created additional items to potentially buttress the new scales.

In Study 2, I administered these new scales to another online community sample (n = 335), while also assessing a range of external variables related to antagonism, self-compassion and emotional distress, and schizotypal personality. I employed a range of analytical techniques to validate the new scales and examine their fine-grained nomological networks.

This research is situated within an extensive discussion of the history of the empathy construct and current controversies regarding its definitional core and boundaries. Among the many theoretical implications generated by these studies, three are of particular note. First, empathic contagion is not a unidimensional construct, as contagion for the positive emotions of others is strongly distinguishable from contagion for negative emotions. Second, empathic approach-avoidance tendencies, which have been neglected in many past studies, appear to be a critical research domain. Third, proto-Rogerian aspects of empathy, particularly tendencies to be supportive confidante listeners for others, should be a renewed area of interest. The new questionnaire scales generated in this project may serve as valuable tools in future research.

The Varieties of Self-Reported Empathic Tendencies

By

Brett A. Murphy
J.D., Harvard Law School, 2011
B.A., Rice University, 2004

Advisor: Scott O. Lilienfeld, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Psychology
2019

Table of Contents

List of Tables

List of Figures

The Varieties of Self-Reported Empathic Tendencies

> The blindness in human beings…is the blindness with which we all are afflicted in regard to the feelings of creatures and people different than ourselves…Hence the stupidity and injustice of our opinions, so far as they deal with the significance of alien lives (James, 1899/2000, p. 267)

A Ph.D. candidate is defending his dissertation, the culmination of years of thinking, working, and hoping, when he begins to feel an anxiety attack coming over him. Though he soldiers on with his speaking, trying to remain outwardly composed, subtle signs of his distress are present: pools of sweat form under his arms and beads of sweat drip down his brow; his hands are slightly shaking; disfluencies leak into his speaking; flashes of anxiety peek out from his eyes.

One of the candidate's lab mates detects that the candidate is starting to panic and feels a slight but only momentary feeling of anxiety in herself as well. This lab mate immediately starts feeling warm caring for the candidate, quickly appraising the situation and assessing how she might be able to help. She focuses on sending him calm, supportive looks, hoping to give the speaker a comforting face to rely upon, reassuring eyes with which to connect. She imagines that, from what she knows of the candidate, this is what might be most helpful to him.

Another of the candidate's lab mates feels the speaker's anxiety and "catches" it himself. As the candidate continues to struggle, this lab mate also starts to sweat. He turns his eyes away from the speaker and tries to redirect his attention elsewhere in the room, but still feels his own anxiety building. At a pause in the presentation, this lab mate quietly leaves the room, with worried thoughts of his own upcoming research presentation on his mind.

Another graduate student audience member clearly perceives the candidate's anxiety, and inwardly smiles in schadenfreude, "Oh, how the mighty have fallen! Not so cocky now, eh!"

Another audience member does not notice the candidate's anxiety, but cares for him and is proud of him, as he smiles at the candidate and appreciates the work he is presenting. Another audience member clearly notices the candidate's anxiety but has no emotional response to it of any kind, his state of boredom remaining uninterrupted.

The term "empathy," first coined in 1909 by Titchener as a translation of a German term ("einfühlung") that had been popularized a few years earlier, has been used by different scholars to capture at least some aspect of all of the different audience member responses just described. This dissertation aims to (a) help unravel a few sources of confusion regarding the construct and (b) provide new self-report measurement scales to modestly assist future research endeavours.

Introductory Sections

At present, empathy is a highly discussed topic not only in academic research (with many hundreds of books and articles published each year referencing it), but also in political discussions, relationship advice columns, religious sermons, classroom education, and other arenas of general public discourse. According to Google's Ngram feature, which analyses the frequency with which a word or phrase appears in published books (normalized for the number of sources per year), the usage of the term began steadily increasing in the 1940s and has continued to steadily increase in use up to the present time (approximately 20 times as frequently used in 2008 as in 1940). Within clinical psychology, empathy has long been a pivotal concept in psychotherapy, at least from the time of Carl Rogers (e.g., Rogers, 1949, 1959). Moreover, empathy deficits have been extensively investigated in relationship to a wide range of psychopathological constructs, such as psychopathy (e.g., Brook & Kosson, 2013; White, 2014), borderline personality disorder (Dinsdale & Crespi, 2013; Ripoll, Snyder, Steele, & Siever, 2013), autism spectrum disorder (Bird & Viding, 2014; Montgomery et al., 2016), narcissism

(Hepper et al., 2014; Watson, Grisham, Trotter, & Biderman, 1984), and schizophrenia (e.g., Derntl et al., 2009; Shamay-Tsoory, Shur, Harari, & Levkovitz, 2007).

Nonetheless, "empathy" remains definitionally controversial, with little consensus as to the meaning of the term (Batson, 2009; Neumann et al, 2015). Some have observed that there are as many definitions of "empathy" as there are empathy researchers (de Vignemont & Singer, 2006). Furthermore, in attempting to distinguish empathy from other terms, particularly "sympathy," researchers struggle to define those other terms as well. The situation is such that, as Debes (2015, p. 287) argued, "At present…'empathy' and 'sympathy' are eclectic concepts, which only the most dogmatic or ignorant pretend to separate objectively and without stipulation." Others have argued that, despite the broad consensus that empathy is critical in social functioning, the lack of definitional agreement and clarity means that the catch-all term itself should be abandoned in favor of a collection of more precise, narrow terms (e.g., Decety & Cowell, 2014).

Furthermore, the different constructs discussed under the umbrella of "empathy" are often substantively indistinguishable from other constructs discussed in psychology, such as emotion recognition and compassion. One major drawback of this is that researchers focused on one specific conception of empathy may often fail to notice and incorporate relevant existing literatures into their work.[1] For example, Murphy and Lilienfeld (2019) speculated that many empathy researchers have failed to notice prior work indicating that individuals are poor judges of their own "cognitive empathy" abilities, possibly because warnings have used terms such as "interpersonal sensitivity" and "mindreading" rather than "empathy."

[1] Cf. Peirce's (1878) "How to Make Our Ideas Clear," which laments the intellectual confusion caused by (a) using different terms for the same construct and (b) definitionally splitting constructs based on trivial differences.

Much of the disagreement about the core and boundaries of "empathy" may subtly emerge from the different questions researchers attempt to use the concept to explore and illuminate. As Wittgenstein (1953, §43) argued, terms can usually be understood as "tools" and the *meaning* of a term is typically substantially revealed in how it is *used* to serve a particular purpose or collection of purposes. Batson (2009), summarizing the field, argued that there are two main questions motivating empathy research: (1) how do human beings come to know/infer what others are thinking or feeling? (and *why* are they are thinking or feeling that way?); and (2) what motivates human beings to emotionally care about others, and subsequently behave in a prosocial manner toward others (especially in regard to *altruistic* motivation, rather than social convention, desire for social reward, etc.)? The first question we might refer to as the "mindreading" issue and the second as the "ethical/moral" issue. Many researchers have aimed to use the empathy construct to deal with both issues simultaneously (e.g., Hogan, 1969; Rogers, 1975; Baron-Cohen & Wheelwright, 2004), whereas others have focused on only one of these two broad functions.

Although they are the two most common motivating questions for discussing empathy, there are other specific motivating issues to be discerned in the intellectual history of the construct. For instance, the term "empathy" is a translation of the German term "einfühlung," which means "feeling into" and was most prominently discussed in the domain of aesthetics: what is the process of "feeling into" art, nature, humor, music, etc., that generates our perceptions of beauty, delight, transcendence, and so on? Is this the same process that causes us to see "inner" beauty in the personalities of others?

In a very different vein, the construct of empathy has been rigorously mined by clinicians in search of answers to the question: how can we most effectively heal or support people in the

face of their emotional struggles (e.g., Rogers, 1975)? How can a clinician feel into a client's inner life in a way that is validating to the client, builds relationship, and generates insight?

Furthermore, other scholars have made use of the empathy construct (or similar constructs) in attempting to answer political questions, such as: how can we promote tolerance and cooperation between different communities? (e.g., Berlin, Debes, 2015; James, 1899/2000; Obama, 2006). How can we feel into the perspectives of those with whom we disagree or with whom we have limited experiential overlap, humanizing them in our eyes and paving the way for peace and understanding?

Of course, these various different motivating questions are not necessarily separate from one another. Nonetheless, the particular ways in which scholars conceptualize empathy tend to be related to the particular questions of primary interest to them. In some cases, however, it is possible that researchers may tend to define empathy in ways that most adapt it to the measurement tools in which they specialize, such as neuroimaging methods, or align it more with the latest hyped concepts, such as mirror neurons, while inadvertently (and perhaps unknowingly) veering away from the underlying "bigger" questions posed.[2]

In defining and conceptualizing empathy, the closest thing to a vernacular is to separate the construct into two broad categories (e.g., Jolliffe & Farrington, 2006; Vachon & Lynam, 2016): "cognitive empathy," the capacity (or capacities) to accurately recognize, infer, or more broadly *understand* the thoughts and feelings of others; and "affective empathy," the emotional (and, for some researchers, behavioral) response to the thoughts and feelings of others. Nonetheless, this common distinction is not embraced by all researchers, with some arguing that empathy should be understood in a unified fashion (e.g., Preston & de Waal, 2002, 2017), some

[2] To amend the words of Tellegen and Waller (2008, p. 267), this can manifest as a "[measurement] drift with conceptual lag."

arguing that either cognitive empathy or affective empathy is not (at least by itself) empathy at all (e.g., Feshbach, 1975), and others arguing for further parsing (e.g., the tripartite framework of Decety & Jackson, 2004).

In terms of cognitive empathy, some researchers focus on processes that do not inherently require any significant "feeling" experience in the observer, such as simple emotion recognition, basic theory of mind, and social inferential reasoning, yielding "abstract, propositional knowledge about the other's mental state" (Preckel, Kanske, & Singer, 2018, p. 1). Others, however, describe a more imaginatively involved form of coming to "know," one in which the empathizer *experiences* the subjective experiences of another, imaginatively sensing the qualia (cf. Nagel, 1974) of the other person, to some limited and highly fallible degree, in the "mind's eye" (e.g., Dekeyser, Elliott, & Leijssen, 2009). In this kind of process, the inner life of the other person begins to feel present and "real" rather than distant and abstract. "This is precisely what empathic knowledge amounts to: experiential, or at least quasi-experiential, knowledge of *what it is like* to occupy another's perspective" (Steinberg, 2014, p. 50).

This is the difference between knowing a person as one knows a car engine or a statistical software package, as an object to be predicted and maneuvered, versus "knowing" a person's subjective experience more "feelingly,"[3] as a consciously experiencing personality rather than just an object in a world of things. Put differently, if we only take cognitive empathy to mean recognition and prediction, this is something that can be achieved by a non-conscious machine with suitable emotion recognition, theory of mind, and inference-refining capabilities (cf. the "theory theory" of our understanding of others; e.g., Carey, 1985; Gopnik and Wellman, 1992, 2012). If, however, we take empathy to require an imaginative experiencing (albeit limited) of

[3] One is reminded of the moment (Act 4, Scene 6) in Shakespeare's *King Lear*, when King Lear asks the blind Gloucester how he sees the world. Gloucester replies, "I see it feelingly."

the consciousness of another, then this would seem to be the bailiwick of only similarly consciously experiencing entities (cf. Philip K. Dick's, 1968, *Do Androids Dream of Electric Sheep?*, which continually references "empathy" as distinguishing humans from androids).

As perhaps another avenue by which to understand this immersive, experiential notion of empathy, consider the virtual reality films that the United Nations Virtual Reality Series has been creating since 2015, with the goal of "pushing the bounds of empathy" (UNVR website). "Clouds Over Sidra," for example, is a brief documentary film in which a young girl guides the viewer around a Syrian refugee camp, narrating about daily life. There is minimal distress shown; the viewer mostly sees people playing, laughing, being together, and going about their lives. There is only minimal sentimentality in the filmmaking. The purpose seems to be to briefly *transport* the viewer into the camp and *humanize* the people who live there. The virtual reality aspect, which can be experienced by any person with a suitable VR headset, is intended to heighten the empathic immersion of the experience (for a proposed taxonomy of the "empathy" aspects in VR, see Fisher, 2017). Limited evidence presented by Schutte and Stilinovic (2017) indicates that the virtual reality form of the film, compared with the normal television screen form of the film, significantly increases viewer engagement, caring for Sidra, and feelings of taking the perspective of Sidra. The enhanced "understanding" hopefully provided by this technology is quasi-experiential, not necessarily about increased ability to predict and inferentially navigate.

Greatly complicating this construct, many scholars, in defining empathy as approximately "the process whereby one person 'feels her/himself into' the consciousness of another person"

(Wispé, 1987, p. 34), have explicitly defined it as distinguished from projection[4], attributing one's own thoughts, feelings, perspectives to the other person (e.g., Dymond, 1950; Rogers, 1957; Stein, 1917). In einfühlung/empathy, one engages in "the imaginative transposing of oneself into the thinking, feeling and acting of another and so *structuring the world as he does*" (Dymond, 1949, p. 343, emphasis added). So, for instance, when putting oneself in Sidra's shoes in the Syrian refugee camp, full einfühlung/empathy would mean imagining how Sidra herself might actually sense and feel her experiences there, not how, say, I , as an American adult male, might feel if I were living there. In imagining how Sidra might experience her world, einfühlung/empathy requires that I limit or modulate the degree to which I, consciously or unconsciously, project my own thoughts and feelings back onto her.

If we conclude, with Wispé (1987, p. 34) and others, that this imaginatively experiential process is the "hard core" of the empathy construct, then we are faced with an uncomfortable truth: the thousands of past studies of empathy, as a general rule, have not *directly* measured the core of the construct (cf. Skinner, 1990). Indeed, it is difficult to imagine how this core of the construct, founded in experiential phenomena in subjective consciousness, could be directly measured, whether via self-report or behavioral observation, with specificity (e.g., distinguishing it from projection, accurate prediction by non-empathic means, non-empathic emotional arousal). Instead, past studies have measured *theorized* close correlates of this kind of experience, such as emotional caring and emotional contagion, including these theorized close correlates as part of a broader empathy construct. We have studied the observable or reportable theorized fringes of the construct, but not the core of the construct itself. Some of the confusion and controversy

[4] In using this term, here and elsewhere, I am not referencing the psychoanalytic concept of "projection" as a defense mechanism, in which individuals deny undesired feelings or characteristics in themselves and project them falsely onto others.

surrounding empathy may be due to a failure to acknowledge and appreciate this inherent conundrum.

In addition to the "understanding" elements of empathy, the broad construct also encompasses emotional elements, historically usually tied to prosociality and caring for the feelings of others. Many researchers (e.g., Bird & Viding, 2014; Bloom, 2017b; Coll et al., 2017; Jolliffe & Farrington, 2006; Preckel, Kanske, & Singer, 2018), however, currently define affective empathy narrowly, as restricted only to feeling the *same* emotion that you perceive another is feeling[5]. In doing so, they exclude compassion and caring from the empathy construct, as these emotions do not isomorphically mirror those of the person who is being perceived.

Oddly, this appears to be emerging as the "standard definition" (e.g., Coll et al., 2017; but see Hall & Schwartz, 2019), especially for cognitive neuroscientists. Others with much broader definitions, however, argue that the emotional response need not be the same as that experienced by the other, but simply "appropriate" (e.g., Baron-Cohen & Wheelwright, 2004; also see Zaki, 2017), or, more specifically, rooted in "concern, kindness, and care to others' suffering" (Stern & Cassidy, 2018, p. 1).

For example, if a father sees that his child is throwing a temper tantrum, some researchers would call it "empathy" only if he also becomes similarly angry, whereas others would call it "empathy" if he becomes calm and compassionately comforting towards his child. In our earlier example of a Ph.D. candidate having an anxiety attack while speaking, one camp would point to the lab mate who anxiously left the room as empathizing the most, whereas the other camp would say that the lab mate who focused on helping ease the candidate's anxiety was the more proper example of empathy.

[5] e.g., "For empathy to have occurred, the empathiser must be in the same affective state as the object of empathy." (Bird & Viding, 2014, p. 521)

The radical diversity in conceptualizations of empathy makes it quite difficult to compare studies and research conclusions (Cuff, Brown, Taylor, & Howat, 2016), perhaps especially for those who do not have expertise in the domain. In drawing broad conclusions about "empathy," such as in meta-analyses (e.g., Vachon, Lynam, & Johnson, 2014), the failure to carefully parse apart the different conceptualizations may be an especially concerning issue. This difficulty is amplified by the fact that the measures researchers use frequently do not facially match with the definitions they present; as assessed by Hall and Schwartz (2019), only about half of empathy studies use measures that align well with the definitions of the constructs they offer. Moreover, some empathy measures do not appear to produce scores that validly measure what the scale is labeled as assessing (e.g., Murphy, 2018, 2019; Murphy et al., 2018a; Murphy, Costello, & Lilienfeld, 2018b; Murphy & Lilienfeld, 2019). There are ample reasons to worry that at least some of the empathy research literature is a house built on sand.

The studies presented in this dissertation have a relatively limited aim and can only hope to incrementally improve the state of empathy research. Given that the overwhelming majority of empathy research has been conducted using self-report questionnaires of trait-level constructs (Hall & Schwartz, 2019), a good starting point is to begin to clear up the confusion in that measurement domain, with the goal of producing conceptually clarified self-report scales possessing improved psychometric properties. Pursuant to that goal, I compiled empathy questionnaire items from a substantial range of measures, created many new items to measure specific potential domains, and conducted an intensive process of distilling this large pool of items into potentially sound self-report scales (at least in comparison with existing scales).

Furthermore, as advocated by Hall and Schwartz (2019), there is a pressing need for systematic, comprehensive investigation as to how different "empathies" relate to one another:

"Fully understanding the structure of 'empathy' as it is operationally defined and measured would be a great step forward" (p. 235). At least in the domain of self-report measures, the dissertation studies I present are a small step towards that goal.

One major limitation of this dissertation research is worth noting at the outset. Although the accuracy or skill of an individual's "cognitive empathy" is, under most conceptualizations of the construct, a critical aspect of (or at least a precursor to) empathy, this dissertation research does not attempt to address it as a self-report measurement domain. This is because, as discussed further below, the meta-analytic findings of Murphy and Lilienfeld (2019) indicate that, at least in terms of the available evidence, individuals' ratings of their own global trait abilities to accurately recognize and/or infer the thoughts and feelings of others appear to be negligibly related to performance in behavioral cognitive empathy tasks. As a result, the two studies presented intentionally excluded such self-report items of cognitive empathy ability, instead focusing only on (1) empathic caring, (2) isomorphic emotion matching, (3) perspective-taking *motivation*, (4) empathic absorption/immersion, (5) empathy avoidance-approach tendencies, and (6) empathic distress. Although aspects of cognitive empathy *ability* are discussed briefly at various points, this construct is not a major focus of this dissertation.

The remainder of the introductory sections of this dissertation, prior to the presentation of the two studies, are intended to elucidate my theoretical approach. First, I present a selective brief review of the intellectual history of the empathy construct. Second, I discuss current issues in the conceptualization and self-report measurement of empathy. Third, I introduce two substantial methodological aspects of the present studies: my use of bifactor modeling and my efforts to address M-Turk sample quality.

Highlights of the Intellectual History of the Empathy Construct

The dizzying variety of conceptualizations of empathy is not new, nor is the occasional despair as to whether the term even still has value. Only a few decades after it was first introduced in English, the psychoanalyst Theodor Reik (1936, p. 192) noted, "Indeed, this expression sounds so full of meaning that we willingly overlook its ambiguity…really the conception of empathy has become so rich in meanings that it is beginning to mean nothing at all." At the same time, many of the current difficulties with conceptualizing empathy, and drawing the borders within and around the construct, are essentially echoes of similar questions and concerns raised decades or centuries ago (and sometimes handled with more philosophical sophistication than currently).

Rather than sketch a full history of the intellectual history of the empathy construct, in the rest of this section, I have selected particular highlights and presented them only briefly. In general, the focus here is on presenting a diversity of conceptualizations and noting particular concerns related to different conceptualizations, many of which continue to be relevant today. The first subsection discusses "sympathy" and "compassion" in the works of a handful of 18th and 19th century philosophers. The second subsection discusses the emergence of the concept of "einfühlung" and its translation and modification as "empathy" in the 20th century.

Before einfühlung: Smith, Schopenhauer, and Nietzsche. Although overshadowed by his monumental *The Wealth of Nations* (1776), the second most-cited book in all of the social sciences (Green, 2016), Adam Smith's work on sympathy is a major event in the intellectual history of moral philosophy. Of all the historical sources related to the empathy construct, the opening paragraphs of Smith's (1759/2009, ps. 13-14) *The Theory of Moral Sentiments* are certainly among the most widely quoted or referenced:

> As we have no immediate experience of what other men feel, we can form no idea of the manner in which they are affected, but by conceiving what we ourselves should feel in

> the like situation…By the imagination we place ourselves in his situation, we conceive ourselves enduring all the same torments, we enter as it were into his body, and become in some measure the same person with him, and thence form some idea of his sensations, and even feel something which, though weaker in degree, is not altogether unlike them…That this is the source of our fellow-feeling for the misery of others, that it is by changing places in fancy with the sufferer, that we come either to conceive or to be affected by what he feels, may be demonstrated by many obvious observations, if it should not be sufficiency evident of itself.

In Smith's view, however, this kind of act is certainly not limited to perceptions of grief or suffering; rather, our fellow-feelings in situations of joy, amazement, and other emotions emerge in the same manner. Noting that "sympathy" has long been discussed in regard to the suffering of others, Smith argues that it "may now, however, without much impropriety, be made use of to denote our fellow-feeling with any passion whatsoever" (p. 15). Unlike many of his predecessors, and many current scholars (see Tell & Pfister, 2016), Smith devoted substantial attention to sympathy with the positive emotions of others. Interestingly, Smith's depiction of "changing places in fancy" is similar to some current "simulation theories" as to how we come to understand the minds of others (e.g., Gallese, 2005; Gallese & Goldman, 1998; also see Ravenscroft, 1998).

Much less frequently quoted or discussed (but see Wondra & Wellsworth, 2015) is that Smith (1759/2009, p. 16, emphasis added) considered this sympathetic psychological process to be strongly tied to how one appraises the situation of the other person, not just to immediate mimicry or contagion:

> If the very appearances of grief and joy inspire us with some degree of the like emotions, it is because they suggest to us the general idea of some good or bad fortune that has befallen the person in whom we observe them…The first question which we ask is, What has befallen you? Til this be answered…yet our fellow-feeling is not very considerable. *Sympathy, therefore, does not arise so much from the view of the passion, as from that of the situation which excites it*.

In other words, fellow-feeling is not primarily fueled by mere emotional contagion or imitation. The mere observation of someone in a particular state, such as joy or fear, only weakly generates emotions in us; it is the appraisal of the person's situation that more powerfully determines our emotional state. For instance, if I look out my bus window and see a woman running with a fearful face, I will tend to experience a flicker of fear. But seeing what she is running from will much more substantially affect my emotions: a rabid Doberman? Department store "loss prevention" agents? A child in a clown costume?

One consequence of this, for Smith (1759), is that in "changing places in fancy" with others, we often feel a different emotional state than that experienced by the person observed. It is possible, for instance, for us to feel worried concern for an individual even if he/she is not worried; for an example from Smith, we could feel concern for an individual who has just committed a grave breach of social manners, even if he himself is blind to the situation.

Of note, Smith points out that some emotions witnessed in others will tend to, even if only slightly, spontaneously elicit similar emotions in us, such as happiness and sadness. Others, however, such as anger, will *only* tend to elicit similar emotions in us, of any degree, if we subsequently appraise the situation to be worthy of anger. In other words, spontaneous "emotional contagion" tendencies will depend upon the type of emotion we perceive in others.

Complicating matters, Smith argues that sometimes our sympathy consists of putting aside our own preferences, beliefs, and values, and imagining what we would feel if we were fully in the place of another with his/her very different such preferences, beliefs, and values. In such cases, we more *fully* take the perspective of the other person. In other cases, though, we sympathize by holding onto our own perspectives. Smith gives the example, for instance, of sympathizing with a person who has lost his sanity and does not perceive the desperation of his

situation. Both kinds of de-centering, whether projection-based or full perspective-taking, are treated as "sympathy" in Smith's account.

For Smith, this "changing places in fancy" was not only the basis of our concern for the well-being of others; it is the basis for all moral judgment. In addition to engaging sympathetic concern for others, this process causes us to approve or disapprove of actions and motives, in both others and in ourselves (and, therefore, be able to judge them as moral or immoral). Absent this capacity, society would fall apart, and there would be no judgments of morality. After presenting his initial theory of the process of sympathy, Smith devoted hundreds of dense pages linking this "changing places in fancy" to the full realm of moral domains present in human life.

A number of other 18th century philosophers discussed sympathy, perhaps most famously Jean Jacques Rousseau and David Hume. Rousseau's comments were far less extensive than Smith's but align roughly with them in a few ways, such as describing sympathy as requiring an act of imagination that takes one outside oneself, identifying with another who is suffering.[6] Hume also wrote extensively on sympathy and, for the most part, his views largely align with those of Smith (Sayre-McCord, 2015). Nonetheless, there are meaningful differences between the two, though most are in reference to moral judgment and approbation, and thus beyond the scope of this dissertation.

Perhaps most importantly for our discussion, compared with Smith, Hume gives far less power to appraisal in generating sympathy, and more to the immediate apprehension of another

[6] E.g., "So pity is born, the first relative sentiment which touches the human heart according to the order of nature. To become sensitive and pitiful the child must know that he has fellow-creatures who suffer as he has suffered, who feel the pains he has felt, and others which he can form some idea of, being capable of feeling them himself. Indeed, how can we let ourselves be stirred by pity unless we go beyond ourselves, and identify ourselves with the suffering animal, by leaving, so to speak, our own nature and taking his. We only suffer so far as we suppose he suffers; the suffering is not ours but his. *So no one becomes sensitive till his imagination is aroused and begins to carry him outside himself*" (Rousseau, 1762/1921, p. 184, emphasis added).

person's emotional state, the idea conveyed by the other which spontaneously generates similar feelings in us (Sayre-McCord, 2015). As a result, Hume's version of sympathy has been characterized as mere "emotional contagion" (Darwall, 1998); this, however, goes a bit too far, given that Hume does have a role for some degree of inference and appraisal (Ilyes, 2017). Nonetheless, Smith's view of sympathy can be viewed as a more cognitively, imaginatively involved process than in Hume (Ilyes, 2017).

In the 19th century, of the enduringly famous philosophers, Arthur Schopenhauer is perhaps most associated with the general domain of sympathy. With Smith, and against the rationalism of Immanuel Kant, Schopenhauer (1840, 1859) also argued that moral behavior rarely emerged from detached rational thought, but, instead, emerged from and was founded upon moral emotion. Schopenhauer's (1840/1903) *On the Basis of Morality* [all subsequent references and quotes are for this particular work] argued that all human conduct could be attributed to three different motivations: Egoism, self-seeking or self-interest; Malice, cruelty or ill will towards others; and Mitleid ("suffering with", translated into English as "compassion"), which he defined (p. 170, emphasis added) as "*direct participation*, independent of all ulterior considerations, in the sufferings of another, leading to sympathetic assistance in the effort to prevent or remove them." Proceeding from the perspective that the moral worth of behavior is judged by the motives generating it, Schopenhauer argued that mitleid is the *sole* basis of morality, and its highest manifestation is in "loving kindness," whereas Egoism and Malice are anti-moral motivations. For Schopenhauer, mitleid was responsible both for restraining oneself from harming another as well as with taking positive action to help another.

As with other scholars before him, though, Schopenhauer held that the overwhelming majority of instances of just or charitable conduct are motivated more by egoism than by mitleid.

Of the average person, Schopenhauer pessimistically admitted that his "conscience" consists: "probably of about one-fifth, fear of men; one-fifth, superstition; one-fifth, prejudice; one-fifth, vanity; and one-fifth, habit" (1840/1903, p. 144). As a result, most behavior that appears caring towards others may be caused by fear of punishment, desire for social reward, dictates of religious convention, etc.

In other words, most occasions of what might appear to be mitleid are actually camouflaged instances of egoism. Nonetheless, Schopenhauer argued that genuine, altruistic mitleid is present in human beings, and that the extent to which it tends to motivate moral character in particular individuals varies as much, and is as important, as individual differences in intellect.

Schopenhauer saw indisputable proof of the existence of this kind of altruistic mitleid in examples of individuals risking their lives for strangers and examples of individuals making sacrifices with clear harm to their own self-interests. In less spectacular forms, it is a "daily occurrence. Every one has often felt its working within himself; even to the most hard-hearted and selfish it is not unknown" (p. 204). Yet it is quite mysterious how or why an interpersonal process of mitleid operates in human beings, overriding self-interest:

> How can that which affects another for good or bad become my immediate motive, and actually sometimes assume such importance that it more or less supplants my own interests, which are, as a rule, the single source of the incentives that appeal to me? Obviously, only because that other person becomes the ultimate object of my will, precisely as usually I myself am that object; in other words, because I directly desire weal, and not woe, for him, just as habitually I do for myself. This, however, necessarily implies that I suffer with him, and feel his woe, exactly as in most cases I feel only mine, and therefore desire his weal as immediately as at other times I desire only my own. But, for this to be possible, I must in some way or other be **identified** with him; that is, the **difference** between myself and him, which is the precise raison d'etre of my Egoism, must be **removed**, at least to a certain extent. (Schopenhauer, 1840/1903, p. 169-170, bold text in original).

Importantly, Schopenhauer argued that mitleid is based in the sufferings of others, not in their pleasures; in his view, the well-being of others only has significant mitleid-based motivation in that it represents an absence of suffering. The positive emotions of others have only a faint impact on our concern for them, in comparison with their negative emotions.

Schopenhauer argued that what distinguishes a compassionate person, in the most general sense, from other people is that "such a person draws less distinction between himself and others than is usually done" (p. 266).[7] Nonetheless, Schopenhauer argued that genuine mitleid requires a continued awareness of the distinction between the self and the other, and may actually be inhibited if the observer is personally just as distressed as the other. Schopenhauer strongly disagreed with the view that:

> …compassion arises from a sudden hallucination, which makes us put ourselves in the place of the sufferer, and then imagine that we are undergoing his pain in our own person. This is not in the least the case. The conviction never leaves us for a moment that he is the sufferer, not we; and it is precisely in his person, not in ours, that we feel the distress which afflicts us. We suffer with him, and therefore in him; we feel his trouble as his, and are not under the delusion that it is ours… (p. 174).

For Schopenhauer, mitleid involves a dissolving of the *motivational difference* between self-interest and other-interest, but not a loss of awareness that the other person's wellbeing is at stake, rather than one's own. This is important to Schopenhauer, given that it means mitleid is not actually an egoistic impulse.

---

[7] "We may here sum up the characteristics of the two human types above indicated. To the Egoist all other people are uniformly and intrinsically strangers. In point of fact, he considers nothing to be truly real, except his own person, and regards the rest of mankind practically as troops of phantoms, to whom he assigns merely a relative existence, so far as they may be instruments to serve, or barriers to obstruct, his purposes; the result being an immeasurable difference, a vast gulf between his ego on the one side, and the non-ego on the other…Whereas the Altruist discerns in all other persons, nay, in every living thing, his own entity, and feels therefore that his being is commingled, is identical with the being of whatever is alive. By death he loses only a small part of himself. Patting off the narrow limitations of the individual, he passes into the larger life of all mankind, in whom he always recognised, and, recognising, loved, his very self; and the illusion of Time and Space, which separated his consciousness from that of others, vanishes." (Schopenhauer, 1840/1903, p. 279)

Finally, Schopenhauer argued that individual differences in the tendency to experience mitleid are "innate, and ineradicable" (p. 283). This view strongly implies that compassion is stubbornly trait-like. Moral education and appeals to the intellect are incapable of improving a person's mitleid-based motivation. As "it is possible to remodel what one does, but not what one wills to do" (p. 246), the most that reason and instruction can do is to help a person's innate mitleid be expressed more consistently and effectively.

Friedrich Nietzsche is the 19th century philosopher most diametrically opposed to Schopenhauer in his views about compassion, and the two are often discussed together in this domain (e.g., Reginster, 2015). His general arguments against the value of compassion, for instance that it represents a "slave morality" (e.g., Nietzsche, 1887) and that a kind of "tough love" is better for suffering people in the long run (e.g., Nietzsche, 1891)[8], are well-known. Some of Nietzsche's particularly intriguing theories regarding the psychological emergence of compassion, however, which are primarily tucked away in his least-popular (Clark & Leiter, 1997) major work *Morgensrothe* (translated as *Daybreak*, 1881/1997), appear to have been overlooked, even by those who have focused on this aspect of Nietzsche's thought (e.g., Reginster, 2015). Although they may be of only limited importance to the larger picture of Nietzsche's campaign against morality, these psychological speculations are particularly important in regard to the narrower domain of empathy conceptualization.

Nietzsche (1881/1997) mocked Schopenhauer's theory of Mitleid, which he characterized as "a mystical process by virtue of which pity makes two beings into one and in this way makes possible the immediate understanding of the one by the other," as "incomprehensible nonsense" (p. 90); this was a bit unfair of a characterization, given that Schopenhauer did emphasize a self-

[8] "But if you have a suffering friend, be not a resting place for his suffering, but a hard bed, a field cot: thus you will profit him best" (1891/1978, §2:3).

other awareness (Reginster, 2015). Where Schopenhauer saw altruism, Nietzsche (1881) saw both conscious and unconscious self-serving, ranging from pride and feelings of superiority to a desire to alleviate one's own personal distress (e.g., our distress at the sight of another's distress "is our own, as the suffering he feels is his own. But it is only this suffering of our own which we get rid of when we perform deeds of compassion," 1881/1997, p. 133).

Nietzsche (1881/1997) also frequently lampooned the internal logic of Schopenhauer's position, such as by pointing out: "Supposing we felt towards another as he feels towards himself – that which Schopenhauer calls sympathy [Mitleid] but which would be better designated as empathy [Einleid] – then we would have to hate him if, like Pascal, he found himself hateful" (p. 63, original words from the German text added). This particular kind of attack continues to appear up to the current day (e.g., Hall & Schwartz, 2019) in response to those who define "empathy" as requiring an isomorphic matching of emotion between perceiver and the person the perceived.

Moreover, Nietzsche argued that mitleid may actually add to the suffering of the world, rather than lightening it. For instance, he pointed out that if such sympathetic affection for others were much stronger in human beings, it would be so distressing that it would cause us to flee at the approach of others, and we would "bestow upon the sympathetic affection the kind of evil names we now bestow upon egoism" (p. 91). In other words, past a certain point of intensity, "catching" the emotions of others causes a breakdown in sociality.

In place of Schopenhauer's approach, Nietzsche (1881/1997, p. 142, original word in German from the text added) offered his own intriguing theory as to how human beings actually come to understand the perspectives of others:

> Empathy [Mitempfindung] – To understand another person, that is, to imitate his feelings in ourselves, we do indeed often go back to the reason for his feeling thus or thus and ask

> for example: why is he troubled? – so as then for the same reason to become troubled ourselves; but it is much more usual to omit to do this and instead to produce the feeling in ourselves after the effects it exerts and displays on the other person by imitating with our own body the expression of his eyes, his voice, his walk, his bearing (or even their reflection in word, picture, music).

In other words, our understanding of others, in Nietzsche's view, emerges out of a reflex to physically imitate. Although he acknowledged the role that appraisal of the situation sometimes plays in this process, he diverged from the views of Smith and others in arguing that we typically do not make the cognitive effort to engage in such appraisals.

Nietzsche, moreover, argued that the pleasure we take in music, in being in nature, and other aesthetic pursuits emerged from much this same process. These aesthetic sensibilities are "the later-born children of empathy [Mitempfindung]" (p. 90). This idea is highly similar to theories that emerged later, such as the einfühlung perspective as popularized by Lipps (1897). Given that Nietzsche, however, was not widely read until many years after his works were written, and that *Daybreak* in particular has long been perhaps the most neglected of his significant works (Clark & Leiter, 1997), it seems that this field of inquiry largely passed by him without notice.

As to how this process emerged in human beings, Nietzsche (p. 90) theorized that it arose out of our quick need to immediately assess other individuals and animals for potentially threatening intent:

> If we ask how we became so fluent in the imitation of the feelings of others, the answer admits of no doubt: man, as the most timid of all creatures on account of his subtle and fragile nature, has in his timidity the instructor in that empathy [Mitempfindung], that quick understanding of the feelings of another (and of animals)…The capacity for understanding – which, as we have seen, rests on the capacity for rapid dissimulation – declines in proud, arrogant men and peoples, because they have less fear.

As explained elsewhere in Daybreak, Nietzsche argued that the first assessment human beings (and other animals) make when encountering another creature is to "size it up" and automatically

assess one's relative capacities of attack and defense, as well as to immediately scan the creature's appearance for indications of danger (e.g., aggression). It is this primitive reflex that Nietzsche posited as the wellspring of mitempfindung, and it is on this basis that he argued that such understanding declines as fear declines.[9]

Finally, as has been frequently noted (e.g., Reginster, 2015), Nietzsche argued that compassion for others is fuelled by a lack of compassion for oneself (e.g., "I say to you: your love of the neighbour is your bad love of yourselves…You cannot endure yourself and do not love yourselves enough"1881/1978, §1:16). This is essentially the opposite of many more recent conjectures (and findings) that self-compassion is positively associated with compassion for others (e.g., Neff & Pommier, 2013); however, the directionality of the association might depend upon which aspects of empathy one is investigating.

The rise of einfühlung and empathy. In translating the term "einfühlung" as "empathy" into English, Titchener (1909) was adapting the concept proposed and popularized by Lipps (1897).[10] Although Lipps is still often credited with originating the term, it actually emerged much earlier, particularly in the writings of Johann Gottfried Herder (Edwards, 2013; Lanzoni, 2018).

The use of this term by Herder in the 18th century, although substantially different than its later use by Lipps, is quite suggestive. For Herder, "einfühlen" (the verb form of einfühlung) represented a way of *intuitively* feeling oneself into intellectual knowledge through immersive imagination, stepping beyond one's own limited personal perspective (Berlin, 1988; Edwards,

[9] The reader may notice that this hypothesis bears some resemblance to Lykken's (1957) fear hypothesis of psychopathy.

[10] Titchener never explicitly mentions his source for einfühlung, but it is a safe presumption that he came to it via Lipps, according to Debes (2015).

2013). For example, in order to properly understand a historical source document, Herder noted that "the whole nature of the soul, which rules through everything . . . colors even the most indifferent actions, in order to share in feeling this, do not answer on the basis of the word but go into the age, into the clime, the whole history, feel yourself into everything — only now are you on the way toward understanding the word" (Herder, 1774/2002, p. 292).

As Berlin (1988/1997, p. 389) described Herder's notion of "einfühlen," it involves penetrating into the lived experiences of other peoples, to "grasp what it must be like to live, contemplate goals, act and react, think, imagine in the unique ways dictated by their circumstances…" An implication of this kind of process, for Herder, was that it lessened the instinct to moral judgment of those who are quite different than one's own community; before engaging in condemnation, one should first exercise einfühlen to gain sympathetic insight (Berlin, 1988/1997). In promoting einfühlen primarily in regard to whole groups of people, usually separated from one by time and/or geography, Herder's use of the term bears a notable family resemblance to some current uses of the term in political contexts (e.g., Obama 2006). Nonetheless, it does not seem that Herder's use of the term *directly* affected the initial rediscovery of the term "einfühlung" roughly a century later.

In 1873, Robert Vischer reinvented the term "einfühlung" in his doctoral thesis on aesthetics, at a time when interest in the psychological basis of aesthetics was emerging (Debes, 2015). Prominent theorists of the time in this domain, such as Vernon Lee and Kit Anstruther-Thompson, argued that our aesthetic sensibilities were founded in physical movements and sensations that shapes, textures, sounds, and so on, generated in us imitatively when we perceived them (Lanzoni, 2018). These physical imitations then led to emotional experience and a sense of identification with the object.

The psychologist Theodor Lipps, however, argued that this was a grave error and that what actually occurred was simultaneously in the mind rather than first only in the body. Lipps (1903) posited that objects prompt not necessarily a physical imitation, but an "inner imitation" within our mind, which arouses kinesthetic feelings in us, which we then project back into the object (whether that object be an acrobat or a still-life painting).

Moreover, Lipps (1903, p. 375) thought this process caused a psychological identification with the object: "In a word, I am now with my feeling of activity entirely and wholly in the moving figure…I am transported into it. I am, so far as my consciousness is concerned, entirely and wholly identical with it…This is esthetic imitation and it is at the same time einfühlung." In Lipps's theory, this loss of distinction between self and object is an *inherent* aspect of einfühlung, which is "that state in which all consciousness of distinction between myself and the object is lost; when I 'feel myself into the object,' as it were; it is a state in which I am identified with the object without even raising the question as to whether there is such an identification" (Jones (1905, p. 640). This loss of distinction between the self and the other, of course, sharply diverges from the empathy-related conceptualizations of Smith, Schopenhaur, and Nietzsche.

Although einfühlung emerged in the field of aesthetics, Lipps boldly extended it to serve as the process by which we come to "feel ourselves into" other people, not just objects. As he asserted, "There are three spheres of knowledge. I know about things, about myself, and about others…The source of the third type of knowledge is…Einfühlung" (Lipps, 1903, p. 187, as presented in English in Allport, 1937, p. 531). The same process of einfühlung operated in regarding other people as in regarding things: a process of inner imitation generated feelings, which were then projected back into the other person. Moreover, Lipps claimed that einfühlung

is not based in in inference by analogy to our own minds, but is "unique, irreducible, and immediate" (Jardine & Szanto, 2017, p. 87).

Confusingly, Lipps generally treated "einfühlung" as roughly synonymous with "sympathy" (Debe, 2015; Jahoda, 2005), although Jahoda (2005) and Edwards (2013) noted that the conflation with sympathy appears to have been primarily when discussing einfühlung toward persons rather than objects. Interestingly, though, as is frequently mentioned (e.g., Jahoda, 2005; Stein, 1986/1917), Lipps specifically equated "sympathy" with what he called "positive einfühlung," but not with what he vaguely referred to as "negative einfühlung." For Lipps, positive einfühlung occurs when the person freely participates in the einfühlung, in a "harmonious" manner. Negative einfühlung, however, is einfühlung which generates discomfort or discord, and which the observer wishes to reject (Jahoda, 2005).

This distinction offered by Lipps, but never substantially explored in his works and enduringly "cryptic" (Debes, 2015), suggests some potential aspects of heterogeneity within the einfühlung (or empathy) construct. Should einfühlung be parsed into appetitive and aversive facets? Or, perhaps, interpersonal approach-generating and avoidance-generating facets? It is unfortunate that this distinction was never clarified in Lipps's theories.

Outside of aesthetics, some of Lipps's most vigorous critics came from the emerging phenomenological movement in philosophy, particularly Edith Stein (Debes, 2015). In 1917, Stein published her doctoral thesis, "Zum problem der Einfühlung." This work is generally translated in English as "On the problem of empathy" (e.g., Stein, 1917/1986) and, when Stein's work is quoted in English, the reader should be aware that she used the term "einfühlung" rather than "empathy." Although this work is usually only glancingly mentioned in historical tracings of the empathy construct (e.g., Edwards, 2013; Lanzoni, 2018; but see Debes, 2015), it offers a

fascinating and useful view of the theoretical struggles that took place in response to Lipps. In comparison with more commonly discussed critical responses to Lipps by phenomenologists, such as by Max Scheler, Debes (2015, p. 313) noted that Stein's work is "more focused and for that reason more instructive."

As Stein broadly defined her subject, "Empathy…is the *experience* of foreign consciousness in general, irrespective of the kind of experiencing subject or of the subject whose consciousness is experienced" (p. 11, emphasis added). Meneses and Larkin (2012, p. 162) argue that Stein's conceptualization is congruent with the definition: "Empathy is to experience another's experience at once, and know that it is another's experience, simultaneously, as part of the same act."

In her extensive examination of einfühlung, Stein was working within the framework of the emerging "phenomenology" movement within philosophy, with her dissertation work supervised by Edmund Husserl (the foundational figure of the movement). Without going into depth about this particular intellectual context (for a more technical philosophical investigation of einfühlung within Husserlian phenomenology, see Manganaro, 2017), two key points are worth noting.

First, as discussed by Stein (1917/1986), phenomenology focuses on subjective experiences in the consciousness of individuals. In considering the nature of "einfühlung" and related states or processes, Stein was concerned primarily with how einfühlung manifests *as it appears in conscious experience*. As asserted by Husserl and Stein, comprehensive analysis of the phenomenological character of a construct must first be conducted before other scientific/philosophical investigations can be justified (the "phenomenological reduction," as described in Stein, 1917/1986). In particular, Stein (1917/1986) argued that any subsequent

scientific psychological investigation of einfühlung must be justified by prior phenomenological analysis:

> …Psychology is entirely bound to the results of phenomenology. Phenomenology investigates the essence of empathy, and wherever empathy is realized this general essence must be retained. Genetic psychology, presupposing the phenomenon of empathy, investigates the process of this realization and must be led back to the phenomenon when its task is completed. (p. 22)

This is similar to the kind of stance taken by some early psychologists in regard to introspection (e.g., "Within the sphere of psychology, introspection is the final and only court of appeal," Titchener, 1896, p. 341).

Second, within Husserlian phenomenology, consciousness is understood as typically directed at or toward something, a concept often discussed in terms of "intentionality" (e.g., Smith& McIntyre, 1971). In Stein's treatment, einfühlung requires that the experiencing person's consciousness is directed at a foreign consciousness, not at his or her own consciousness. As a result, for instance, Stein (1917/1989, p. 23) excluded simple emotional contagion from the einfühlung construct, arguing that "it is certain that as we are saturated by such 'transferred' feelings, we live in them and thus in ourselves. This prevents our turning toward or submerging ourselves in the foreign experience, which is the attitude characteristic of empathy."

Stein also excluded projection and simulation accounts from her conceptualization of einfühlung, classifying them, instead, as other non-einfühlung ways of knowing about others. So, for her, Lipps's account of einfühlung, as unconsciously imitating another and then projectively attributing one's subsequent feelings to the other, is excluded because it fails to account for how we experience the strange "otherness" of others; it is merely an experience of oneself (Manganaro, 2017; also see Zahavi & Rochat, 2015, for a more general discussion of this phenomenological critique). For similar reasons, Stein excluded much of the more effortful kind

of simulation account, offered by Smith and others, wherein we imagine what we would feel in another's situation and then projectively attribute our feelings to the other (it is worth noting, though, that this would not exclude cases offered by Smith where the person more fully adopts the perspective of the other, rather than merely projecting).

In her phenomenological investigations, Stein observed that foreign consciousness can often be experienced as truly foreign, with knowledge of the other emerging without imitation or projection. Though imitation and simulation-projection may explain some of our experiences of others, there is still much room for Steinian einfühlung to explain other instances. I suppose the way to respond to this claim, on Stein's terms, is for us to examine our own conscious experiencing of others: in our phenomenal consciousness, how do *we* experience the egos, personalities, and so on, of others we encounter?

In the initial approved version of her thesis, Stein devoted substantial attention to charting the intellectual history of the general concept of einfühlung. Unfortunately, this part of her doctoral work was not included in the published version and appears to no longer be extant (McDaniel, 2016). The content of the text, however, seems to indicate that her historical treatment extended beyond uses of the term "einfühlung" and encompassed prior work on "sympathy," such as that of Adam Smith, whom she briefly discusses in the published text.

In the foreword to the published version, Stein (1917/1989) explained that her historical treatment encompassed aesthetic, cognitive, and ethical lines of thought. Though she examined them separately in her (now lost) historical treatment, she observed that the problems of einfühlung were largely the same in all of them and speculated that prior conceptual failures in each separate thread might be overcome by merging them into a single, basic inquiry.

In Stein's thesis, einfühlung is described as a way of knowing what the other is experiencing, not as a way of responding to that knowledge (e.g., not compassion). This way of knowing is not founded in deductive or inductive reasoning, nor the kind of inference via inner imitation espoused by Lipps, but is primarily directly "given" in an intuitive and experiential manner.[11] This is not to say that effortful appraisal, inference, and controlled reasoning do not play a role in shaping the contours of how einfühlung unfolds, but they are assisting processes, not replacing. This is much in the same way that our own inner experiences appear to be "given" immediately to us, but can be elaborated through reflection (e.g., "I am feeling anxious right now, but I need to reflect for a bit in order to understand what and why I am feeling"). This is very similar to the view held by Lipps, but it is seemingly "radicalized" by Stein (Jardine & Szanto, 2017, p. 87), with Stein refusing to reduce einfühlung to inner imitation and projection.

In Stein's view, this way of knowing is sui generis, involving intuitive, direct understanding of the other as an "I," a human personality with an inner life, not as merely an object. In other words, it is a way of knowing that has an inherent aspect of "humanizing" the other. As a consequence, this way of knowing is not necessarily emotionally "cold" or ethically detached; my supposition is that this aspect may be what led Stein to feel that the ethical line of inquiry and the epistemological line of inquiry could be merged.

Stein, however, did not attempt a scientific psychological explanation as to how this capacity develops and operates in human beings. It is difficult to conceive how knowledge about others can be directly "given" to us, emerging out of something more or different than inferential processing of some kind.

[11] Reflecting on the discussion of einfuhlung by Lipps and by his phenomenologist critics, Allport (1937, p. 533) stated that "the theory of empathy is a peculiar blend, and must in fact be regarded both as a theory of inference and as a theory of intuition, depending somewhat upon the coloring given it by different authors."

As C.S. Peirce (1868), the enduring scientist-philosopher (and, arguably, the first American experimental psychologist, e.g., Cadwallader, 1974), argued in his "Questions Concerning Certain Faculties Claimed for Men," all instances of what may appear to be direct intuition are actually products of inferential processes within the mind. Though something may appear in one's phenomenal consciousness (which Peirce, 1905, later termed the "phaneron," to be studied through "phaneroscopy") to have been directly given to the mind, the appearance actually arose from inferential processes of which one is simply unaware. As Peirce (1868) demonstrated, such as by discussing the blind spot on the retina of the eye, this is true even of sense perceptions, which are mediated by inferential processes even though they appear to be directly given to us. Peirce's conclusion was that the entirety of one's phenomenal consciousness is inferentially mediated, not directly given via irreducible intuition.

It is unclear whether, in describing the experience of einfühlung as directly given rather than inferential, Stein meant this only in reference to its *appearance* in phenomenal consciousness and would have accepted that it ultimately arose from inferential processes outside of consciousness. Nonetheless, it may be worth noting that her stance (and those of other proponents of einfühlung) bears a subtle family resemblance to epistemological claims put forth in other domains, such as Poincare's (1908) claim that mathematical discovery requires non-deductive intuitive insight. For example, consider Albert Einstein's (1932, p. 10) argument for intuitive einfühlung in physics research:

> The supreme task of the physicist is the discovery of the most general elementary laws from which the world-picture can be deduced logically. But there is no logical way to the discovery of these elemental laws. There is only the way of intuition, which is helped by a feeling for the order lying behind the appearance, and this Einfuehlung is developed by experience.[12]

[12] As a side note, in appreciation to some of the sincere scientists who have instructed me in graduate school:
> Many kinds of men devote themselves to Science, and not all for the sake of Science herself. There are some who come into her temple because it offers them the opportunity to display their particular talents.

To Stein, einfühlung is a complex phenomenon that can apparently take many forms. Yet, she defined the "highest level of the consummation of empathy" as "where we are 'at' the foreign subject and turned with it to its object," in a manner of "inner participation" in a foreign consciousness (p. 12). In the sense, she agreed with Lipps, at least as to what constitutes einfühlung in its highest form.

Stein (1917), however, disagreed with Lipps in his contention that, in einfühlung, the two consciousnesses necessarily merge into one. Although a kind of self-forgetfulness may sometimes occur, such that one has lost sight of his or her own "I" and has become completely absorbed in the experiencing of another, this is not an aspect of einfühlung itself. It is, however, something that can only occur as a result of einfühlung; it cannot occur on its own without the aid of einfühlung.

Stein also distinguished "einfühlung" from sympathy. To Stein, einfühlung does not require that one's own emotional state mirror that of the foreign consciousness. "Sympathy," however, is emotional mirroring that *may* emerge out of and alongside einfühlung. In other words, if I direct my consciousness towards your subjective consciousness, experientially comprehend your joy, *and also* feel joy myself, then I am in sympathy with you. For example, imagine you are at an opera, experiencing a sense of rapture during a particular aria. You see a man in front of you clearly also deeply aesthetically engrossed. You comprehend his pleasure as

---

> To this class of men science is a kind of sport in the practice of which they exult, just as an athlete exults in the exercise of his muscular prowess. There is another class of men who come into the temple to make an offering of their brain pulp in the hope of securing a profitable return. These men are scientists only by the chance of some circumstance which offered itself when making a choice of career. If the attending circumstance had been different, they might have become politicians or captains of business. Should an angel of God descend and drive from the Temple of Science all those who belong to the categories I have mentioned, I fear the temple would be nearly emptied. But a few worshipers would still remain—some from former times and some from ours. To these latter belongs our Planck. And that is why we love him. (Einstein, 1932, p. 7)

his own, while you are also experiencing much the same pleasure in your own person. This would constitute "sympathy" in Stein's usage. Importantly, for Stein, the emotional mirroring in sympathy is distinct from that in emotional contagion; in sympathy, one's consciousness is directed toward the foreign consciousness, whereas it is turned toward one's own "I" in emotional contagion.

Stein provided numerous examples and hypotheticals in her thesis, and one in particular seems illustrative of the difference between einfühlung and sympathy, at least as she has defined the terms. It is also fascinatingly historically ironic[13].

> I experience every subject who I empathically comprehend as experiencing a value as a person whose experiences interlock themselves into an intelligible, meaningful whole. How much of his experiential structure I can bring to my fulfilling intuition depends on my own structure….I can be sceptical myself and still understand that another sacrifices all his earthly goods to his faith. I see him behave in this way and empathize a value experiencing as the motive for his conduct. The correlate of this is not accessible to me, causing me to ascribe to him a personal level I do not myself possess. In this way, I empathically gain the type of a *homo religiosus* by nature foreign to me, and I understand it even though what newly confronts me here will always remain unfulfilled (Stein, 1917/1989, p. 115).

In other words, a non-religious person can, through einfühlung, come to experientially understand, to some incomplete extent, what it is like to be a person with strong religious sentiments, even if sharing those sentiments is not possible for that person. In our day, we could say that a fervent Democrat can come to better understand the humanized perspective of a supporter of Donald Trump through einfühlung, even if sharing feelings in "sympathy" is not possible. This is highly similar to the kind of pluralistic imaginative insight emerging out of Herder's original use of the term einfühlen. As we will discuss later, though, this kind of non-

[13] An avowed atheist, of Jewish birth, Stein converted to Catholicism in the 1920s and entered a convent in 1934. She was murdered at Auschwitz in 1942. She was beatified by Pope John Paul II in 1987 and canonized as Saint Teresa Benedicta of the Cross in 1998. For an excellent review of her life and formation as a philosopher, see MacIntyre (2007).

matching (i.e., the two are not sharing the same emotion) dynamic would be excluded from the empathy construct by many current theorists (for a contemporary phenomenology-influenced critique in this vein, see Zahavi & Rochat, 2015).

A final note regarding Stein's dissertation: she posited that God, if such a thing exists, "can comprehend people's lives in no other way [than einfühlung]. As the possessor of complete knowledge, God is not mistaken about people's experiences, as people are mistaken about each other's experiences" (Stein, 1917/1989, p. 11). Theological concerns aside, the notion that God understands individuals through complete, infallible einfühlung points to what we might consider as a regulative ideal, an idealized perfection of a process that can be used as a tool to better understand a concept.

For example, consider what a specific civil rights activist might have felt and thought during a specific 1960s sit-in at a segregated lunch counter. Perfect einfühlung for this activist in this episode, such as Stein proposes for a possible God, would mean knowing her full subjective experiencing: the *smell* of the coffee, the chaotic *sounds* of the angry white people accosting her, the *coolness* of the counter as her hands held on to it, the sensation of her *heart pounding* in her chest, the *orange-ness* of the juice spilled in front of her, her shifting/flowing *emotions*, and everything else in her stream of consciousness. More than that, perfect einfühlung would mean knowing how this activist perceived the society in which she lived, the thoughts and feelings that led her to bravely take a stand at that lunch counter. In other words, this God would completely know what it was *like to be her* in these moments.

Instead of a God, one could perhaps approximate this regulative ideal by considering the virtual reality and consciousness simulation (or, perhaps, consciousness "recording") capabilities of a hypothetical civilization far more technologically advanced than our own. Through

sophisticated tools, such a civilization may allow for einfühlung experiences that somewhat approach god-like degrees of immersive completeness.

Regardless of whether God or such god-like technologies are real or possible, though, contemplation of their potential einfühlung capacities may help illustrate the idea of "knowing" a foreign consciousness in this kind of einfühlung manner, rather than merely knowing something as a complex object. Doing so also clearly highlights the limitations of human beings in this regard. Though some people may "feel themselves" into the experiences of others much more than does the average person, there is nonetheless a wide gap of "blindness" remaining in us all.

Although Titchener (1909) coined the term "empathy," he did relatively little with it and is not generally further discussed in relation to the construct (Debes, 2015). His description of the term's meaning, though, is worth noting: "We have a natural tendency to feel ourselves into what we perceive or imagine…This tendency to feel oneself into a situation is called empathy" (Titchener, 1915, p. 198). Although empathy was frequently used in literature on aesthetics in the coming decades (e.g., Brown, 1915; Ducasse, 1928; Howes, 1913; O'Neill, 1914), its extension into other domains of psychological inquiry was infrequent and eclectic. Debes (2015) and Edwards (2013) have speculated that the limited interest in einfühlung and empathy within scientific psychology, in the decades after its brief burst into discussion, most likely was related to a collapse of interest in introspective methodologies in the early 20th century.

For one such eclectic example of its discussion, though, about the same time as Stein was finishing her dissertation, the American psychiatrist Elmer Southard, a President of the American Medico-Psychological Association and student of William James (Lanzoni, 2018), was perhaps the first to bring the new English term "empathy" into the field of clinical psychology. In his 1918 article in the *Journal of Abnormal Psychology*, titled "the Empathy Index in the Diagnosis

of Mental Diseases," Southard (1918) argued that "empathy" should not be reserved only for use in the fields of aesthetics and art, and that it should be recognized as a method by which clinicians research, diagnose, and treat individuals.

Southard defined empathy as "how far we read or feel ourselves into the person, group, nation, or state" (p. 200), describing it as an effortful process, requiring imagination and an intentional stance, not simple emotional reflex. Moreover, though empathy has a humanizing element, it "is more intellectual than emotional. Though empathy readily leads the sympathies, it may well stop at a coldly rational view of its object…The human interest of sundry newspaper stories is often a matter of empathy rather than sympathy" (p. 203).

Southard's "empathy" was based in projection: how easily can one project one's own thoughts and feelings into another person or group of people? In his use, it does bear some relationship to its use in aesthetics: it borders on asking "how pleasant or unpleasant, beautiful or ugly" is this individual's personality to you?

Southard hypothesized that the degree to which clinicians could empathize in this way with patients suffering from various forms of psychopathology could become a valuable research domain, and that, eventually, different forms of psychopathology could be characterized on a metric or ordinal "empathy index." To this end, Southard (1918, p. 207, emphasis added) provided a series of questions for clinicians to be asked, to assess their empathy towards a patient, such as:

> How far can you read or feel yourself into the patient?
> Can you identify yourself with the patient?
> Disregarding your sympathies and abhorrences, can you empathize with the patient?...
> Can you put yourself in his place?...
> Does this patient's reaction seem intrinsically *human* or is there something extrinsic and *non-human* about the reaction?

Although the distinction may be rather subtle in practice, this "empathy index" was meant to measure how much psychological common ground a clinician could see between himself and a patient, not to measure how well the clinician believed he could predict a patient's thoughts, feelings, and behaviors. As pointed out by Lanzoni (2018, p. 106), this index, illustrated by its use in a book of case studies entitled *The Kingdom of Evils (*Southard, Jarrett, & Pound, 1922), "served as a crude likeability or attractiveness scale," reflecting common gender, racial, and other biases.

A few years after Southard introduced "empathy" to clinical psychology in the U.S., Gordon Allport began discussing "empathy" in a broader context of personality psychology (e.g., Allport, 1924, 1929), and in a manner that went much further beyond the term's use in aesthetics. For instance, Allport (1924) extended the empathy construct to the question: how do we usually holistically assess the total personalities of others, rather than only individual trait domains? Although noting that substantial advances were being made in conceptualizing and measuring individual trait domains, he argued that "the concrete personal life never has to do with a single capacity or trait in isolation" (Allport, 1924, p. 133) and that the ways in which different traits interacted with one another in totality was not captured by this divided trait method.

Instead of focusing on individual trait domains, he argued for new attention to the "undivided personality," giving credit to gestalt psychologists. As for how to best analyze this undivided personality, he turned to the method of einfühlung or "empathic process" proposed by Lipps. Though he did not have a strong belief as to the particular mechanisms underlying empathy, Allport (1924, pp. 139-140) stated, "Whatever the genetic[14] source of this process may be, we are forced to recognize its supremacy in determining our apprehension of personality.

[14] This word, as here used by Allport, refers to developmental rather than heritable origins.

When we fail to empathize with a person we fail to understand him; conversely, the greater our emphatic (sic) ability the greater is our capacity for understanding."

Allport continued to refine his understanding of empathy in the coming decades. In the decade prior to his death, for instance, Allport (1961, p. 536) held to an enduringly valuable definition of empathy as "the imaginative transposing of oneself into the thinking, feeling, and acting of another," while also continuing to contend that empathy remained a riddle as "the nature of the mechanism is not yet understood" (Allport, 1968, p. 30).

Allport (1924, p. 140), following Lipps, noted that empathy is "characterized by its unconscious functioning. Hence it is that this process customarily lies outside the limits of introspective examination." In other words, although individuals tend to rapidly and holistically make empathic assessments as to the personalities of those around them, they will struggle to identify *how* they actually make these assessments and, perhaps, to what *extent* they make them (whether in frequency or in depth). Though not explicitly mentioned by Allport (1924), a direct implication of this, of course, is that self-reports of some facets of empathy may be beyond the metacognitive abilities of respondents.

It is perhaps worth mentioning at this point that einfühlung/empathy, at least when promoted as a method for scholarly inquiry, came to be viewed by some as a form of unscientific intuitionism (e.g., Sarbin, 1944). The philosopher C.I. Lewis (1929, p. 410), for instance, argued that the notions of einfühlung (or empathy, he used both words) of his day represented a "native longing of humanity to transcend the bounds of subjectivity; to know our object not only in the pragmatic sense of successful prediction and control but in a deeper sense of somehow coinciding with its nature… characteristic of mysticism, of intuitionism." (p. 410). Similarly, in psychology, R.B. Cattell (1937, p. 130-131) argued, "The intuitionist desires to know another

personality directly, by empathy instead of in an objective fashion….in practice its use leads in the end to errors more impressive than its successes…. propounded as an independent method of arriving at psychological knowledge, it would seem to be a pure illusion."

At some point around the same time Allport promoted it as a method of personality assessment, the term "empathy" was first introduced to the general American public. For instance, in 1927, a quirky New York Times piece (Ware, 1927) questioned a number of psychologists (e.g., E.L. Thorndike and J.M. Cattell) for their opinions about the public's newfound love of trivia quizzes. Within the piece, "empathy" was introduced as "one of the newest pet words of psychologists" and defined as, "Empathy is sympathy carried a step further. Sympathy is feeling for a thing, whereas empathy is feeling into the thing to such an extent that you become part of that thing yourself." As argued by A.A. Brill, a psychiatrist and translator of Freud, empathy was key to the attraction of both prizefights and also the new trivia question books: empathy allowed one to feel connected with the fighters in the ring and also to feel connected with the great minds of the world through trivia books (Ware, 1927).

Despite the whimsical introduction in the nation's most popular newspaper, though, according to Google Ngram, frequency of usage of the term "empathy" remained essentially the same (very little) until beginning to steadily rise starting in 1942 (and continuing to rise to the present day). As documented by Lanzoni (2018), it was not until the 1950s that the term surged into wider use in American life (with publications frequently including an accompanying definition to assist readers).

Empathy's popularity within academic psychology, or lack thereof, largely tracked its popularity in general usage. Between 1909 and 1948, Buchheimer (1963), for instance, found only a handful of rare references to empathy in scientific psychology.

Outside of psychology, however, the construct continued to be elaborated in sociology, particular in the domain of ethnography and interviewing, even though it was not necessarily termed einfühlung or empathy (Edwards, 2013). As described by Edwards (2013), ethnographic methods as promoted by the "father of American ethnography," Franz Boas, which were influenced by the ideas of Herder, involved stepping outside one's own perspective and participating, even emotionally, in the perspectives of those one is studying. As Burgess and Cottrell (1939, p. 334), a past and a future president of the American Sociological Association, explained:

> This method, if it may be called a method, is that of empathy, of sympathetically entering into the experiences and attitudes of another person through the medium of an interview or a personal document and thus, for the time, identifying oneself with the other and taking his role. By this procedure the student of human behavior is able both to appreciate the attitudes and acts of the subject from the subject's point of view and to understand them in the perspective of knowledge of other cases.

The re-emergence of empathy into scientific psychological investigation, as a topic of general import rather than merely a methodology, was linked to its discussion in sociology. In 1949, Cottrell and Dymond, a young psychologist who had experience with ethnographic interviewing methods, lamented that essentially no research attention had been paid to empathy within academic psychology and argued that empathy was of critical importance in: analyzing and treating personality disorders, interpersonal dynamics and communicative processes, and self-insight. Cottrell (1950, p. 708) suspected that the failure to investigate empathy stemmed from the fact that it is inherently difficult to study with the empirical tools then available, while also observing that empathy needed to be adequately defined rather than left "in the realm of mysticism" or "circular reasoning."

Cottrell and Dymond (1949, p. 357) announced, "whatever the reasons, it is our contention that research in this area can no longer be avoided." Specifically, they argued that

dispositional trait empathy was a crucial area of inquiry. Reporting on a series of studies which were admittedly "crude" and "none too rigorous" (p. 357), Cottrell and Dymond (1949, p. 359) tentatively indicated that people high in trait empathy "appeared to be emotionally expressive, outgoing, optimistic, warm people who had a strong interest in others," whereas those low in empathy were "rather rigid, introverted people whose emotional life appears inhibited but who are subject at times to poorly controlled outburst of emotional behaviour."

In preparing her doctoral thesis on empathy, Dymond noted that very few people knew the term and she had to define it in nearly every conversation (Lanzoni, 2018, personal interview). In examining the use of the terms "empathy" and "einfühlung" for discussion in her dissertation, mostly in the field of aesthetics and some scattered uses in psychoanalytic literature, she concluded that common conceptualizations of the term, as *projecting* one's own feelings into something or someone else, were inappropriate (Lanzoni, 2018). As a result, she formed her own definition, positing empathy, in Dymond (1949, p. 343), as "the imaginative transposing of oneself into the thinking, feeling and acting of another and so structuring the world as he does."

Importantly, like Stein (1917), Dymond (1950, p. 344) emphasized that empathy must be distinguished from projection: "Projection seems to be an antithetical process to empathy since projection involves the attribution of one's own wishes, attitudes and behavior to some thing, or some one other than the self. If projection is involved, therefore, the thoughts and feelings of the self are attributed to the other rather than those of the other being experienced." The degree to which empathy cannot be contaminated by projection, however, has never been clearly specified. Even highly empathic experiences likely involve some limited degree of projection.

Dymond (1949), in one of the first attempts to generate a test of empathic accuracy, had dyads rate six basic personality trait domains for (a) how they rated the other person, (b) how

they thought the other person would rate him/her, and (c) how they thought the other person would rate himself/herself. The degree of correspondence between how they predicted the other person would rate and how they actually rated was used to calculate accuracy scores.

Although the validity of Dymond's empirical methods was strongly questioned (e.g., Cronbach, 1955; Lindgren & Robinson, 1953; Murstein, 1957), it appears that her work did help spur further attention to the empathy construct (cf. Buchheimer, 1963, who credited Cottrell and Dymond with helping to reinvigorate empathy research). Other researchers jumped into the fray, exploring ways to objectively disentangle empathy from projection, but with little success: "by the end of the decade, validity measures for empathy tests in industry, the clinic, and the laboratory all came up wanting" (Lanzoni, 2018, p.185). Wispé (1987) argued that the mounting failures of tests of empathic ability spurred a shift in focus toward the measurement of emotional empathy (e.g., empathic concern) instead. As observed by Mehrabian and Epstein (1972), though, self-report measures that emerged failed to demonstrate adequate validity, whereas proposed physiological measurements lacked specificity to empathy as opposed to other aspects of emotional experience.

In 1951, Dymond was hired by Carl Rogers, and the two worked together on conceptualizing empathy and other processes in therapeutic practice (Lanzoni, 2018), producing, for instance, the classic *Psychotherapy and Personality Change* (Rogers & Dymond, 1954). Throughout the 1950s, empathy rose from the basement dustbin of psychology up to a place of popularity on the mantle.

Among psychologists, Carl Rogers is perhaps the figure most closely associated with the empathy construct. As Wispé (1987, p. 29) concluded in the 1980s, "Without doubt, the present popularity of empathy as a construct comes from Rogers's emphasis on it," while she also

concluded that his definitions of empathy were consistent with the general sweep of the history of the construct.

In the early years of his career as a practicing psychologist, Rogers, under substantial influence from the ideas of Otto Rank and others, came to believe that the most effective therapeutic approach entailed developing an emotional connection with the client by reflecting the client's feelings back to them, in a non-judgmental, non-moralizing manner; speaking much less and listening much more; typically refraining from offering interpretation or advice; and being sympathetically engaged with the client, but without being too personally wrapped up in the client's feelings (Lanzoni, 2018).

Much of Rogers's non-directive approach, which he started referring to as "client-centered therapy" in 1945, strongly resembled "empathic processes"; it was not until 1948, however, that he began explicitly referring to "empathy" as a part of his approach (Lanzoni, 2018). Over the years, Rogers continually refined his understanding of "empathy" as a driver of therapeutic change (Rogers, 1973, 1975). At a middle stage in his explorations of the construct, for instance, Rogers (1959, pp. 210-211) defined empathy as follows:

> The state of empathy, or being empathic, is to perceive the internal frame of reference of another with accuracy and with the emotional components and meanings which pertain thereto as if one were the person, but without ever losing the 'as if' condition. Thus it means to sense the hurt or the pleasure of another as he senses it and to perceive the causes thereof as he perceives them, but without ever losing the recognition that it is *as if* I were hurt or pleased or so forth.

In other words, empathy is more than just accuracy; it means to *sense* the world as if one were the other person but accompanied by a clear understanding that the sensations and feelings are the other's rather than one's own. In this sense, empathy is a controlled process: one seeks a balance between (a) being emotionally detached and (b) becoming too overly emotionally entangled.

Importantly, Rogers stressed that authentic empathy is a process that tends to be positive and validating for the person with whom one is empathizing (e.g., Rogers, 1951, 1975). For Rogers, empathy is not merely accurate prediction of a person's thoughts and behaviors, but an emotional stance toward the other person, one which is sincerely interested in the other's inner life, for the other's sake rather than for one's own. As a result, Rogerian empathy is inherently bound up with positive regard and care for the other person (Bozarth, 1999, 2009). Moreover, Rogerian empathy is not only a way of gaining knowledge, but a way of building a validating relationship. Ultimately, this relationship itself is healing: "The words – of either client or counselor – are seen as having minimal importance compared with the present emotional relationship which exists between the two" (Rogers, 1951, p. 172).

One resulting difficulty, in narrowing Roger's conceptualization of empathy, is that it is difficult to discern where the empathy construct ends and another construct, such as non-judgmentalism, begins. As will be discussed further below, it appears that Rogers perceived them to all be related: a person cannot fully empathize with another unless he or she cares about the other person to some degree, is able to put aside his or her impulses to judge the other person, and is secure and emotionally regulated enough to immerse him or herself in the inner world of the other without becoming distressed or agitated.

As explained by Lanzoni (2018), the proselytizing of empathy by Rogers, his students, and his colleagues was so successful that the empathic approach to psychotherapy eventually lost its distinctiveness, as empathy became a generally accepted and adopted aspect of most other approaches to psychotherapy. At the same time, empathy came to be viewed as a broadly positive term, though perhaps vaguely and inconsistently defined, among much of the general public (Lanzoni, 2018).

As a final stop in this brief historical treatment, consider the 1975 issue of *The Counseling Psychologist, 5*(2), dedicated to the topic of empathy. In this issue, different authors presented markedly varying definitions of the construct. Interestingly, the battle lines drawn in the articles in that issue are similar to those in 2019, giving the uneasy impression that, despite thousands of research articles examining empathy in the last two generations (cf. Hall & Schwartz, 2019), the field may not be substantially closer to convergence on conceptual definitions and distinctions than it was in the 1970s (and may actually be further away from convergence).

Iannotti (1975, p. 22, emphasis added), for example, defined empathy broadly as "responsiveness of an individual to the feelings of another person. Individuals who are high in empathy are those who frequently and *appropriately* respond to the feelings of others." Iannotti distinguished empathy and sympathy by adopting the perspective of Wispé (1968, p. 44): "In empathy, I try to feel your pain. In sympathy, I know you are in pain, and I sympathize with you, but I feel my sympathy and my pain, not your anguish and your pain." Iannotti (1975) also lamented that the different definitions and measurement methods of "empathy" made it difficult to integrate across studies.

Feshbach (1975, ps. 25-26), in contrast, offered a much more limited definition of empathy that "restricts the empathy reaction to a match in affective response between subject and object….if a child feels angry after witnessing another person's manifestation of anger, that process would be labeled as empathic." Iannotti (1975), however, noted that the isomorphic matching advocated by Feshbach "makes the distinction between sympathy and empathy difficult to discern" (p. 23).

Furthermore, Feshbach (1975) strongly argued that the social understanding of others, by itself, should not be considered sufficient to constitute "empathy." Given the longstanding centrality of emotion in the empathy construct, Feshbach (p. 25; also see agreement by Iannotti, 1975) argued that "when empathy is defined solely in cognitive terms, it has little theoretical utility beyond that contributed by the cognitive functions themselves." Moreover, some aspects of interpersonal cognition should be excluded from the empathy construct altogether, such as "projection," given that it attributes characteristics of the perceiver to the person perceived, whereas empathy should move in the opposite direction (Feshbach, 1975; also see agreement by Iannotti, 1975).

Hogan (1975, p. 15) adopted a "role-theoretical" definition of empathy as "the process of representing to oneself the expectations that others hold with regard to one's behavior," as exemplified in statements such as "is socially perceptive of a wide range of interpersonal cues" and "seems to be aware of the impression he makes on others." As a result, Hogan closely associated empathy with agreeableness, even-temperedness, likeability, and social tact, characteristics strongly encompassed in his (1969) Empathy Scale (cf. factor analyses by Johnson, Cheek, & Smither, 1983), which appears to have been the first widely used self-report empathy questionnaire. In crafting his Empathy Scale, Hogan was led by descriptions and ratings of the "empathic" person by both psychologists and also laymen, finding that they largely agreed with one another; as a result, his Empathy Scale might be considered more a reflection of "empathy" as it was generally used in society, rather than a reflection of a particular nuanced academic definition.

Rogers (1975, p.4, emphasis added), however, gave the most nuanced, multi-faceted definition of empathy, not just in this particular issue of *The Counseling Psychologist* but in much of the broader literature. Though lengthy, it should be presented in full:

> The way of being with another person which is termed empathic has several facets. It means *entering the private perceptual world of the other and becoming thoroughly at home in it*. It involves being sensitive, moment to moment, to the changing felt meanings within this other person, to the fear or rage or tenderness or confusion or whatever, that he/she is experiencing. It means *temporarily living in his/her life*, moving about in it delicately without making judgments, sensing meanings of which he/she is scarcely aware, but not trying to uncover feelings of which the person is totally unaware, since this would be too threatening. It includes communicating your sensings of his/her world as *you look with fresh and unfrightened eyes at elements of which the individual is fearful*. It means frequently checking with him/her as to the accuracy of your sensings, and being guided by the responses you receive. *You are a confident companion* to the person in his/her inner world. By pointing to the possible meanings in the flow of his/her experiencing, you help the person to focus on this useful type of referent, to experience the meanings more fully, and to move forward in the experiencing.
>
> To be with another in this way means that *for the time being you lay aside the views and values you hold for yourself* in order to enter another's world without prejudice. It some sense it means that you lay aside your self and *this can only be done by a person who is secure enough in himself* that he knows he will not get lost in what may turn out to be the strange or bizarre world of the other, and can comfortably return to his own world when he wishes.

This Rogerian view of empathy is directed primarily at therapists, but Rogers argued that it can be re-articulated in "terms which are perfectly understandable by contemporary youth, or citizens of a beleaguered inner city" (p. 5). At the same time, though, Rogers noted that his definition, as he specifically offered it, is "hardly an operational definition, suitable for use in research" (p. 4); it would need to be re-formulated for such purposes.

Importantly, Rogers argued that empathic understanding is not separate from actual caring or compassion: "It might seem that we have here stepped here into another area, and that we are no longer speaking of empathy. But this is not so. It is impossible accurately to sense the perceptual world of another person unless you value that person and his world – unless you in

some sense care" (p. 7). In order to truly step into another person's shoes, one must be able to see the person as a human personality worthy of care.

Moreover, Rogers (1975) indicated that the caring of the observer, as perceived by the subject of empathy, is what leads the subject to lower his/her guard and better allow the observer to enter into his/her private experiential world. In other words, although caring may be theoretically distinguishable from empathic understanding, the two might be so strongly entangled that it makes sense to understand empathic caring as part of the empathy construct.

Though the articles in the issue defined empathy differently, all strongly related it to moral development, prosociality, and other overwhelmingly positive correlates. For example, as Aspy (1975, ps. 13-14) passionately stated:

> [Empathy] is a universally crucial ingredient of healthy human development which, in turn, is essential for a peaceful, productive human race. Therefore, there should be a massive effort in empathy training with all kinds of government and foundation support. Why not establish an empathy thrust to counterbalance our threat to seize Arab oil by force if necessary?....in light of the destructive potential of our military weapons, the amelioration of human tensions through increased levels of empathic understanding appears to be one of the eminently sane alternatives to a holocaust.

From the perspective of the 1980s, reflecting on the history of the empathy construct, Wispé (1987, p. 17) concluded that "there have been some – relatively slight – differences of emphasis and connotation in the term 'empathy' from one field to another, but there has also been a remarkable consistency of usage and meaning." Specifically, across many perspectives, she discerned that "the hard-core meaning of Einfühlung/empathy has always been the process whereby one person 'feels her/himself into' the consciousness of another person" (p. 34). At the time, Wispé (1987) seemed to indicate that the Rogerian view of empathy was dominant for many psychologists, and quite consistent with the history of the construct.

Current Conceptualizations of Empathy

Whether Wispé (1987) was correct or not in concluding that the usage and meaning of empathy had been remarkably consistent up into the 1980s is highly debateable. What is clear, though, at this *current* stage in the history of the construct, is that there is incredible divergence between scholars as to the usage and meaning of "empathy," differences so extreme, in some cases, that empathy is now even perceived by some as a harmful psychological phenomenon (Bloom, 2017a; Prinz, 2011). I suspect that part of the present confusion may be due to the field of empathy research having often neglected some of the lessons of its intellectual history.[15]

Empathic understanding ("cognitive empathy"). Cognitive empathy, as it is discussed in "empathy" research, is an exceptionally heterogeneous construct. In coming to recognize, predict, and understand others, human beings rely on a wide range of capacities and tendencies, such as: basic theory of mind; the capacity to emotionally and cognitively simulate such thoughts and feelings within oneself; broader abilities in generating theories about the world; recognition of different kinds, intensities, and mixtures of emotions from verbal and non-verbal cues; imagination; emotion regulation, ranging from suppression to re-appraisal; general executive functioning and working memory; and, perhaps most importantly, *motivation* to expend cognitive effort in the pursuit of gaining such knowledge of others.

Given that the empathy construct has long been tied to emotional experiencing of some kind, and that some processes allow knowledge to be acquired about another's internal state without any emotional experience on the part of the perceiver, it is unclear whether many aspects of "cognitive empathy" should even be discussed as empathy per se (e.g., Cuff et al., 2006). Such processes as facial emotion recognition might only make sense for inclusion within the broader

[15] For instance, the two different works that have caused particular concern in empathy research circles in recent years, both titled "Against Empathy" (Prinz, 2011, and Bloom, 2017a), never reference einfühlung, Rogers, Lipps, etc.

empathy construct by virtue of the fact that they interact with or regulate affective aspects of empathy (cf. Baron-Cohen & Wheelwright, 2004; Strayer, 1987), not that they are "empathic" on their own (far from a new idea, cf. Feshbach & Feshbach, 1969). Other researchers prefer to keep this kind of interpersonal accuracy separate from the empathy construct altogether, partly because such skills can be used for malignant purposes (e.g., conning and manipulating) rather than caring motivations primarily associated with empathy (Hall & Schwartz, 2019).

Many of the objections or concerns about including "cognitive empathy" facets within the empathy construct appear to be focused on ways of understanding others that are not inherently immersive and experiential, such as simple emotion recognition skills, basic theory of mind capacities, and detached forms of social inferential reasoning. Yet, "the process whereby one person 'feels her/himself into' the consciousness of another person" (Wispé, 1986, p. 34) has long been at the heart of the empathy construct. This particular way of "understanding" cannot be excluded from the construct. It must be given pride of place at the center.

Instead of understanding another in terms of abstract, propositional statements, this kind of empathic understanding means "to join, become immersed in, dwell on, feel into…, or step under or inside" the experience of another (Dekeyser, Elliott, & Leijssen, 2009, p. 117; also see Colman, 2009). Dekeyser, Elliott, and Leijssen (2009, p. 113) give an example of this in a therapist-client context: "As the therapist listened, he let himself be carried away into the client's experience: he felt a tickling, tingling sensation in his stomach; he remembered the feelings of his own successes…; he 'ran a movie in his head' of the client striding out of the kitchen, head held high, accompanied by a sense of pride and happiness." As argued by Rogers (1975) and others, this kind of understanding is not an invention reserved only for psychotherapy; it is a long-standing human capacity/tendency (though one which can be trained or cultivated). Perhaps

this kind of empathy could be referred to as "**einfühlung empathy**," to emphasize (a) its core meaning of immersively feeling oneself into the subjective consciousness of another and (b) its intellectual historical lineage.

Although this "einfühlung empathy" way of understanding others has been argued to produce more accurate knowledge of others (e.g., Rogers, 1959), its value and meaning is not ultimately bound to it having a resulting predictive accuracy higher than alternative ways of understanding others. Rather, it is a particularly "human"[16] and "humanizing" process of gaining understanding. As quoted in Buchheimer (1963, p. 66), Cartwright (the married name of Rosalind Dymond) pointed out that "people might be empathic without being able to predict and might be able to predict accurately without the process being an empathic one."[17]

It seems, though, that the centrality of "the imaginative transposing of oneself into the thinking, feeling, and acting of another" (Allport, 1961, p. 536) within the empathy construct has eroded dramatically, at least outside of clinical psychology. For example, consider the recent empathy review article by Coll et al. (2017), co-authored by a number of prominent empathy researchers. Coll et al. (p. 132) stated, "While there is as yet no consensus as to the precise definition of empathy, most researchers (at least in the field of cognitive neuroscience and psychology) agree that empathy involves the adoption of another's affective state so that both the empathizer and the empathic target…are in a similar state." The article repeatedly refered to this as the "standard definition" of the field.

---

[16] Which is not to deny that some elements of it, though, are present in other animals (e.g., de Waal, 2008). However, as suggested by Rochat (2002), it is reasonable to theorize that human beings acquire more diverse ways of empathizing than do other animals.

[17] Also see Cottrell (1950, p. 708): "Prediction is not necessarily based upon the kind of knowledge and understanding which comes from an alleged empathic response. Actuarial knowledge provides the basis for prediction without the kind of understanding which we assume results from empathy."

In Coll et al. (2017), understanding the other person is only regarded as relevant to empathy in terms of whether the emotional state of the other is accurately identified. The process of emotion identification is described purely in terms of emotion recognition, theory of mind, and inferential reasoning. There is nothing at all in the article that would lead a naïve reader to have any clue that empathy might, alternatively, be framed in terms of the imaginative experiencing of the conscious experiencing of another. Empathy is nothing more than "Identify the Emotion → Feel Roughly the Same Emotion" in this new "standard" view.

As suggested by Debes (2015), Edwards (2013), and Lanzoni (2018), the fate of the empathy construct has, at various points, been pulled in different directions depending upon which methodologies and movements were in vogue at the time, as well as depending upon the perceived limitations of different empirical measurement methods. Einfühlung fell into neglect when introspective methods lost popularity. Empathy from a Rogerian perspective was once in vogue, but, partially due its successful spread throughout psychotherapy circles, lost its distinctiveness as a "hot" field of research inquiry. The emergence of emotional empathy self-report questionnaires in the 1970s and 1980s as the dominant method of investigation appears to have been due to perceived failures of empathic ability tests (which significantly re-emerged in research in the 1990s; e.g., Ickes, Stinson, & Bissononnette, 1990; Nowicki & Duke, 1994; Stinson & Ickes, 1992) and failures of psychophysiological methods to attain specificity.

At present, neuroimaging appears to be leading the field of empathy research. For illustration, even after excluding a large number of fMRI studies of "empathy," Timmers et al. (2018) selected 128 (!) different recent functional neuroimaging whole-brain studies of "empathy" in response to stimuli, almost always aversive, which they meta-analyzed. The substantial majority of neuroimaging studies of empathy involve: presenting participants with

stimuli (such as an image of a hand or foot in a painful situation), typically instructing the participant to imagine the pain experienced by the other person or to share in it with the other person, comparing brain region activation in this condition with that in a neutral or baseline condition (e.g., passive viewing with no instruction), and frequently employing some kind of presumed moderator variable (e.g., race or gender of target).

As reviewed by Shamay-Tsoory and Lamm (2018), two seminal studies (Decety & Jackson, 2004; Singer et al., 2004) demonstrated that watching another experience pain activated much of the same "pain matrix" (especially the ACC and the anterior insula) that is activated by experiencing pain oneself. Following upon these studies, hundreds of additional studies have since been conducted replicating and exploring a wide range of moderators (e.g., race bias, psychopathology, administration of various chemical substances) of this kind of brain pattern activation (Shamay-Tsoory & Lamm, 2018). Although the external validity of this kind of methodology has been criticized (e.g., Schilbach et al., 2013) and more naturalistic neuroimaging designs for the study of empathy are emerging (e.g., Ashar, Andrews-Hanna, Dimidjian, & Wagner, 2017), one might conjecture that the rise of the "Identify the Emotion → Feel Roughly the Same Emotion" definition is partly due to its conformity with the basic methodology of much of the flood of empathy neuroimaging studies over the past 15 years.

"Einfühlung empathy," in contrast, is difficult to imagine being remotely as easily and specifically assessed in neuroimaging studies as is "pain empathy." More than that, it is difficult to envision a clear way to measure it *specifically* in any experimental paradigm; this difficulty, which extends throughout the history of the construct in psychology, has never been adequately surmounted. We can measure more observable *theorized* close correlates of the construct (e.g., empathic caring, self-reported perspective-taking, attentive listening, ability to reflect another's

experiences back to him/her in a way that he/she feels deeply understood.). The “einfühlung empathy” process itself, though, is likely to be stubbornly resistant to quantification given that: (a) aspects of it are not fully in conscious awareness; (b) it is inherently multi-faceted, recruiting a full range of cognitive, emotional, and sensory processes, much like dreams and daydreams; (c) it is not directly observable in outward behavior, and individuals have little way of comparing themselves with others in terms of frequency and intensity of such experiences; and (d) it is difficult to distinguish from projection.

Perhaps, however, if a large number of functional neuroimaging studies were conducted with subjects carefully instructed or guided to intentionally undertake this kind of “einfühlung empathy” approach in a wide variety of experimental designs, meta-analyses might reveal various brain activation correlates (and perhaps also be able to distinguish them from projection). At the moment, though, Thorndike’s dictum, "Whatever exists, exists in some quantity and can (in principle) be measured” (as cited in Cattell, 1965), hangs over this “einfühlung empathy” construct like an ominous cloud.

Affective empathy (caring vs. matching). In conceptualizing the “affective” aspects of empathy, many scholars define empathy as being characterized by caring for others, such as: “an other-oriented emotional response elicited by and congruent with the perceived welfare of someone else” (Batson et al., 2005, p. 486); “the capacities to resonate with another person’s emotions…and responding with the appropriate prosocial and helpful behaviour” (Oliveira-Silva & Goncalves, 2011, p. 201); and “a cognitive and emotional understanding of another’s experience, resulting in an emotional response that is congruent with a view that others are worthy of compassion and respect and have intrinsic worth” (Barnett & Mann, 2013, p. 230).

The *aim* of connecting empathy to caring for others is perhaps best illustrated by the empathy-altruism hypothesis promoted for decades by Batson and colleagues (e.g., Batson et al., 1981; Batson, Ahmad, Lishner, & Tsang, 2016). Like Schopenhauer, some researchers believe that human beings are capable of being *genuinely* altruistic, with at least some portion of our caring and helping behaviors motivated by the welfare of others, independently of or even against our own self-interests (no matter how subtle or unconscious). Yet, there are Nietzschean arguments against the existence of altruism at all turns, such as those discussed by Murphy, Watts, and Lilienfeld (2016) in theorizing how psychopathy might actually contribute to some seemingly self-sacrificially heroic behaviors, emerging out of egoistic "costly signaling" bids rather than other-oriented care.

If psychological altruism does exist, then what might generate it? Some, from Kohlberg (1976) to Kant (1785), have posited principled moral reasoning, rather than emotion. But the idea that reasoning itself leads to altruism has frequently been found implausible by others (e.g., Schopenhauer, 1840; Smith, 1759; also see Haidt & Joseph, 2004). Others have argued that altruism must emerge from internalized societal values (e.g., Staub, 1974), but one wonders whether this could truly explain the caring behaviors of all people whose altruism goes against the collective feelings of their particular communities (e.g., some Holocaust rescuers in Germany).

Instead, scholars like Schopenhauer (1840) and Batson, Ahmad, Lishner, and Tsang (2016) posit an emotional source for much (or all) of altruism, rooted in sympathy, empathy, or a related construct. Societal organizations, like the United Nations Virtual Reality Series, attempt to play to this empathy in order to increase feelings of caring. For many, myself included, this question of altruism is what draws one to the concept of empathy in the first place. It is perceived

(or hoped) to be a conceptual tool to be *used* for the purpose of helping explain the issue of altruistic caring, and that tends to become part of its *meaning*. The worry, of course, is that this ends up devolving into circular reasoning. Nonetheless, this overarching aim can likely partly explain why the term "empathy," for many psychologists and laymen alike, encompasses aspects of caring.

This other-oriented caring domain has frequently been referred to, broadly, as "empathic concern" (Davis, 1983; Batson, 1991; Hall & Schwartz, 2019). Of all non-cognitive aspects of empathy, it is by far the most commonly investigated in research (Hall & Schwartz, 2019. As pointed out by Hodges and Biswas-Diener (2007), out of all the proposed *emotional* aspects of empathy, this is the element that perhaps most corresponds to everyday use of the term by laypeople. Similarly, it has been observed to be a broadly agreed-upon aspect of empathy among clinical and counselling psychologists (Hatfield, Rapson, & Le, 2009). As Rogers (1975) argued, this kind of caring for others may be strongly entangled with the tendency to empathize oneself into the experiential perspectives of others, such that a certain degree of caring is necessary for empathic understanding to fully flower. It could also be argued that the process of empathic understanding inherently tends to lead one to care about the other, at least to some degree.[18]

Nonetheless, it must be admitted that stepping into the experiencing of another does not automatically relate to caring for that person. For instance, on some occasions, individuals take pleasure in the pain of others, and the increasing "reality" of that pain (at least up to some point) may increase the emotional satisfaction of the observer (Heilbrun, 1982). For example, Wang, Lilienfeld, and Rochat (2019) have described different aspects of "schadenfreude," such as: taking pleasure in the pain of those we perceive as personal rivals; taking pleasure in the pain of

[18] As the comedian Richard Pryor once mused: "It's easy to love somebody. That's all you got to do. Sit with 'em a little while. Talk to 'em" (Pryor, 1983).

those whom we deem deserving of punishment, fuel for a sense of retributive justice; and taking pleasure in the pain of those perceived as out-group members. Without at least a minimal capacity to identify the emotional states of others and be emotionally affected by the emotions of others, schadenfreude would be blocked. In this sense, some degree of "empathy" appears to be necessary for schadenfreude, at least depending upon how one defines the construct of empathy.

Wang et al. (2019, p. 7), however, argued that schadenfreude is enabled by "dehumanization," which is "the process by which a person or social group is perceived as lacking attributes that define what it means to be human (i.e., humanness)." Wang et al. argued that a core aspect of the empathy construct, *motivation* to perceive the minds of others, is reduced when dehumanizing others; to the extent that motivation to perceive the minds of others increases, those others will be increasingly "humanized" and schadenfreude will slacken. In a converging sense, then, the core of "einfühlung empathy" may tend to obstruct schadenfreude, even though related facets of the broader empathy construct may theoretically facilitate schadenfreude. When einfühlung empathy is extensive, even the most disapproved individuals may come to be regarded with some degree of compassion.[19]

Similarly, the extent to which one steps into the perspective of another appears to be inversely related to the "fundamental attribution error" (term introduced by Ross, 1977), the tendency to overlook situational influences when evaluating the behavior of others; this error indirectly leads us to exaggerate dispositional influences on such behavior. As suggested by some lines of research (e.g., Hooper et al., 2015; Joireman, 2014; Regan & Totten, 1975), this

[19] On an autobiographical note, my curiosity about the empathy construct first clearly emerged in the mid 2000s, while living and working in New Orleans with Catholic monastics who spent much of their time building relationships with death row inmates. In particular, Sister Lilian Flavin was perhaps the most "empathic" individual I had ever encountered, with a tremendous capacity for generous understanding, which was subtly contagious when she spoke of how life might feel like to others, whether prisoners, impoverished families, victims of police brutality, etc. It is difficult for me to imagine her experiencing schadenfreude.

error is ameliorated by empathic perspective-taking. Thus, when presented with individuals with whom we disagree or whom we perceive as committing moral violations, einfühlung empathy may tend to reduce our tendency to attribute their behavior to dispositional qualities and, instead, allow us to perceive situational forces that may have led them to think, feel, and act in certain ways (and "feel" the influence they might have had).

In most investigations, the broad domain of "empathic concern" is measured and interpreted as a single construct. As discussed further below, for instance, self-report scales of empathy generally lump all elements of this domain into a single scale, such as the Empathic Concern scale of the Interpersonal Reactivity Index (Davis, 1983). It is possible, however, that a number of meaningfully distinguishable facets of empathic caring are contained within this broad construct, and that discussing and measuring them separately could offer incremental pragmatic value. For instance, potential facets include:

- Caring for the positive emotions of others, such as wanting to bring joy to others
- Caring for the negative emotions of others, such as wanting to alleviate another's distress
- Empathic restraint, wanting to avoid causing distress to others
- General attitude towards empathic caring, holding empathic caring as a personal value in oneself and others

Other theorists extend the emotional response of empathy even more broadly than just caring or concern, encompassing general "appropriateness": e.g., empathy "…entails an appropriate affective response in the observer to the other person's mental state" (Baron-Cohen & Wheelwright, 2004, p. 168) or "…requires individuals to choose the best socioemotional response (e.g., by soothing a sad person without being as sad as this person)" (Carre et al., 2013,

p. 680). For example, in Baron-Cohen and Wheelwright's (2004) Empathy Quotient questionnaire, items related to empathic concern are accompanied by items related to "social skills," such as "I find it hard to know what to do in a social situation" and "I often find it difficult to judge if something is rude or polite." Though this kind of conceptualization resembles that of Hogan's (1969, 1975) "role-theoretical" perspective on empathy, encompassing social tact and skill, most scholars do not extend the empathy construct to include such aspects.

The narrow "isomorphic matching" definition. Especially in recent years, many scholars have moved to a definition of empathy that explicitly excludes compassion and caring from the empathy construct, essentially intensifying the "matching" definition held by Feshbach (1975). Interestingly, though, Feshbach and Feshbach (2009, p. 85, emphasis added) appear to have substantially softened the "matching" theory of Feshbach (1975), broadening the definition to require only "some degree of correspondence" in shared affect, resulting in the "affective ability to experience emotions in an *appropriate manner*."

Nonetheless, the "matching" definition of affective empathy is gaining ascendancy while seeming to also tighten. As a representative example, de Vignemont and Singer (2006, p. 435, emphasis added) defined empathy as follows: "(i) one is in an affective state; (ii) this state is *isomorphic* to another person's affective state; (iii) this state is elicited by the observation or imagination of another person's affective state; (iv) one knows that the other person is the source of one's own affective state." This kind of definition appears to be quickly gaining popularity (e.g., Bloom, 2017b; Lamm, Rutgen, & Wagner, 2019; Lockwood, 2016; Preckel, Kanske, & Singer, 2016), and, as discussed earlier, some now regard it as the "standard definition" in psychology (Coll et al., 2017).

Although some have posited a distinction between empathic matching for positive and negative emotions (e.g., Telle & Pfister, 2016; Sallquist, Eisenberg, Spinrad, Eggum, & Gaertner, 2009), most discussions of the narrow matching construct of empathy treat it as a single construct. In particular, existing self-report scales combine contagion for positive emotions and contagion for negative emotions into a single measurement dimension. Moreover, as pointed out by Telle and Pfister (2016), most research in line with this construct has focused only on empathy for negative emotions.

As Cuff et al. (2017) pointed out, though, individuals may have quite different capacities (or tendencies) in regard to sharing the positive and negative emotions of others; a person could be highly susceptible to positive contagion, but not for negative contagion, or vice versa. Murphy et al. (2018b) further argued that, as broad trait domains, positive emotionality and negative emotionality are largely orthogonal and, as has been pointed out frequently over the centuries (Smith, 1759; Stein, 1917), the emotional tendencies of the observer moderate the degree to which he or she will tend to sympathize/einfühlen with specific emotions. Moreover, whereas appetitive contagion will theoretically tend to motivate individuals to approach others, aversive contagion will likely tend to cause them to avoid one another (Murphy et al. 2018b; cf. Nietzsche, 1880). Finally, "presuming that susceptibility to emotionality in oneself is correlated with contagious susceptibility to emotionality of the same valence in others, contagion for negative emotions might be correlated with chronic emotional distress, whereas contagion for positive emotions might be associated with well-being" (Murphy et al. 2018b, p. 2).

Consistent with these predictions, Murphy et al. (2018b) factor analysed an existing empathic contagion scale (the EI; Jordan, Amir, & Bloom, 2016) and observed that, unbeknownst to the scale designers, appetitive contagion and aversive contagion items formed

separate factors. Furthermore, consistent with their predictions, they observed that appetitive contagion and aversive contagion generally demonstrated opposing relationships with negative emotionality, emotional distress, adult attachment, and personality disorder features. Due to the limited number of appetitive contagion items present, however, the EI is not suitable for producing reliable and robust scales of these two constructs.

Beyond this issue of parsing empathic matching into multiple distinguishable domains, many deeper questions and concerns have been raised regarding this narrow definition of empathy. For instance, in requiring a response that is "identical or very similar to the other's emotion" (definition in Eisenberg, Fabes, & Spinrad, 2006, p. 647), just how close of a match must there be between the emotions experienced by the two parties? Scholars (e.g., Batson, 2009; Iannotti, 1975) have repeatedly noted that this is never made clear by those embracing this restricted definition.

For an analogy, consider matches between colors. In order to consider two colors as "matching," do they need to be largely identical? Or is it "very similar" enough that they simply be roughly in the same domains of the basic and secondary colors (red, blue, and yellow; orange, green, and purple)? Similarly, if a person is experiencing a blend of anger, pain, and fear, must an observer feel that same *mixture* of emotions? Or is it enough if he/she simply experiences some rough approximation of only anger? Depending upon how close of a "match" is required, in this view, empathy could be extremely rare or extremely common.

Moreover, what is the justification for specifying that there be a match in the first place? As Wondra and Ellsworth (2015, p. 412) skeptically wondered, "are vicarious emotions that match the target's feelings and those that do not match so different?" Is the main psychological process driving isomorphic emotion matching critically different than that which produces non-

matching emotional resonance? Under the view that emotions are heavily generated by appraisal (cf. Smith, 1759), Wondra and Ellsworth (2015) concluded that there is little justification for deeming matching and non-matching vicarious emotions as fundamentally distinct.

A defender of this special status of matching emotions might argue that empathy can be tied, in restricted fashion, to only processes originating in physical mimicry and emotional contagion. This, however, would mean that empathy will generally not occur outside of situations where a person is immediately confronted with an expressed emotional state of another (Wondra & Ellsworth, 2015). Yet, some of the main proponents of this restricted "matching" view of empathy explicitly extend empathy beyond what is directly observed, such as to what one imagines another might feel, even if that person is not present (e.g., Lockwood, 2016). Bloom (2017b, p. 25), for instance, in defining empathy as isomorphic matching, tied it to inferring (or appraisal) and specifically delineates it from emotional contagion: "Unlike in emotion contagion, this person does not have to be present, or even exist – we can have this feeling toward fictional characters." This conceptual tweaking appears to emerge out of a desire to allow this narrow definition of "empathy" to occupy conceptual territory that is not already claimed for the basic process of emotional contagion.

If empathy is not limited to mimicry or emotional contagion, then it is unclear why we should consider matching and non-matching vicarious emotions as emerging from importantly distinguishable processes. Wondra and Ellsworth (2015, p. 415) offered a wonderful illustration:

> Consider our emotional reactions to horror films. You can feel scared for characters who know that a murderer is stalking them in their home, but you can also feel scared for characters who are clueless about the murderer's presence. Is the cause of your fear very different in these two cases? Does a new empathy-generating process take over from some other process once the clueless characters notice the murderer and become scared too?

Moreover, if we understand empathy as being based in "other-mindedness," as the experiencing of a foreign consciousness as *different* from one's own (e.g., Stein, 1917), then the isomorphic matching definition runs the risk of undermining this other-oriented intentionality (and perhaps the self-other distinction, discussed below). As Zahavi and Rochat (2015, p. 544, emphasis added) pointed out:

> To insist that the empathizer must have the same (kind of) state as the target, is to miss what is distinctive about empathy, namely the fact that it is a special form of other-directed intentionality, one that allows the other's experiences to disclose themselves as other rather than as own…Rather, to empathically experience, say, the emotion of another *necessarily differs* from the way you would experience the emotion if it were your own.

Finally, the logic of this narrow emotion-matching definition of empathy can appear to break down when subjected to reductio ad absurdum. For example, as pointed out by Hall and Schwartz (2019), would we call it "empathy" if a person feels murderous rage after observing a person feeling murderous rage? Or, as Nietzsche ridiculed, what would we call it if we encountered a self-hating person and, in isomorphic matching, hated him too? If a therapy client comes to an appointment in a delusional euphoric state, would a therapist only be empathizing if he/she also experienced euphoria?

These reductio ad absurdum arguments point to a more basic problem with the narrow matching definition. As Keltner and Haidt (1999) and Keltner and Kring (1998) have argued, the rapid interpersonal transmission of emotions between individuals can be understood as based in evolved social functionality. As Keltner and Kring (1998, pg. 326) explained, "Emotions provide information about interacting individuals' emotions, intentions, and relational orientations. Emotions evoke complementary and similar emotions in others that motivate behaviors that

benefit social relationships. The perception of emotion and anticipated elicitation of emotions in others serve as incentives for certain social behaviors."

Keltner and Kring (1998) noted that, although research attention has been heavily directed toward instances in which emotional responses are isomorphic to that of another person, for many emotions the spontaneous functional response will tend to be complementary rather than isomorphic (e.g., feeling fear in rapid response to seeing another's anger expression). This is similar to Smith's (1759) discussion of differences in spontaneous sympathetic matching for anger as opposed to sadness.

If rapid emotional resonance is understood to have evolved not primarily for direct mirroring but primarily for functional responding, then narrowing empathy to only isomorphic matching seems even more lacking in justification. Not only does the mirroring requirement separate empathy from compassionate caring; it also seems to undercut the "social skills" views of empathy held by people like Hogan (1969) and Baron-Cohen and Wheelwright (2004). It is precisely this mirroring requirement that allows scholars like Bloom (2017a) to come out against empathy, expounding upon its maladaptivity. Perhaps the mother who is calm when hearing her baby cry no longer needs to be seen as "unempathic" in comparison with the mother who starts crying herself…

Taking this functionality critique further, consider the situation of psychotherapy, which was the motivating context for much of the study of empathy for decades. If emotion-matching is required for empathy, then it may tend to *remove it* as a functional component of therapist behaviour. From its earliest major uses as a concept in therapeutic practice, empathy has often been associated with refraining from experiencing the same emotions as a client (e.g., "the counselor is perceiving the hates and hopes and fears of the client through immersion in an

empathic process, but without himself, as counselor, experiencing those hates and hopes and fears," Rogers, 1949, p. 86), or at least strongly regulating such matching so that the therapist is capable of remaining fully present and attentive to the client's inner life.

It would, of course, be unfair to say that the proponents of the isomorphic matching conceptualization are "against" compassionate caring in any way (e.g., Bloom, 2017a). In defining empathy as an isomorphic mirroring of another person's emotional state, for instance, Preckel, Kanske, and Singer (2018, also see Singer & Klimecki, 2014) described compassion and empathic distress as different potential *outcomes* of empathy; compassion, characterized by helpful complementary rather than isomorphic emotional responses; and empathic distress, characterized by detrimental isomorphic emotion matching. Others have similarly posited compassion as a potential outcome of this kind of isomorphic empathy (e.g., Marsh, 2018; Christov-Moore & Iacoboni, 2014).

As mentioned by Marsh (2018, p. 112), though, "one as-yet unaddressed puzzle is how the negative social inputs encoded during emotional empathy for other's suffering are transformed into the positive prosocial motivation that drives care and altruism in response." Put differently, what is the internal process that determines whether an initial instance of isomorphic emotional contagion leads to empathic distress, compassion, or another type of response (e.g., empathic joy, schadenfreude, or apathy)? This is, in some ways, an alternative restatement of the classic question posed by Schopenhauer and others: how is that the welfare of another becomes a motive for me, much as my own welfare is usually my motive?

Whatever this undefined process might be, it would appear to be *critical* to understanding and conceptualizing the broader empathy construct. As pointed out on many occasions, the difference between empathic distress (sometimes referred to as "personal distress") and

compassion (often referred to as “empathic concern”) is that the latter is other-oriented and the former is self-oriented (e.g., Batson, 2009).

If we are faithful to the idea that empathy requires an other-focused intentionality, rather than an I-focused intentionality, then empathic caring appears to be a *continuation* of empathy, whereas empathic distress is a termination of empathy. Why not, then, extend the term “empathy” to include the unknown mechanism that converts emotional resonance into compassion, rather than distress, and keeps the empathic encounter moving forward?

In many accounts, such as that of Preckel, Kanske, and Singer (2018), this transition/conversion process is not labeled or elaborated. Christov-Moore and Iacoboni (2014) speculated that it is mediated by higher-order cognitive processes. Bloom (2017b) hypothesized that compassion results due to reduced activation of neural systems responsible for mirroring the experiences of others. Many others have attributed it generally to emotional regulation (e.g., Decety & Jackson, 2004) or to emotional over-arousal (e.g., Eisenberg & Fabes, 1992). On closer examination, these offerings do little to actually present an explanation; there is nothing inherent in emotion regulation, lack of over-arousal, or higher-order cognitive processes that would tend to convert the welfare of another into the motivating goal.

Marsh (2018), however, speculated that it may rather be driven by neurobiological structures evolved for caregiving. In other words, whatever it is that makes a mother calmly comfort her crying infant rather than burst into tears herself transforms emotion-matching into other-oriented caring rather than self-oriented distress.

Yet, many scholars have argued that the broader domain of empathy can be understood as evolving out of the mammalian caregiving system (e.g., de Waal, 2008; Preston, 2013; Stern & Cassidy, 2018). In these accounts, the capacity for empathy emerged out of the need to build

emotional bonds with kin and facilitate caring towards them, and the underlying evolved mechanisms came to also serve similar purposes within broader groups of non-related or lesser-related individuals.

This position has two particularly important implications. First, if this kind of account is true, then whatever mechanism transitions emotional resonance into compassion might be reasonably characterized as an extension of the evolved functionality of empathy in general. If so, then there would seem to be little justification for excluding other-oriented non-matching emotions from the empathy construct.

Second, if empathy should be properly understood as emerging out of caregiving, then perhaps the emotional aspects of empathy would be better conceived as related to bonding and relationship-building more than to prosociality or general "caring" per se (cf. Main, Walle, Kho, & Halpern, 2017). Einfühlung empathy, for instance, may be a tendency that has evolved in order to "glue" individuals together in a non-transactional fashion, allowing for increased emotional intimacy and attachment. The broader ripple effects of empathy, such as towards acquaintances and even strangers, may stem from a weak but valuable sense in which empathy causes us to feel an inkling of emotional "relationship" to them.

If empathy is more about emotional relationship to others than about pity or compassion, then empathy research might need to be adjusted quite a bit, particularly in self-report questionnaires, which rarely have significant content directly related to emotional intimacy or relationship tendencies. It may be that empathy is more associated with being a "safe" and supportive person to whom other people feel comfortable opening up to about their vulnerable hopes and fears, a "nurturing" person who helps other feel seen and understood and valued, than with being a person who gives the most to local charities or feels the most emotion when seeing

another person in pain. In some ways, this would simply constitute a return to more Rogerian (or proto-Rogerian) conceptualizations of empathy, which have been generally neglected in self-report empathy questionnaires.

Self-other distinction. One aspect of the definition of empathy that has been relatively stable as a consensus view among empathy researchers, since at least the 1950s until the present day, is the self-other distinction (e.g., Coplan, 2011; Decety, 2015; Decety & Jackson, 2004; Eisenberg & Eggum, 2009; Eisenberg & Strayer, 1987; Singer & Steinbeis, 2009; but see de Waal & Preston, 2017, whose model of empathy extends across mammalian species and does not support the self-other distinction as a requirement). Although this is phrased differently across sources, Decety and Meyer (2008, p. 1053, emphasis added) presented a simple and representative example: "The psychological construct of empathy refers to an intersubjective induction process by which positive and negative emotions are shared, *without losing sight of whose feelings belong to whom*."

The extent to which persons must keep in sight the source of their feelings in the other person, though, has yet to be clearly specified. Must our conscious attention be focused on the distinction, or can it simply weakly exist on the periphery of our conscious attention? Must the self-other awareness be strong and constant, or can it be more modest and fluctuating? In our daily lives, our feelings are continually being affected by resonance with those around us, but the degree to which we are consciously aware of this contagion/mirroring/responding usually floats somewhere in a gray area between full awareness and total lack of awareness.

This self-other distinction as a component of empathy may have emerged from multiple sources in the intellectual history of the construct, but it seems to have certainly been pushed by those seeking to differentiate empathy from projection (e.g., Dymond, 1950) and by

psychotherapists aiming to strike an emotional balance between alienation and dysregulated over-identifcation (e.g., Rogers, 1949). Beyond that, though, it extends back into the 18th and 19th centuries in moral philosophy (e.g., Schopenhauer, 1840).

Though this distinction is important both in regard to general empathic understanding and also general empathic caring, it is a particularly important concern in terms of the narrow isomorphic matching construct that some currently consider the entirety of affective empathy (e.g., Bloom, 2017b). For example, Singer and Klimecki (2014, p. R875, also see Decety & Lamm, 2006; Eisenberg & Eggum, 2009; Gerdes, Segal, & Lietz, 2010) noted that "if this self–other distinction is not present, we speak of emotion contagion, a precursor of empathy that is already present in babies." Although some argue that isomorphic emotional contagion should simply be included as "empathy" (e.g., Darwall, 1998), doing so would tend to leave very little distinctive territory left for isomorphic matching alone to be called "empathy."

Psychologically speaking, however, these two requirements, the matching requirement and the self-other distinction requirement, may be in conflict, especially at a trait level. As an individual becomes more aware that his/her anxiety/frustration/sadness is coming from an external source, and focuses more of his/her attention on this distinction, the more easily the person will be able to engage in cognitive appraisal to shift his/her emotional state to a non-matching state. In other words, as self-other awareness rises, isomorphic matching may tend to stall out in intensity and/or duration.

For what might be a relevant example, consider the functional neuroimaging study of Jackson, Brunet, Metlzoff, and Decety (2006), who showed research participants images of hands and feet in painful situations, and instructed them to take the perspective that it was either their own hand/foot or the hand/foot of a specific (but unfamiliar) person. Interestingly, though

both conditions activated brain regions associated with the processing of pain, the "self" condition elicited generally greater activation than the "other" condition, as well as higher perceived pain ratings. Specifically, in reference to the right anterior insula, subsequently meta-analytically identified by Timmer et al. (2018) as a distinguishing correlate of "pain empathy" in this kind of paradigm, the "self" condition elicited significantly higher activation than the "other condition." In other words, this key brain region for "matching empathy" for pain appears to be *reduced* in activation as the self-other distinction increases.

To my knowledge, no measure of empathy exists that attempts to quantitatively assess individuals for differences in regard to this self-other distinction, at either a state or trait level. As Hall and Schwartz (2019) pointed out, this theoretical issue is certainly in need of empirical study, although measuring it may be difficult in practice. Thorndike's dictum comes back into view again.

Neglected domain: empathic approach-avoidance. Empathy is not necessarily automatic. Motivation plays a large role (Weisz & Zaki, 2018, p. 70): "Empathy is a social bridge that allows people to connect with each other, but it is not an automatic response to others' suffering. Instead, it is a motivated phenomenon, reflecting approach and avoidance motives that encourage people to engage in or avoid it." In fact, it is likely that motivation to approach or avoid empathic encounters may be the single best predictor of the frequency and/or intensity of an individual's empathic experiences.

In addition to regulating their emotional responses when perceiving the emotions of others (such as in Rogerian empathy), individuals are also able to regulate how often and how intensely they enter into empathic encounters in their daily lives. Although this aspect of empathy, at both the state and trait level, is at least to some degree implicit in many discussions

and operationalizations of empathy, Weisz and Zaki (2018; also see Zaki, 2014) have argued for a more explicit focus on it as a domain of the empathy construct. In their treatment, which is focused mostly on neuroscience investigations of empathy, they lay out largely state-related aspects of empathic approach-avoidance:

> At least two types of motives interact with context to facilitate or inhibit empathy. *Avoidance motives* drive people to feel *less* empathy. For instance, people are motivated to avoid empathy if it will lead to costly helping, if it will be exhausting, if it interferes with obtaining a desired outcome, like during zero-sum competition. Conversely, *approach motives* encourage people to feel *more* empathy. People are often motivated to empathize more when they want to share others' positive states, when empathy is socially desirable, or when empathy strengthens their social ties (Weisz & Zaki, 2018, p. 67).

Although these state-focused motives certainly depend upon situational contexts, the kinds of motives described by Weisz and Zaki may also generate a trait-like consistency of empathic approach-avoidance.

Paralleling the process model of emotional regulation described by Gross (1998), empathy can be approached or avoided both antecedent to a full emotional empathic response as well as subsequent to the emergence of that response (also cf. Gross & John, 2003). Zaki (2014) provided an excellent account of the various stages at which empathy can be approached or avoided. Individuals are able to choose, at least in many cases, whether they will enter into empathic encounters (*situation selection*; e.g., go talk to the grieving co-worker or not). When an individual does enter into an empathic encounter, whether voluntarily or involuntarily, he or she can also choose to change it into a different kind of encounter (*situation modification*; e.g., try to change the subject, crack jokes) or to divert attention away from its salience (*attentional deployment*; e.g., "zone out" or start thinking about a grocery list). Similarly, an individual can engage in *cognitive reappraisal* to diminish the empathic salience of the encounter, such as focusing on placing blame on people for their own troubles orfocusing on the "bright side" (e.g.,

"your husband might have run off, but you still have the kids!"). Alternatively, an individual can choose to *suppress* their emotional response to the empathic encounter. In sum, there are a myriad of ways in which individuals can engage in approach or avoidance behaviors in relation to emotional empathy (Zaki, 2014), much as with any other kind of emotional process (Gross & John, 2003).

Despite the seemingly central importance of this approach-avoidance domain, empathy research rarely assesses it (but see approach-avoidance in studies conducted by Batson and colleagues; e.g., Shaw, Batson, & Todd, 1994). In particular, although there are scattered instances of relevant items in some empathy questionnaires, no extant self-report empathy questionnaire aims to specifically capture this domain. This dissertation aimed to remedy this deficit.

Current Self Report Measures and Potential Areas of Improvement

"If we want to know how people feel: what they experience and what they remember, what their emotions and motives are like, and the reasons for acting as they do—why not ask them?" (Allport, 1942, p. 37)

The quality of empathy research hinges on the construct validity, conceptual clarity, and psychometric characteristics of the measures used to assess aspects of the broader construct. Although a range of behavioral tasks are available to investigate cognitive empathy ability (e.g., Dyck, 2012; Dziobek et al., 2006; Nowicki & Duke, 1994), as well as a few to investigate affective empathy in a limited fashion (e.g., Dziobek et al., 2008), self-report questionnaires are *by far* the most common method of measurement in research (Hall & Schwartz, 2019).

Paulhus and Vazire (2006), in a review of self-report data, provided a number of arguments for the value of such reports. For instance, individuals have dramatically greater quantity and breadth of information about themselves than others have about them; this is especially the case in regard to feelings and behaviors that are private. Similarly, individuals have access to internal thoughts and feelings that are not clearly available to others, or are commonly restrained in outward display (e.g., anxiety, shame, Machiavellian intentions). Furthermore, individuals are more reliably interested in themselves than in others, and, therefore, would seem to have a greater motivation to respond effortfully.

Particularly in regard to self-reporting of competencies or abilities, however, many concerns have been raised as to whether individuals can accurately rate their own abilities (e.g., Keefer, 2015; Zell & Krizan, 2014). As argued by Zell and Krizan (2014) in a meta-synthesis of prior meta-analyses, the self-report validity of an ability domain strongly depends on whether individuals generally receive clear and consistent feedback about their performance in that particular domain. Unfortunately, "nonverbal skill is a domain in which people receive little to no direct feedback. Although people often make judgments about whether others are telling the truth or lying from nonverbal cues (e.g., eye gaze), they rarely receive objective feedback indicating whether others are actually telling the truth or lying… This lack of feedback may lead people to remain ignorant of their true nonverbal skill for most of their lives" (, p. 117).

Unsurprisingly, as demonstrated by Murphy and Lilienfeld (2019), self-reported ratings of cognitive empathy *ability* lack adequate validity evidence to support their use as proxies for actual cognitive empathy ability. A number of empathy questionnaires, especially those developed since 2004, contain "cognitive empathy" scales composed of items asking individuals to rate their own abilities at recognizing and understanding the thoughts and feelings of others

(e.g., EQ: Baron-Cohen & Wheelwright, 2004; BES: Joliffe & Farrington, 2006; ACME: Vachon & Lynam, 2016; QCAE: Reniers et al, 2011; ESE: Olderbak et al., 2014). Sample items include "I can easily work out what another person might want to talk about" (QCAE Perspective-Taking scale), "I can usually tell how people are feeling" (ACME Cognitive Empathy scale), and "I can often understand how people are feeling even before they tell me" (BES Cognitive Empathy scale).

Although most studies reporting on cognitive empathy ability rely upon such self-report measures (cf. Hall & Schwartz, 2019), a number of theorists have argued that individuals cannot validly self-report their own trait abilities to recognize or infer the thoughts and feelings of others (e.g., Davis & Kraus, 1997; Ickes, 1993; Realo et al., 2003). Despite this warning having emerged from findings generations ago (e.g., Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979; Zuckerman & Larrance, 1979), many empathy researchers appear to have overlooked this danger to the validity of self-report measures of cognitive empathy ability as proxies for behaviorally assessed ability. As suggested by Murphy and Lilienfeld (2019), this may be due to such caveats frequently being couched in terms such as "social perceptual confidence," "general nonverbal decoding ability," and "interpersonal sensitivity," rather than explicitly in terms of "cognitive empathy."

Moreover, limited data indicate that self-reported cognitive empathy is positively correlated with narcissism (Ames & Kammrath, 2004; Ammirati, 2013) and self-reported "sense of power" (Catterson, Naumann, & John, 2015), suggesting that self-report cognitive empathy scores may be more associated with social self-esteem, overconfidence, or other related variables, than to actual cognitive empathy ability. Furthermore, evidence tentatively suggests that the domain of self-reported cognitive empathy ability may be undermined by the well-

documented Dunning-Kruger effect (cf. Kruger & Dunning, 1999), with poor performers dramatically overestimating their CE abilities and high performers often (although less dramatically) underestimating their CE abilities (Ames & Kammrath, 2004; Ammirati, 2013).

Murphy and Lilienfeld (2019) meta-analyzed the relationships between self-report scales measuring cognitive empathy ability (including scales not explicitly labeled as "cognitive empathy"), on the one hand, and behavioral measurements of cognitive empathy ability, on the other. Our meta-analytic results, based on random effects models, from 85 studies (total N = 14,327) indicated that self-report cognitive empathy ability scores account for only approximately 1% of the variance in behavioral cognitive empathy assessments (meta-analytic effect size estimate, r = .09). Moreover, this relationship was not significantly different from that demonstrated by affective empathy scores (r = .12). Murphy and Lilienfeld also observed that effect sizes were not moderated by self-report empathy domain, gender composition, unisensory versus multisensory behavioural stimuli presentation, child versus adult samples, or normative versus clinical/forensic nature of the sample.

The meta-analytic results observed by Murphy and Lilienfeld (2019) were consistent with similar findings from much smaller meta-analyses, relying heavily upon older generations of research instruments, conducted by Davis and Kraus (1997) and Hall, Andrzejewski, and Yopchick (2009). As discussed by Murphy and Lilienfeld, though, the behavioral assessments of cognitive empathy have severe limitations themselves and often do not correlate highly or even moderately with one another (see meta-analysis by Schlegel, Boone, & Hall, 2017). Nonetheless, Murphy and Lilienfeld (, p. 7) "caution researchers and practitioners against substantial reliance on self-report CE scale scores as proxies for CE abilities, at least until and unless adequate

evidence is adduced demonstrating their validity for this purpose." As a result, I avoided an attempt to craft such a scale in this dissertation work.

Outside of the domain of self-report cognitive empathy ability, the most obvious concern with empathy scales is that self-report measures of empathic caring and empathic tendencies will be heavily contaminated by individual differences in the tendency to want to look good in front of others, or perhaps more importantly, to feel good about oneself. As Batson et al. (2016, p. 488) argued, "attempts to measure these differences by standard retrospective self-report questionnaires seem suspect at best. Such questionnaires are more likely to reveal the degree of desire to see oneself and to be seen by others as empathic rather than to provide a valid measure of one's proclivity to be empathic."

This worry extends to any self-report rating method of highly evaluative content, such as antagonism, general personality traits (especially agreeableness and conscientiousness), and personality disorder features. Although a number of studies and reviews have indicated that the fear of contamination from such response biases has been exaggerated in at least some contexts (e.g., Chan, 2009; McCrae et al., 1998; McCrae & Costa, 1983; McGrath, 2010; Watts et al., 2016), others indicate genuine overall reason for concern (e.g., Konstabel, Aavik, & Allik, 2006).

At the same time, observer reports of empathy may often have their own serious limitations (e.g., Murphy, 2018, 2019), physiological measures of empathy may still lack specificity (e.g., as noted by Cassidy & Stern, 2018), and the few behavioral measures available (e.g., Dziobek et al., 2008) are so limited in what they measure that they likely have the same concerns leveled at behavioral measures of cognitive empathy in the absence of aggregation (cf. Epstein, 1979). In other words, despite serious limitations, self-report measures may still be the

"least worst" currently available approach for assessing trait-like empathic feelings and tendencies (though not abilities).

Especially given the strong evaluative content of empathy items, I presume that empathy self-reports and other evaluative self-report measures at least modestly reflect this kind of bias (and this may be exacerbated by "satisficing" tendencies in questionnaire responding, as discussed further below in the section on M-Turk sample issues). As described below in the section on bifactor models, though, this relates to an additional potential benefit of employing bifactor modeling with accompanying incremental validity approaches. As will be explained further: any *systematic* evaluative response bias will tend to be encompassed within the general factor of an empathy item pool. As a result, the incremental validity of specific content empathy domains, above and beyond this general factor, will theoretically be substantially cleansed of substantial systematic response bias (in addition to reflecting what is unique to the specific content domains rather than shared with the item pool as a whole).

Despite these concerns, self-report measures of affective empathy appear to demonstrate meaningful convergent validity with measures of many other constructs associated with empathy deficits (e.g., Baron-Cohen & Wheelwright, 2004; Vachon & Lynam, 2016). They have been frequently observed to predict charitable donations (Davis, 1983; Jordan, Amir, & Bloom, 2016; Verhaert & Van den Poel, 2011) and *may* also substantially correlate with performance in behavioral affective empathy tasks (e.g., Dziobek et al., 2008). Still, the available evidence for external validity is sparse due to a dearth of prior investigations[20], and generally focuses on very limited contexts, such as willingness to donate a small amount of money to a charitable cause.

[20] Per personal communication with Dr. Hall, it seems that there may not currently be enough available data to conduct a meaningful meta-analysis, for instance.

The Interpersonal Reactivity Index (IRI: Davis, 1983) has long been the most commonly employed self-report empathy questionnaire for adult participants, but a wide variety of alternative adult measures also exist, such as the Questionnaire Measure of Emotional Empathy (QMEE: Mehrabian & Epstein, 1972), the Balanced Emotional Empathy Scale (BEES: Mehrabian, 1996), the Empathy Quotient (EQ: Baron-Cohen & Wheelwright, 2004), the Basic Empathy Scale (BES: Joliffe & Farrington, 2006), the Toronto Empathy Questionnaire (TEQ: Spreng, McKinnon, Mar, & Levine, 2009), the Empathy Assessment Index (EAI: Lietz et al, 2011), the Questionnaire of Cognitive and Affective Empathy (QCAE: Reniers et al., 2011), the Emotion Specific Empathy questionnaire (ESE: Olderbak, Sassenrath, Keller, & Wilhelm, 2014), the Affective and Cognitive Measure of Empathy (ACME: Vachon & Lynam, 2016), the Empathy Index (EI: Jordan, Amir, & Bloom, 2016), and the Empathy Components Questionnaire (ECQ: Batchelder, Brosnan, & Ashwin, 2017). Most of these questionnaires are primarily composed of items from other measures, so the total number of individual items across measures is much smaller than the number of different questionnaires might suggest.

Despite many attempts to craft valuable empathy questionnaires, there remains substantial room for improvement. First and foremost, a number of potentially distinguishable and valuable emotional empathic tendency constructs have been largely overlooked across the sweep of measures. Some affective empathy scales include one or two items that could potentially emerge as distinguishable and valuable content domains of their own if an adequate number of other similar items were included in factor analyses (e.g. empathic restraint/repair items in the ACME Affective Resonance scale). Some scales could potentially be shown to comprise subcomponents with very different nomological networks. For example, unidimensional emotional contagion scales should likely be parsed into contagion for aversive

and appetitive emotions, as limited evidence provided by Murphy et al. (2018b) indicates that these domains, though substantially positively correlated, demonstrate sharply diverging nomological networks.

Other scales that have been heavily criticized on validity grounds may have the potential to emerge as more valuable components of the empathy construct if items were revised or replaced. For instance, the Fantasy scale of the IRI is meant to assess immersive interest in the stories of others, but almost all the items relate to books or movies; the scale may be more valuable if the items are revised to reference actual person-to-person sharing. Finally, one theoretically important domain (cf. Weisz & Zaki, 2018; Zaki, 2014) is almost entirely ignored: individual differences in approach-avoidance behavior tendencies towards empathic encounters.

A second potential area for improvement is that none of the extant empathy questionnaires has been developed or refined using item response theory (IRT) methods. In the one study thus far to evaluate the measurement precision of existing empathy self-report measures at varying latent trait levels, Murphy et al. (2018a) observed that empathy scales may generally have weak measurement precision at high levels of the latent traits. This could pose problems for empathy research with ostensibly high-empathy populations, such as therapists and other helping professionals, or attempting to identify “high empathy” participants from larger participant pools.

Previous attempts at developing empathy questionnaires have generated a large body of both questionnaire items and conceptual hypotheses. Any attempt to produce new scales should aim to analyze and integrate as much of this prior development work as possible. Although many previous development attempts have utilized items borrowed from other measures (e.g., Reniers et al., 2011; Spreng et al., 2009; Batchelder et al., 2017), these prior attempts have generally not

made targeted efforts to generate new or revised items specifically for potentially overlooked content domains, especially those theorized and explored in this dissertation.

Main content dimensions of existing empathy scales. In perhaps the most rigorous multi-dimensional exploration of existing item pools thus far, Reniers et al. (2011) factor-analyzed 65 items drawn from four empathy questionnaires. The resulting QCAE (Reniers et al., 2011) contains five scales, based upon the five dimensions observed in the factor analyses. Although I do not adopt the scale labels offered by Reniers et al., the five measurement domains (shown in Table 1) they identified can serve as a reasonable starting point for discussing the potential domains that might be observed in this dissertation project.

Metacognition of mindreading ability. Although a large number of studies have relied on self-report scales to measure cognitive empathy ability, the findings of Murphy and Lilienfeld (2019) suggest that such self-report measures may be largely invalid as proxies for actual ability, at least as operationalized by laboratory indices. As a result, this construct and item type was not be included in the empathy item pool for this dissertation project

Perspective-taking reasonableness. Researchers have most frequently relied on the Perspective-Taking (PT) scale of the IRI (Davis, 1983) to measure cognitive empathy. Sample items from that scale include “I try to look at everybody's side of a disagreement before I make a decision,” “I believe that there are two sides to every question and try to look at them both,” and “Before criticizing somebody, I try to imagine how I would feel if I were in their place.” The QCAE’s (Reniers et al., 2011) Online Simulation scale is primarily composed of items from the IRI PT scale, with a few similar items from other measures also included.

Although researchers generally classify these two scales as measuring cognitive empathy, it is probably more accurate to view them as measuring affective motivation to be fair,

reasonable, and understanding towards others. Murphy and Lilienfeld (2019) meta-analytically demonstrated that scores from these scales, much as with self-reported cognitive empathy ability scales, bear negligible associations with behaviorally assessed cognitive empathy ability. These scales do, however, substantially correlate with self-report measures of affective empathy (e.g., Reniers et al., 2011; Vachon & Lynam, 2016), so much so that they sometimes appear to be indistinguishable from empathic caring (e.g., Alterman, McDermott, Cacciola, & Rutherford, 2003; Siu & Shek, 2005). In other words, this domain may actually be more aligned with "affective" than "cognitive" elements of empathy.

This construct domain was included in the empathy item pool for this dissertation project. In addition to the existing items for this domain, I wrote new items designed to directly reference the potential interference of strong emotions (e.g., "If I start to feel frustrated with someone, I focus on understanding that person's point of view"). In daily life, perspective-taking reasonableness often does not occur in an emotionally neutral state; rather, it occurs or does not occur depending upon whether a person successfully makes the effort to step outside of his or her own negative emotions towards others in order to see things from a different perspective.

Empathic contagion. Multiple empathy questionnaires contain scales corresponding to the restrictive "isomorphic matching" conceptualization of empathy, which Murphy et al. (2018b) broadly referred to as "empathic contagion" (BES: Joliffe & Farrington, 2006; EI: Jordan, Amir, & Bloom, 2016; QCAE: Reniers et al., 2011; ESE: Olderbak et al., 2014). Sample items include "After being with a friend who is sad about something, I usually feel sad" (BES), "I can't watch shows in which an animal is being hunted because I feel nervous as if I am being hunted" (EI), and "I am happy when I am with a cheerful group and sad when the others are glum" (QCAE).

In these scales, contagions for different emotions, or for general valences of emotions (negative versus positive), are not generally differentiated at the scale total level. Instead, all contagion items are aggregated into a single dimension. One study, Olderbak et al. (2014), created contagion scales for each of six primary Ekman emotions, but the authors did not factor analyze them together.[21]

As described earlier, Murphy et al. (2018b) preliminarily observed that contagion for aversive emotions and contagion for appetitive emotions, though positively correlated, sharply diverge in their nomological networks, especially in relation to emotional distress, adult attachment, and personality disorder features. In this dissertation project, I took efforts to ensure that adequate numbers of items were available to potentially create strong scales for both aversive and appetitive empathic contagion.

Empathic caring. Although many researchers exclude compassion and caring from the empathy construct because such emotions are not isomorphic matches of the emotions experienced by the other (e.g., Jordan, Amir, & Bloom, 2016), most empathy questionnaires contain a general affective empathy scale primarily related to empathic concern or caring. This is, by far, the most common aspect of affective empathy measured in prior studies (e.g., Hall & Schwartz, 2019).

As compared with other kinds of empathy scales, though, the item content within these unidimensional scales tends to be quite facially heterogeneous, especially across measures. The IRI Empathic Concern (EC) scale includes items such as "When I see someone being taken advantage of, I feel kind of protective towards them" and "I am often quite touched by things

[21] I elected not to heavily utilize this item pool, as I was primarily interested in splitting contagion only into aversive and appetitive aspects and thought that I had sufficient coverage without its 30 items; in hindsight, this decision ended up being a major mistake.

that I see happen." The ACME Affective Resonance (AR) scale includes items such as "It makes me feel good to help someone in need" and "I get excited to give someone a gift that I think they will enjoy." The QCAE Proximal Responsivity scale includes only four items, all taken from the EQ, such as "Friends talk to me about their problems as they say that I am very understanding" and "I often get emotionally involved with my friends' problems." The item pools from unidimensional global empathy measures, such as the EQ and TEQ, contain even more heterogeneous collections of items broadly related to empathic caring.

Based solely on examining item contents, there are a number of potentially distinguishable item types within these pools, such as:

• Empathy in situations of positive emotionality, such as wanting to bring someone joy or to share in another's pride

• Being an understanding emotional confidante for others, being an attentive and supportive listener

• Feeling compassion for people who are suffering, even if not personally known

• Empathic restraint/repair, avoiding hurting others' feelings or desiring to undo harm one has caused to others

• Valuing empathic traits in oneself and others, relative to other traits/attributes

Despite the facial heterogeneity of item types, aside from the ECQ, empathy questionnaires do not parse these affective empathy items into different scales (except for distinguishing emotional contagion from empathic concern, e.g., QCAE). One possibility is that these various item types are, indeed, practically indistinguishable constructs. Another possibility, though, is that they have failed to manifest as different factors in previous analyses because the different potential domains have not been assessed with a sufficient number of items. In large

item pools, distinguishable constructs assessed with only a few items will frequently fail to emerge as separate factors in factor analyses (Bandalos & Finney, 2010). Nonetheless, as described below in the section on bifactor models, even if these domains emerge separately as factors in factor analyses, they may still not demonstrate *practical* value in being separately measured.

This dissertation project attempted to ensure that these potentially distinguishable domains were adequately represented in factor analytic investigation by (a) including item pools from across the range of empathy questionnaires and (b) writing new items for these potential domains. My hypothesis was that multiple affective empathy dimensions would emerge from these various items, provided that each content domain has a sufficient number of items written for it.

Fantasy/absorption. The IRI Fantasy (FN) scale is arguably an odd empathy scale. It aims to measure tendencies to become imaginatively absorbed in the feelings and action of fictional characters, such as in books and movies, and it is inspired by the idea that being empathically engaged with fictional characters can foster empathy and also serve as a manifestation of it (Davis, 1983). The FN scale is composed of items such as "I am usually objective when I watch a movie or play, and I don't often get completely caught up in it" and "Becoming extremely involved in a good book or movie is somewhat rare for me."

It appears to capture not only a person's relative love for books and movies, but also something akin to Tellegen's absorption construct (Tellegen & Atkinson, 1974), with which it correlates substantially (Murphy et al., 2018a; Wickramasekera & Szlyk, 2003). Partly because the FN scale does not directly relate to interpersonal functioning, many researchers exclude it

from analyses utilizing the IRI (e.g., Hatcher et al., 1994; Jolliffe & Farrington, 2004). The QCAE Peripheral Responsivity scale is composed primarily of items from the FN scale.

Although many researchers do not regard the FN scale as a valid measure of the empathy construct (e.g., Joliffe & Farrington, 2004), it may point towards a fruitful alternative measurement domain. The FN scale focuses on absorption in books and movies, but human empathy and human stories have existed far longer than the printing press and Hollywood, and reframing these kinds of items in terms of person-to-person communication may illuminate a valuable avenue of empathy measurement.

Rather than using existing Fantasy items, I revised existing items (e.g. "Becoming emotionally involved in another person's story is very rare for me") and wrote new items designed to capture empathic interest in directly hearing the stories of others in personal conversation (e.g., "When someone reveals his or her stories to me, I instinctively imagine what it might feel like to be that person" and "When people open up to me, I can't help but feel imaginatively involved in the stories they tell").

Of all the different domains of self-reported empathy, this potential domain, at least facially, appears to most closely resemble "einfühlung empathy." Yet, the items do not distinguish between "einfühlung empathy" and projection (a major difficulty in item generation noted, for instance, by Bryant, 1987). I was unable to craft items that facially did so in a way that did not make them too cumbersome and confusing for a self-report measure. Moreover, given that the degree to which one projects one's feelings onto others is probably well beyond reliable metacognition, and that regulation of projection is a skill/ability that likely does not receive consistent performance feedback, I highly doubted that self-reports would be valid proxies for such distinguishing.

Personal distress/emotional regulation *(not examined by Reniers et al., 2011)*. The IRI Personal Distress (PD) scale, which correlates highly with trait neuroticism and negative emotionality (e.g., Hawk et al., 2013; Murphy et al., 2018a), is composed of items like "In emergency situations, I feel apprehensive and ill-at-ease" and "I sometimes feel helpless when I am in the middle of a very emotional situation." Only two of the PD items refer to the presence of another person (e.g., "When I see someone who badly needs help in an emergency, I go to pieces"), which presumably explains in part why the items were not analyzed by Reniers et al. (2011).

Although this scale does not appear to be a manifestation of empathy, it points to a potentially valuable empathy-related domain, which can be referred to as "empathic distress." Empathic encounters can be difficult to endure and individuals presumably differ in their ability to emotionally regulate their self-oriented distress levels. The aim of the IRI PD scale, as explained by Davis (1983), was to capture the extent to which individuals experience self-oriented emotions of distress as opposed to more helpful other-oriented feelings (cf. Batson, Fultz, & Schoenrade, 1987). Though the PD scale has failed to demonstrate convergent validity with the empathy construct (e.g., Murphy et al., 2018a) this original aim may still be worth continued efforts in item revision and new item development.

I revised the IRI PD scale items so that they explicitly referenced other people and empathic encounters, rather than just "emergencies" (e.g., "When people share their deep emotions, I feel apprehensive and ill-at-ease" and "I feel scared when I am in a conversation with someone who is opening up about his/her deepest worries"). I also wrote new items to more sufficiently cover this content domain.

New content domain: empathic approach/avoidance. Although some items from existing measures presumably relate, at least indirectly, to individual differences in the tendency to be avoidant towards empathic experiences, none of the existing questionnaires significantly and directly appears to measure this theoretical domain.

In particular, almost none of the items currently used appears to relate to approach-avoidance tendencies regarding situation selection, situation modification, and attentional deployment (with a few scattered exceptions, e.g., "When a friend starts to talk about his/her problems, I try to steer the conversation towards something else," found in the TEQ). In order to potentially remedy this gap, I wrote a substantial number of new items for this broad domain, such as "When people start talking about their own painful feelings, I sometimes only half-listen" and "If I can tell that someone needs a supportive listener, I go out of my way to give them my time and attention." My prediction was that this domain would be an especially strong predictor of other constructs conceptually related to empathy.

The 11 potential domains which I hypothesized, and for which I attempted to provide adequate item coverage, are presented in Table 2. All the items administered and considered for inclusion in new scales are listed and organized in Appendix A.

Employing Bifactor Modeling in Scale Creation

When considering scientific constructs, determining the extent to which concepts should be lumped together or split apart has long been a common dilemma. Identifying heterogeneity is critical for the fine-grained investigation and discussion of natural phenomena (and intellectual concepts in general). Yet, at some point, the splitting apart must terminate due to practical considerations, well before infinitesimal differences generate infinite categories (cf. Peirce, 1878). From Teddy Roosevelt (1895, p. 687; cf. Lydekker, 1897) arguing that the "excessive

multiplication of species based on trivial points of difference" was interfering with effective discussion in zoology to Berg (2018) arguing that the excessive splitting of postdoctoral job titles is interfering with empirical study of professional training paths, there is a point at which parsing heterogeneity may hurt more than help.

This dilemma is front-and-center in any attempt to create or revise a psychological measure. Should schizotypal personality be measured and analyzed as a single dimension, three broad separate dimensions (positive symptoms, negative symptoms, disorganized symptoms), or seven distinguishable narrow dimensions (cf. Cohen et al., 2010)? Is it acceptable to use a total score for Neff's (2003) Self Compassion Scale, or should only the six facet scores be analyzed and presented (cf. Neff, Whittaker, & Karl, 2017)? When evaluating an extensive body of empathy questionnaire items, should a researcher create broad domain scales, narrow facet scales (how many?), or both?

The strongest appeal of broad, heterogeneous content scales is that they are often (depending upon the specificity of the outcome variable) superior in *prediction*, demonstrating more pronounced relationships with external criteria than narrow content scales (McGrath, 2005; Smith, McCarthy, & Zapolski, 2009). Narrow, more homogenous content scales, on the other hand, are often superior in *representation*, as the interpretive meanings of scale scores are less ambiguous. In terms of measure construction, evaluation, and interpretation, this basic tradeoff has been recognized since at least the time of Loevinger's (1954) "attenuation paradox," the observation that as the internal consistency of a scale is increased, it may often come at a cost of reduced construct validity. Nonetheless, when a multidimensional domain is parsed into sub-domain scales, the collection of those scales (e.g., in multiple regression) tends to provide

stronger statistical predictive validity than a single "total score" of them all combined (McGrath, 2005).

From a theoretical standpoint, in measure *construction*, the goals of (a) maximizing the value of a total score and (b) maximizing the values of narrower subscales are not well aligned. Many of the items that are the most predictive indicators of the overall latent dimension are unimpressive indicators of any particular narrower subdomain and/or are items that cross-load across multiple sub-domains, not distinguishing well among them. Many of these items are actually relatively "pure indicators" of the core of the overall construct (whether the construct is substantive, artifactual, or both to some degree). Researchers focused only on creating distinguishable subscales will often tend to discard them without recognizing their particular value. On the other hand, selecting items based on their being the strongest loading ones in a unidimensional EFA will lose sight of heterogeneity.

In recent decades, there has been a strong trend toward the creation of new multidimensional measures and the revision/reinterpretation of existing unidimensional measures into multidimensional forms. Factor analytic findings of multidimensionality, whether through exploratory factor analyses (EFAs) or confirmatory factor analyses (CFAs), are frequently the lynchpin of this multidimensional trend. Nonetheless, in very recent years, a counter-movement has begun to emerge, arguing that many of the narrow subscale/facet scores of psychological measures are substantively and psychometrically superfluous. For example, Reise, Bonifay, and Haviland (2013, ps. 132, 136) have opined that "findings of multidimensionality do not guarantee that subscales can provide meaningful and reliable information about subdomains that is unique from the general construct….there is a superabundance of examples of personality and psychopathology measures where an essentially unidimensional domain has been broken up

unnecessarily into subscales." In other words, just because a broad domain splits statistically into a collection of narrower sub-domains in factor analytic models does not necessarily mean that all sub-domains have practical value in psychological research. The "better-fitting" model is not necessarily the more useful measurement model.

Bifactor models constitute the methodological basis for many of these critiques of multidimensional measures. In a conventional bifactor model, the covariance among a set of items is modeled as accounted for by both a superordinate general factor (reflecting the common variance of all items) and also a number of specific factors (reflecting the residual correlations between sets of items). Conventionally, the model assumes that the general factor is uncorrelated with the specific factors and that the specific factors are uncorrelated with one another (for a review, see Markon, 2019).

For example, consider a battery of intelligence items in a bifactor model. The general factor represents the common variance across all intelligence items, whereas the specific factors represent unique subdomains of intelligence (e.g., "verbal," "spatial") that are not fully captured by the general factor. The general factor and the specific factors are both *directly* associated with the individual items, competing to explain common variance. In other words, for each item, the variance in correct/incorrect responding will be partially explained by the general factor of intelligence and partially explained by a specific factor, such as spatial reasoning.

Although bifactor models have origins in Spearman's (1904) models of intelligence, and were extensively described by Holzinger and Swineford (1937), they have been only rarely used until the last decade (for a review, see Reise, 2012). Instead, psychological researchers have overwhelmingly relied upon Thurstone's (1935) correlated factors (CF) model. The CF model does not prohibit superordinate factors (Thurstone, 1944), but it models them as relating directly

to subordinate factors rather than to the items themselves. In CF models, the associations between superordinate factors and the items are indirect, mediated by the subordinate factors.

The most common use of bifactor models involves comparing bifactor CFAs to other CFAs to judge the factorial structure of data. Although many have compared bifactor CFAs to non-bifactor CFAs to argue that various construct domains have a bifactor structure (e.g., Anderson & Marcus, 2019; Caspi et al., 2014), a number of studies have indicated that bifactor CFAs are biased to produce better fit (e.g., Bonifay & Cai, 2017; Reise, Kim, Mansolf, & Widaman, 2016). Reviewing past findings, Markon (2019, p. 62) concluded, "On average, a given bifactor confirmatory model will tend to fit to any given random data set better than a nonhierarchical confirmatory model, and it will fit almost, but not quite, as well as an exploratory factor model. This…reflects the ability of hierarchical models to capture chance features of any given sample that might not replicate." As a result, a number of scholars have strongly cautioned against comparing bifactor CFAs to other CFA models to make claims about the "best-fitting" model structure (e.g., Bonifay, Lane, & Reise, 2017; Watts, Poore, & Waldman, 2019).

Bifactor models, however, are "invaluable for informing the degree to which a measure yields an univocal total score…and, relatedly, the extent to which subscales representing theoretically distinct constructs (i.e., group factors) yield reliable scores after accounting for the general factor" (Bonifay et al., 2017, p. 184). Because bifactor models relate the general factor directly to the items, a range of valuable psychometric indices can be computed with bifactor structures (for reviews, see Rodriguez, Reise, & Haviland, 2016a; 2016b). For instance, the Explained Common Variance (ECV) is the ratio of variance explained by the general factor divided by the total variance explained by the general plus the specific factors (Stuckey &

Edelen, 2015). The ECV can be computed both for the whole collection of items as well as for individual items.

The most prominently utilized bifactor indices, though, are based upon McDonald's (1999) coefficient omega (ω), which is the model-based latent variable analogue to Cronbach's (1951) alpha reliability (Revelle & Zinbarg, 2009). Its importance is due to the fact that, in bifactor structures, ω can be parsed into ωH (omega hierarchical) attributed to the general factor and ωHS (omega hierarchical subscale) attributed to the specific factor. If a subscale has a very high ωH and a very low ωHS, that is an indication that it is almost entirely a reflection of the general factor alone. If a scale (or factor) has a very low ωH and a very high ωHS, that is an indication that it is almost entirely a reflection of specific factor content.

The bifactor omega perspective presents a Scylla and Charybdis dilemma. If a subscale has an extremely low ωHS, then it is possible that the total test score of the multidimensional measure is actually a stronger predictor of the specific factor construct than the specific factor construct's subscale (Haberman, 2008). If a subscale has an extremely high ωHS, however, then that is an indication that the subscale should not be viewed as a narrow facet of a more general construct (Reise et al., 2013).

As a general rule of thumb, if a general factor has an ωH above .8, it is an indication (though imperfect) that the degree of unidimensionality in a "multidimensional" measure is so high that it is acceptable to consider an overall total score as essentially unidimensional (Reise et al., 2013; Rodriguez, Reise, & Haviland, 2016a, 2016b). Nonetheless, if certain other indices are strong, undimensionality can be reasonably adopted even if ωH is below .80 (Reise, Schienes, Widaman, & Haviland, 2013).

Rough benchmarks have not yet been set in regard to the strength of ωHS that would indicate a strongly distinguishable subscale. Reise et al. (2013, p. 137) stated, "Tentatively, we can propose that a minimum would be greater than .50, and values closer to .75 would be much preferred, but that is a subjective guideline." However, subsequent research has indicated that a .50 benchmark would be so stringent that very few subscales in multidimensional measures would be able to meet it (e.g., Rodriguez et al., 2016b)

Rodriguez et al. (2016b) retrieved bifactor loadings from 50 different recently published articles in personality and psychopathology literatures, all of which had examined multidimensional measures and reported that bifactor CFAs were well-fitting. Based upon the retrieved loading patterns, they calculated bifactor indices for the various measures. Surprisingly, they found that the average ECV of the "well-fitting" multidimensional measures was .67; the average ωH of the general factor was .80 and the average ωHS of subscales was only .27 (the median would be substantially lower). In other words, across the many different multidimensional measures: the general factor accounted for substantially more than half of the explained variance; the measures tended towards unidimensionality; and the specific factors were responsible for only a comparatively small degree of the scale internal consistencies.

In the last few years, a rapidly increasing number of studies have been published using omega-based bifactor indices, often to either support the use of total scores (e.g., Neff, Whittaker, & Karl, 2017) or to caution against specific subscale scores (e.g., with the Woodock-Johnson IV, Dombrowski, Beaujean, McGill, & Benson, 2018). Unless a subscale has a truly minimal ωHS, though, such bifactor indices cannot dispositively conclude that a subscale is superfluous or a total score is better than facet scores; a specific factor may only contribute a small degree of explained common variance, yet that small amount may be highly valuable.

Nonetheless, at a minimum, the disconcerting trends for subscale distinguishability in bifactor indices highlight that, when a subscale's reliability is largely attributed to a general factor, meaningful substantive interpretation of the subscale's correlates requires separating the subscale's specific factor common variance relative to that of the general factor (Reise et al., 2013). In other words, in ambiguous situations vis-à-vis bifactor indices, the value or superfluity of a subscale should ultimately hinge on whether its specific factor content adds incremental value, above and beyond the general factor, in statistically predicting important external correlates.

In CF approaches, the external validity focus is usually on (a) demonstrating that subscales have convergent and discriminant validity and (b) demonstrating that different subscales relate differently to external correlates, that they possess "differential validity."[22] This same approach can be taken with different measures altogether, not just subscales within a measure. For example, some research has used differential validity approaches to assess whether superficially similar constructs such as self-esteem, neuroticism, locus of control, and generalized self-efficacy possess substantial differential validity (Judge, Erez, Bono, & Thoresen, 2002). An implication of the bifactor perspective, though, is that scales might demonstrate differential validity only because they are saturated with a shared general factor to different degrees, not because their specific factor contents actually relate differently to external correlates (there could be "differential validity" even if all their specific factors are totally inert in relating to the external variables).

A bifactor approach to external validity can add particular value, at least as a supplement to traditional perspectives on differential validity. If a specific factor demonstrates meaningful

[22] Of course, it is possible to examine incremental validity of subscales above and beyond a higher-order factor in a CF framework.

incremental value above and beyond the general factor in relating to theoretically important external criteria, that would be an indication that it should be measured and analyzed even if its ωHS is low (for a related perspective, see Ferrando & Lorenzo-Seva, 2019). Moreover, by partitioning out the variance explained by the general factor for all subscales under investigation, the differential validity of the different subscales can be investigated in a more fine-grained manner. Such approaches certainly cannot replace traditional CF perspectives, though; if subscales relate substantially differently to a general factor, that is a form of differential validity in itself. Nonetheless, partialing out the general factor from specific factors can be valuable in at least a supplemental fashion, helping to more clearly elucidate the meaning and value of specific content domains.

Moreover, the nature of the general factor in a bifactor model is that it will tend to capture systematic response bias (e.g., Lorenzo-Seva & Ferrando, 2018; cf., Backstrom Bjorklund, & Larsson, 2009; also cf., Davies, Connelly, Ones, & Birkland, 2015), which would include systematic socially desirable responding and evaluative self-deception. As a result, in controlling for the substantive content of the general factor in an incremental validity approach, one is also likely controlling, at least to some extent, for such response biases when examining the unique variances of specific content domain scales.

To date, the overwhelming number of studies utilizing bifactor modeling for psychometric purposes have been conducted using (a) confirmatory factor analyses with (b) existing measures. Until recently, the only available method for exploratory bifactor analysis (EBFA) was Schmid and Leiman's (1957) method, commonly known as the Schmid-Leiman, or SL, transformation. This procedure essentially fits a basic EFA, then adds a higher-order factor, and then relies upon certain "ideal situation" assumptions to translate the higher-order factor into

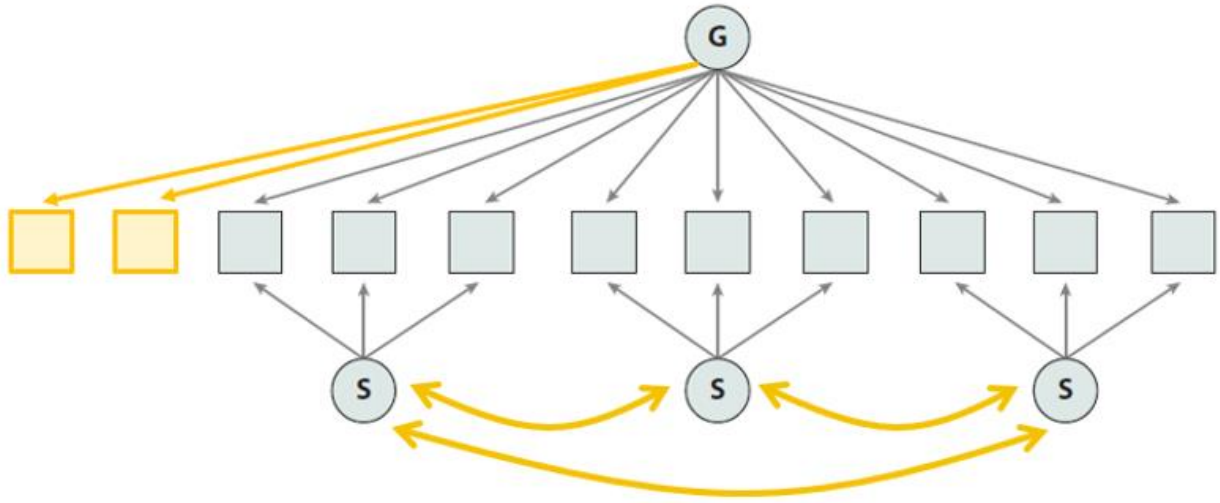
a general factor that directly relates to the items themselves. Of note, all factors are orthogonal in the SL transformation method. Although the SL transformation will generally be somewhat inaccurate, given that its constraints are rarely satisfied with real data, it will usually not be terribly far off the mark; as a result, it has been argued that it can still be used with some confidence to at least identify overall patterns of trivial and non-trivial loadings (Reise, Moore, & Maydeu-Olivares, 2011). In situations where there are "pure indicators" of the general factor (items that load on the general factor but not appreciably on any specific factor) or where there are specific factor cross-loadings, the SL method is not very accurate. For example, SL transformations tend to fail to identify pure indicators, instead forcing them to saliently load on a specific factor (Abad et al., 2017). In ideal situations, though, with no pure indicators or cross-loadings, the SL transformation is highly accurate (Abad et al., 2017).

Motivated by the limitations of the SL transformation, Jennrich and Bentler (2011) were the first to introduce an alternative exploratory bifactor method, one which also orthogonalizes all factors but directly rotates to a bifactor solution rather than translating a EFA. Nevertheless, this method has been substantially criticized on a number of technical grounds (e.g., Mansolf & Reise, 2016) and, on average across simulation situations, actually appears to be less accurate than the SL transformation (Abad et al., 2017). Very recently, a number of other new methods have been presented (e.g., Abad et al., 2017; Lorenzo-Seva & Ferrando, 2018; Waller, 2017). Although research is currently underway comparing these various new competitors (e.g., in the lab of Niels Waller), preferred methods and best practices have yet to be fully established. There is still ample room for improvement (Lorenzo-Seva & Ferrando, 2018).

Given the known and unknown limitations of existing methods, the safest approach is to employ multiple different exploratory bifactor methods and compare them for convergence (e.g.,

Chin, Buchanan, Ebesutani, & Young, 2018). This is perhaps especially important when working with relatively small sample sizes. At the same time, it is also important to select methods that, at least on the basis of the limited evidence available, appear to work best with the kind of underlying structure of the data (e.g., presence or absence of pure indicators).

An interesting issue is that existing methods (e.g., Jennrich & Bentler, 2011; Waller, 2017) tend to rely on orthogonalization of all factors and evaluation studies tend to only consider situations in which specific factors are orthogonal to one another (e.g., Abad et al., 2017). As noted by Markon (2019), in order for specific factors be correlated with one another in an admissible bifactor model, at least (a) some pure indicators of the general construct must be present and (b) they must be allowed to be modeled as such (e.g., a bifactor CFA cannot specify that they load on any specific factor).[23] The figure below depicts this requirement: the basic orthogonal model is in black, whereas an oblique model is specified by adding the yellow components.



[23] Markon (2019) does not give a line of justification for this claim and it is unclear as to why this is the case. I have explored it in my datasets, however, and observed that oblique bifactor models fail if I remove pure indicators.

In choosing to not examine oblique EBFA methods, multiple reasons are cursorily given by researchers (e.g., Abad et al., 2017; Morin, Arens, & Marsh, 2016). First, when compared with more parsimonious orthogonal models, oblique models may pose interpretation problems. Second, the greater complexity of oblique methods can lead to problems with model convergence. Third, oblique rotations are simply not common in bifactor model research and valuable statistical indices (e.g., ωH and ωHS) are well understood in orthogonal models rather than oblique models.

As argued by Lorenzo-Seva and Ferrando (2018), however, it makes sense to conduct an exploratory bifactor analysis using oblique rotation to determine whether the factors are meaningfully correlated after partitioning out the general factor. If they are not meaningfully correlated, then an orthogonal method has more justification. If the specific factors are correlated, though, and their correlations can be viewed as substantive rather than as merely construct-irrelevant nuisance, then orthogonality between the factors can be artifactual. As Lorenzo-Seva and Ferrando (2018) pointed out, especially in personality assessment domains, the general factor can, to some degree, reflect response biases such as extreme responding or acquiescence. We can add to that brief list other response biases, such as malingering, social desirability, or response patterns that emerge out of "satisficing" questionnaire responding. In other words, though far less investigated than methods which restrict the specific factors to being orthogonal to one another, and very rarely used used in prior published bifactor analyses of extant measures, bifactor analysis allowing for correlated specific factors should frequently be used in research projects (assuming that there are pure indicators in

the model to allow it). This is especially likely to be true when analyses are exploratory and researchers have no prior basis for expecting orthogonality of the specific factors.

In the case of empathic caring, and probably many other heterogeneous construct domains, for instance, an assumption of complete orthogonality between specific factor domains seems conceptually unrealistic (unless the item pool is so saturated with the general factor that the added value of the specific factors is trivial or minimal). If the residual variances of common factors are meaningfully correlated, then the parsimony gained by orthogonal methods may carry a significant cost in terms of proper substantive interpretation (much as with the costs and benefits of oblique and orthogonal rotations in more conventional factor analysis models, though to a lesser degree of distortion severity).

Only one oblique bifactor method, the oblique variant of the JB EBFA presented by Jennrich and Bentler (2012), has been in existence for a few years and has been tested in oblique situations in multiple published articles. As Jennrich and Bentler have preliminarily demonstrated with multiple real datasets, using an oblique rotation with JB EBFA methods appears to produce more interpretable loading matrices with fewer cross-loadings than its earlier orthogonal variant. If the underlying data, however, produces an oblique rotation with exact bifactor structure (each item loading on one and only one specific factor, with zero loadings on other specific factors), then oblique JB EBFA will be undefined; Jennrich and Bentler (2012) suspected that this generates distortions when an oblique rotation approaches exact bifactor structure even if it is not truly exact bifactor structure. They argued, however, that this problem is likely to be very rare when working with real data, which will almost never produce an exact bifactor structure. I suspect, however, that it may be a more common difficulty, though, when working with multidimensional scales that have been created by dropping cross-loading items.

In direct comparison with the PEBI exploratory bifactor model with correlated factors offered by Lorenzo-Seva and Ferrando (2018), the only other such oblique model investigated in peer-reviewed literature, the JB EBFA with oblique rotation was extremely accurate in estimating general factor loadings, but was less accurate than the PEBI method in estimating specific factor loadings (Lorenzo-Seva & Ferrando, 2018). Of note, however, its accuracy in estimating specific factor loadings increases substantially as the number of specific factors in the model increases, with excellent accuracy for models with 5 or more specific factors (Lorenzo-Seva & Ferrando, 2018).

Although a substantial number of recent studies have employed exploratory bifactor models in evaluating existing measures (e.g., Bryan & Harris, 2019; Chen, Osman, Freedenthal & Gutierrez, 2019; Grosz et al., 2019; Persson, Kajonius, & Garcia, 2019; Sunderland et al., 2019; Vilagut et al., 2019), virtually all of them employing either the SL transformation or the orthogonal variant of the JB EBFA, very few research projects have utilized bifactor models in crafting new measures. The value of bifactor modeling, though, may be even greater during initial stages of measurement construction than in later examinations; it is better to notice and address problems proactively at the beginning of the process.

Bifactor modeling can be used to:

- Assist with determining how many factors to extract when generating scales from a large item pool
- Assist with determining the psychometric distinguishability of subscales from the outset
- Ensure that items selected for subscales best measure the specific content domain under consideration

- Allow the identification of pure indicators of the core construct. In EFA approaches to measure construction, these items will often be discarded, given that they will tend to not load highly on any factor and/or will have substantial cross-loadings.

The very few measure construction efforts to use EBFA thus far, though, are informative. For example, Ebesutani et al. (2012) used an SL transformation to create a shortened version of the Anxiety scale of the Revised Child Anxiety and Depression Scale. For each of the five subscale domains, they retained three items that loaded well on both the specific factor and the general anxiety factor. This resulted in a shortened scale that was both usable in a unidimensional form while also retaining the multidimensional breadth of the broad construct.

Passarelli et al. (2018) used a targeted rotation of an initial SL transformation to create a new facial emotion recognition test. Observing that almost all items failed to load on any of the six specific factors (representing the six basic Ekman emotions), but instead all operated as pure indicators, they concluded that the item pool was best used unidimensionally, and they employed IRT methods to reduce the item pool.

In another published attempt, Garfield et al. (2019) used EBFA to create a 20-item measure of father involvement in child health. With an initial 47 item pool, they used the orthogonal variant of JB EBFA, extracting a general factor and 4 specific factors. Many items loaded well on the general factor, but not on any specific factors. Interestingly, they only retained items that loaded substantially on a specific factor, discarding pure indicators of the unidimensional construct.

The work of Deveney et al. (2019), though, *indirectly* illustrates the value of retaining and using pure indicators of a general factor. They theorized that the construct of trait irritability

can be empirically distinguished from closely related constructs such as trait reactive aggression and trait anger proneness. To distill trait irritability to its core features, they used Jennrich and Bentler's (2011) orthogonal bifactor rotation, extracting one general and three specific factors, to analyze the aggregated item pool of three different irritability questionnaires. They then interpreted the general factor as a measure of the "core" features of irritability and used its factor scores as a "core irritability" variable in subsequent analyses.

If other researchers, following upon this study, wished to measure the core features of irritability, in a way that minimizes other specific factor elements of existing measures, they would not necessarily need to administer all three of the questionnaires. In Deveney et al. (2019), five of the top six items loading on the general factor were pure indicators. There were a number of other strong pure indicators as well. Theoretically, a non-specific general factor scale could be fashioned from these pure indicators, potentially with the use of IRT as well. Although Deveney et al. (2019) did not fashion such a scale, their work gives an example of the potential value of doing so.

EBFA can potentially be used to generate scales measuring broad constructs, but in a way that reduces multidimensionality and focuses on the substantive core of the overall construct. Minimizing multidimensionality in this way not only assists with enabling IRT (which requires adequate unidimensionality), but also provides other benefits, such as better interpretability in regression "partialling" situations (cf. Sleep et al., 2017). As opposed to simply taking the top-loading items from a unidimensional EFA, selecting strong pure indicators from EBFAs may often be superior for some conceptual purposes.

If an extensive, comprehensive pool of empathy questionnaire items produces a strong general factor, a valuable scale can potentially be fashioned from only pure indicators of the

general factor. This scale will represent the "core" of the broader construct. Given that it will have minimized specific factor content, and also have no item overlap with specific factor scales, it can serve as a high-quality tool against which simultaneously created subscales, and other scales (including new scales that have yet to be conceived and created), can be tested for incremental validity. In future studies, it can serve as a strong approximation of the general factor from a comprehensive item pool, or as an "anchor" for a new general factor, without necessitating that a massive item pool be continually administered in each new study.

In this dissertation research, I relied on EFAs and CFAs. I supplemented them, though, with follow-up bifactor analyses. When conducting initial views of overall patterns of trival and non-trivial loadings, I relied upon SL transformations. When engaged in more fine-grained EBFAs, such as for final item selection, I employed a combination of more sophisticated orthogonal and oblique methods and compared them for convergence.

Methodological Overview of Present Studies

The approach to scale construction and validation taken in this dissertation research attempted to conform to best practices and recommendations, most notably as described by Clark and Watson (1995, 2019), with further influence from other important conceptual reviews and recommendations (e.g., Loevinger, 1957; McGrath, 2005). To craft high-quality self-report questionnaires, an adequately comprehensive item pool must be generated, with potentially overinclusive coverage rather than underinclusive coverage of potential content domains, and with adequate items available to potentially reliably represent each theoretical content domain. This item pool should then be subjected to factor analytic procedures to preliminarily determine the number and interpretability of latent dimensions underlying a pool (or multiple pools) of items. Further conceptual (e.g., close reading of items for substantive content) and psychometric

investigation (e.g., scale reliability analyses) must be applied to determine which items should be retained for use in scales.

The construct validity of the scale measurements (both convergent and discriminant) should then be investigated by examining their associations with relevant external variables and comparing the associational patterns of different scale measurements with one another (differential validity). Although this last consideration is often neglected in initial creation studies, which frequently administer only the item pool for the potential new scales, "anchor" scales of convergent and discriminant constructs are highly recommended in the initial creation phase, to investigate the meaning of the scales from the outset. Study 1 aimed to meet these various initial goals.

Factor structures identified in the exploratory creation phase should be confirmed or disconfirmed in a subsequent study (ideally, in multiple studies with different sample characteristics). Where deficiencies are noted or suspected in the scales, further studies should attempt to ameliorate and/or more strongly characterize potential problem areas; this process may require multiple study iterations beyond the initial creation study[24]. Additional studies should further expand the observed nomological networks of the scale measurements. Additional studies should also aim to reevaluate, or at least revisit, the *constructs* interpreted as characterizing the scales (and, beyond the scales, the theoretical constructs themselves); for instance, initial labels or descriptions of scales may need to be modified based on nomological network results in further studies. Study 2 aimed to meet these goals, to the limited extent possible in a single initial validation study.

[24] "We have emphasized the iterative process of scale development. Phrased in its most extreme form, scale development ends only when a measure is "retired" because, owing to increased knowledge, it is better to develop a new measure of a revised construct than to modify an existing measure" (Clark & Watson, 2019, p. 10).

Both studies were approved by the Emory University IRB. Funding ($2,500) was provided by the Laney Graduate School.

Study 1. In Study 1, I administered an extensive collection of empathy questionnaires items, alongside new or revised items I generated, to 800 M-Turk participants. I also administered the HEXACO-60 broadband personality inventory (Ashton & Lee, 2009), to initially investigate the nomological networks of the preliminary scales I subsequently created.

After substantial participant screening and exclusion (described in depth below), I first used the ICLUST (Revelle, 1979; Revelle, 2015) procedure, as well as unidimensional EFA, to (a) evaluate the suitability of the item pool for bifactor modeling and (b) determine the degree to which the item pool should be initially parsed into clusters for further factor analytic investigation. I identified a large cluster of items pertaining to a general domain of empathic caring (and a small cluster highly correlated with it), as well as a smaller cluster of items pertaining to empathy-related negative emotionality, which was minimally correlated with the large empathic caring cluster. Given the conceptual bifactor approach I subsequently utilized, I analyzed the two major clusters separately in factor analyses.

With the large cluster of general empathic caring items (combined with the highly correlated small cluster), I used a combination of basic EFAs and multiple methods of EBFAs to identify the multidimensional factor structure of the item pool and then identify items for inclusion in preliminary scales. For the broad empathic caring domain, I first generated a unidimensional scale of 15 “quasi-pure” indicators of the general empathic caring construct, with the assistance of item response theory (IRT) methods to attempt (ultimately, with only modest success) to provide adequate measurement precision across all levels of the latent trait. Next, I constructed six specific content scales (perspective-taking; confidante listening; appetitive

contagion; empathic restraint; empathy avoidance; empathic restraint; and empathic absorption), each representing narrow facets within the broader empathic caring construct. I employed bifactor indices to preliminarily evaluate the distinguishability of the specific content scales from the general factor of empathic caring.

With the smaller cluster of empathy-related negative emotionality items, I determined that a bifactor approach was not merited conceptually (nor psychometrically), for reasons discussed below. I relied upon basic EFAs to identify the multidimensional factor structure of the item pool and identify items for inclusion in three preliminary scales (empathic intimacy discomfort/distress; anxiety contagion; and sadness contagion).

Finally, I examined the nomological networks of my newly created scales in regard to broadband personality traits, as assessed by the HEXACO-60. I compared my "general factor of empathic caring" (GFEC) scale with a broad but unidimensional empathic caring scale, the TEQ (Spreng, McKinnon, Mar, & Levine, 2009). For all nine specific content scales, I regressed each on my new GFEC scale, to examine their unique variances, above and beyond the general factor of empathic caring. Although I analyzed the zero-order correlations between the specific content scales and the HEXACO dimensions (and with one another), I focused on their unique variance residual correlations in my interpretive evaluations of the preliminary scales.

The findings of Study 1 raised a number of concerns regarding multiple of my newly created scales. I concluded that one specific content scale ("empathic absorption") was fatally flawed. I concluded that other major concerns should be further investigated and/or ameliorated in Study 2.

Study 2. In Study 2, I administered the preliminary scales created in Study 1, along with the TEQ, to a sample of 372 M-Turk participants. I also administered a limited number of new

items intended to buttress the quality of the various specific content scales. To further investigate the nomological networks of my new scales, I administered self-report measures of a number of constructs conceptually related to empathy, such as emotional well-being, aggression, Machiavellianism, psychopathy, and schizotypal personality.

After participant screening and exclusion, I examined the internal consistencies of each scale, to identify items that should be dropped or added. I then conducted confirmatory and exploratory factor analyses of the scales, to confirm or disconfirm the factor structures established in Study 1. Using IRT models, I examined the measurement precision of all my new scales, as well as the TEQ, across the range of latent trait levels.

In extending the nomological networks of the scales, I relied upon the following measures: the Satisfaction with Life Scale (SLS; Diener, Emmons, Larsen, & Griffin, 1985), the Penn State Worry Questionnaire (PSWQ: Meyer, Miller, Metzger, & Borkovec, 1990), the Forms of Self-Criticising and Self-Reassuring Scale (FSCRS; Gilbert et al., 2004), the Buss-Perry Aggression Questionnaire (BPAQ; Buss & Perry, 1992), the Machiavellian Personality Scales (MPS; Dahling, Whitaker, & Levy, 2009), the Meanness and Disinhibition scales of the Triarchic Psychopathy Measure (TriPM; Patrick, 2010), and the Schizotypal Personality Questionnaire Brief-Revised (SPQ-BR; Cohen, Matthews, Najolia, & Brown, 2010). These scales, their subscales, and acronyms used are presented in Table 3. I formed a priori predictions for the subscales of these various measures based upon related prior findings in the research literature, initial personality associations observed in Study 1, and theorized conceptual links. As with Study 1, in evaluating the nomological networks of the specific content scales, I analyzed zero-order correlations but focused on relationships demonstrated by their unique variances above and beyond the GFEC scale.

The findings in Study 2 generated a range of interpretive conclusions about the heterogeneity of the empathy construct, particularly the distinction between empathic caring and empathy-related negative emotionality. Findings also generated recommendations for future use of my new scales in research, as well as future directions for their potential improvement.

Data analytic methods across both studies. When working with individual items, whether in exploratory factor analyses, estimations of numbers of factors, or in scale-level internal consistency analyses, I employed polychoric correlations rather than Pearson correlations, consistent with the ordinality of the 4-point Likert scale response options (cf. recommendations by Sellbom & Tellegen, in press).

For EFAs, I used minimum residual factoring with oblimin rotation[25], implemented with the PSYCH package in R (Revelle, 2017). Although maximum likelihood (ML) factoring may be superior when variables are normally distributed, minimum residual factoring is less sensitive to violations of normal distribution (Briggs & MacCallum, 2003; Revelle, 2017), a common feature in ordinal data. With ordinal data with four or fewer response options, ML is decidedly inferior (as summarized in Sellbom & Tellegen, 2019). In interpreting the meaning of factors in EFAs, I followed Tabachnik and Fidell (2001), who recommended .32 as a minimum item loading for salience. In determining whether a factor has adequate reliable indicators to form a measurement scale, Osborne and Costello (2005), following a review of the literature, recommended a minimum of at least 3 items with loadings above .5. I used this benchmark in my analyses.[26]

[25] Though there are other oblique rotation methods, they tend to produce roughly the same results (Fabrigar et al., 1999).

[26] As pointed out by Clark and Watson (2019), however, narrow dimensions should have higher minimum loading thresholds than broader dimensions.

For initial exploratory bifactor analyses, I utilized SL bifactor estimation (Schmid & Leiman, 1957), implemented with the *omega* function in the PSYCH package. Although the traditional SL method relies upon assumptions that are rarely met, and only achieves an extremely high level of accuracy in certain unusual situations, it is the most widely employed exploratory method and is generally regarded as adequate for establishing preliminary overall conclusions (e.g., Reise, Moore, Maydeu-Olivares, 2011; personal communication with Dr. Niels Waller). For more fine-grained (e.g., scale item selection) exploratory bifactor examination, I employed three recently developed methods and compared them for convergence: iterated SL estimation (Abad et al., 2017) with a partial specification threshold of .20, the same method with a partial specification threshold of .05, and the oblique variant of the biquartimin JB EBFA (Jennrich & Bentler, 2012). SLi analyses were performed using the FUNGIBLE package in R (Waller, 2019), whereas JB EBFAs were performed using the PSYCH package. Given that no EBFA method has yet emerged as a clear "gold standard" (Lorenzo-Seva & Ferrando, 2018), the safest approach is to employ multiple methods and examine their results for convergence (e.g., as done in Chin, Buchanan, Ebesutani, & Young, 2018).

Specific factor loadings in bifactor models tend to be weaker than their counterpart loadings in EFAs, given that partialling out the general factor removes a substantial amount of reliable variance. A loading of .2 or higher is conventionally regarded as salient (e.g. default specification threshold in SLi, Abad et al., 2017; SL in PSYCH package, Revelle, 2017), but an equivalent benchmark for a reliable indicator does not appear to have been determined. Without much empirical basis for deciding upon a benchmark for this, I chose .3 as a threshold, reasoning that items that loaded that strongly may be likely to manifest as at least salient indicators across different studies.

Confirmatory factor analyses were performed in the LAVAAN package in R (Rosseel, 2012), using the WLSMV estimator, which is most appropriate for working with ordinal data (Sellbom & Tellegen, in press). Bifactor CFAs with correlated specific factors were conducted alongside bifactor CFAs with orthogonal factors, but only when pure indicators were present in the item pool (as recommended by Markon, 2019). In evaluating model fit indices, such as CFI and RMSEA, I followed the rough (and comparatively stringent) benchmarks proposed by Hu and Bentler (1999); I kept in mind, however, that these benchmarks cannot, on their own, be fully dispositive (e.g., Sellbom & Tellegen, 2019).

When determining the degree of multidimensionality within item pools (including at the scale level), I relied primarily on Horn's (1965) parallel analysis method and Velicer's (1976) minimum average partial (MAP) test. Although historically the most widely used method of estimating the number of latent factors, Kaiser's (1960) criterion of retaining factors with eigenvalues above 1 tends to overestimate the number of latent factors underlying a set of variables (Fabrigar et al., 1999). Horn's (1965) parallel analysis technique is perhaps the most generally reputable method for estimating the number of factors to retain. In Horn's parallel analysis, eigenvalues are calculated for randomly generated variable sets with the same number of cases and items as the real variable set. The eigenvalues obtained with the real dataset are compared with those obtained at random; factors in the real dataset with eigenvalues larger than their randomly generated counterparts are presumed to be non-illusory factors. The parallel analysis method has been observed to be highly accurate compared with other common methods (e.g., Fabrigar et al, 1999; Ruscio & Roche, 2012).

Velicer's (1976) minimum average partial (MAP) test, although more likely to be inaccurate than the parallel analysis technique (Ruscio & Roche, 2012), is an additional widely

used method for estimating the number of factors, which has also been observed to perform substantially better than the Kaiser criterion (e.g., Velicer, Eaton, & Fava, 2000). In this technique, factors are partialled out of the correlation matrix, one by one, while the mean partial (i.e. residual) inter-correlation of the items is tracked. The estimated number of factors is equal to the number of factors partialled out from the correlation matrix once the minimum average partial inter-correlation is obtained. As highlighted by researchers using bifactor indices (e.g., Rodriguez, Reise, & Haviland, 2016b), though, these methods cannot conclusively establish the number of reliable, substantively distinguishable factors within a variable set; closer conceptual examination is required rather than "theory-blind" mechanistic decision-making (for a similar perspective, see Sellbom & Tellegen, in press).

When selecting items for a scale, "the goal is to measure one thing (i.e., the target construct)—and *only* this thing—as precisely as possible" (Clark & Watson, 2019, p.10). The most commonly used measure of scale reliability is Cronbach's (1951) coefficient alpha (which, with ordinal items, should be adjusted to an ordinal alpha, based upon the polychoric correlation matrix). Although lower alpha levels are frequently described as "acceptable," a level of .80 is the preferred benchmark for adequate internal consistency (Clark & Watson, 2019). Alpha reliability is highly sensitive to test length, given that measurement precision increases with aggregation. For lengthy scales, it is a poor measure of *homogeneity*, though, as a sufficiently large number of very disparate items can achieve a "strong" alpha reliability level (Clark & Watson, 2019; McGrath, 2005).

Clark and Watson (1995) advocated reliance on the "average item inter-correlation" (AIC), but have more recently (2019) argued that the AIC is weaker than first supposed, given that strong items can obscure the problems with very weak items and, also, that substantial latent

heterogeneity can still be present even when AIC is strong. As a result, Clark and Watson (2019) now recommend examining all items individually in terms of their correlations with other scale items. In this dissertation research, when evaluating individual scales, I employed ordinal α, but also examined (a) each item's correlation with the scale (corrected for item overlap) and (b) the unidimensionality of the scale (e.g., via Horn's parallel analysis).

I also employed item response theory (IRT) techniques when evaluating scales, primarily in Study 2. Classical test theory assumes that test reliability is constant at all levels of the latent construct, from the lowest levels to the highest. This assumption, however, may often be quite unwarranted; a psychological measure, for instance, may have acceptable overall measurement precision, but very poor measurement precision at particular ranges of the latent construct. Filling this empirical gap, IRT is a psychometric modeling paradigm that evaluates reliability at different levels of the latent construct, both at an overall test level and in reference to individual items (De Ayala, 2013).

Consider a standardized test for statistical competence, with a collection of a multiple-choice questions asking students to match a term with its definition. Some questions will be relatively easy, with most students getting them correct. Other questions will be very hard, with only the best students consistently getting them correct. In other words, item will vary in *difficulty* (occasionally referred to by other terms, such as "severity" or simply "location", Thomas, 2019). Beyond difficulty, items will also differ in how well they relate to the overall content of the test. Some items, whether easy or hard, may be highly related to statistical competence (e.g., define 'correlation'), whereas other items may be much less related to the overall construct (e.g., define 'h index'). The degree to which an individual item reliably

discriminates between test-takers higher versus lower on the overall latent construct discrimination is described, in IRT terms, as the item's *discrimination*.

Despite the obvious value of IRT techniques, they have been infrequently used in personality and psychopathology research (Reise & Waller, 2009), but are gradually becoming more common (Thomas, 2019). In the only IRT examination of self-report empathy scales to date, Murphy et al. (2018a) observed that empathy scales tended to have good measurement precision from low to moderate levels of empathic traits, but very poor measurement precision at high trait levels.

To examine the measurement precision of scales and items in these dissertation studies at varying latent trait levels, I relied on the unidimensional "graded response model" (GRM; Samejima, 1969), which is a two-parameter IRT model (2PL; evaluates both difficulty and discrimination) appropriate for use with ordinal response options as opposed to simply binary response options (e.g. correct or incorrect). GRM models generate item information curves, which graphically depict an item's discrimination value (how much information it provides) at different levels of the latent construct. The individual item information curves for items in a scale are summed to create an overall test information function (TIF) curve. In the TIF curve, the conditional standard error of measurement for a given latent construct level (e.g., 2 SDs above the mean) equals the inverse square root of the information level at that latent construct level. A TIF of 10 at a specific latent construct level is equivalent to a marginal reliability of .9 at that level, whereas a TIF of 5 is equivalent to that of .8 (Embretson & Reise, 2000). In almost all cases, I used GRM models only with items from a single scale. In one case, however, I employed a multidimensional 2PL model with a JB EBFA model of the overall empathic caring item pool;

I compared this for convergence with a basic limited-item unidimensional GRM, however, given the uncertainties of bifactor modeling.

Examination of nomological networks was performed in IBM SPSS Version 25, using Pearson correlations between total scale scores. To (partially) account for the multiplicity of tests conducted, I used a significance level of $p < .01$ to identify salient associations. In addition to examining zero-order correlations, I also examined associations for the unique variances of scales, above and beyond a general factor of empathic caring. As described in the Methods section of Study 1, I created unique variance variables by regressing each specific content scale, one at a time, on a scale measuring the general factor of empathic caring (the GFEC scale) and then saving the standardized residuals. This method allowed me to examine the incremental value added by each specific content scale, above and beyond the "core" of the broad empathic caring construct.

This method also allowed me to better assess the particular nomological networks of the different specific content constructs. Nonetheless, as Lynam, Hoyle, and Newman (2006), echoing Meehl (1975) and others, pointed out, "it is difficult to know what construct an independent variable represents once the variance shared with other independent variables is removed" (p. 329). The unique variance of a particular variable may capture something different than what its original label/description/interpretation indicated (e.g., error variance, construct-irrelevant variance; Lynam et al., 2006; Verona & Miller, 2015). The interpretative difficulties are exacerbated as shared variances with an increasing number of variables are partialed out, as is common in many multivariate regression studies (e.g., what is the substantive meaning of the residualized variable of psychopathic trait meanness, after partialing out shared variance with 7

other psychopathy-related variables?). Moreover, the difficulty is especially problematic when the constructs partialed are multidimensional (Sleep et al., 2017).

These concerns are certainly not eliminated in this dissertation research, but they are comparatively minimized: specific content scales generated were narrow in content and the GFEC scale is notably unidimensional. As a result, the substantive content of the unique variances of specific content scales can probably be assumed to be reasonably close to the substantive content meaning of the original scales (at least in comparison with other common partialing examples in psychological research).

Related to this, I was particularly interested in examining whether the general factor of empathic caring obscured the underlying relationships between specific content scales and external correlates, a situation referred to as "suppression" (e.g., Sleep et al., 2017). Statistical suppression is "equivalent" to mediation (MacKinnon, Krull, & Lockwood, 2000), but a suppressor variable is one which increases the predictive validity of another variable when included in a regression equation (either increasing the strength of an association or causing the directionality of the relationship to change signs), whereas a mediator variable decreases the predictive validity of the other variable. Suppressor variables are rare, and frequently difficult to replicate, but, when they are robust, they can help demonstrate the importance of distinguishing between related variables (e.g., Paulhus, Robins, Trzesniewski, & Tracy, 2004; Watson, Clark, Chmielewski, & Kotov, 2013).

In investigating potential suppression effects of the general factor on the specific content scales, I conducted bootstrapped mediation tests using the PROCESS macro in SPSS (Hayes, 2012) with bootstrapped samples of 1000 and their 99% confidence intervals. I did not test all associations for suppression, only those in which the residual variable exhibited a significant

association that was either stronger than or in the opposite direction of the zero-order association. In determining statistical significance and effect sizes, I examined the completely standardized indirect effects, alternatively referred to as the "index of mediation" by Preacher and Hayes (2008), and their 99% confidence intervals.

Addressing Concerns with M-Turk Samples

The two studies conducted in this dissertation research both utilized online community samples recruited via Amazon Mechanical Turk (M-Turk). M-Turk samples have become widely used in psychological research, drawing on more demographically diverse samples than university study pools, and allowing for the speedy and inexpensive collection of large data sets (for a short review, see Buhrmester, Talaifar, & Gosling, 2018). Additionally, M-Turk samples are generally higher in indices of distress-related psychopathology and lower in emotional well-being than student samples (e.g., McCredie & Morey, 2018; Stone et al., 2019), potentially allowing for improved examination of psychopathology constructs.

M-Turk samples, however, have been properly regarded with suspicion, due to the potential presence of four different problematic respondent types: (1) non-human bots; (2) incentivized inattentive respondents, participants who compromise quality to quickly complete tasks and increase their effective rate of pay; (3) "professional" income-generating respondents who may be highly practiced (non-naïve) in particular tasks (Chandler, Mueller, & Paolacci, 2014); and, finally, (4) foreign respondents masquerading as North American respondents, with poor proficiency in the English language (Moss & Litman, 2018). Although potential bot respondents have been thought to be a major concern, Moss and Litman (2018) of the TurkPrime company have observed, in multiple studies, that foreign workers (particularly from India, using

server farms) masquerading as North American participants are a much larger concern, with actual bot activity minimal or non-existent.

Highly practiced M-Turk participants are not of particular concern to this dissertation research, given that no common behavioral tasks were employed. Unreliable or low-effort respondents, though, whether low-effort scammers, respondents with limited English proficiency (LEP), or automated bots, pose a serious threat to research projects such as this dissertation, especially in regard to factor analytic methods. Unfortunately, in M-Turk samples, with subjects unsupervised, distracted and/or multi-tasking, and financially incentivized to complete tasks as quickly as possible, unreliable responding is unsurprisingly common.

For instance, in a series of studies, Deetlefs, Chylinski, and Ortmann (2015) estimated that approximately 11% of M-Turk respondents could be classified as providing substantially unreliable response patterns. Other studies have found similar rates of careless/inattentive responding (15%, Barends & de Vries, 2019; 13%, Lian et al., 2019; 11-13%, Wood, Harms, Lowman, & DeSimone, 2017). Nonetheless, studies comparing M-Turkers to student samples in regard to attentive responding are mixed, with some finding M-Turkers to be substantially less attentive (e.g., Aruguete et al., 2019; Goodman et al., 2013) but others finding no differences (Behrend, Sharek, Meade, & Wiebe, 2011; Casler et al., 2013; Hauser & Schwarz, 2015). In an extensive review of inattentive responding findings in both M-Turk samples and other sample types, Thomas and Clifford (2017) concluded that M-Turk attentiveness quality is not significantly different than in other samples; inattentiveness is a substantial problem in research in general.

Although it is often assumed that low-quality responding will only increase measurement error, this is not necessarily the case if low-quality responses are systematically correlated with

substantive content domains. If respondents are negligent in the reflective effort they devote to tasks, for instance, one common consequence is that measures administered may appear to be spuriously high in internal consistency (provided all items within the measure are coded in the same direction), generate spuriously high correlations with other measures, and appear to be high in convergent validity yet deficient in discriminant validity (for discussion, see Hauser, Paolacci, & Chandler, in press; also see findings in Wood et al., 2017). This can occur because participants engage in "straight-lining," marking the same response to rows of items. It can also, however, occur because participants respond to items in a rapid evaluative (heuristic) manner rather than a more reflective manner, failing to attend to more subtle differences between superficially similar items and failing to effortfully integrate evidence/memories when deciding upon a response. For example, some participants may rapidly register item contents as little more than "good" and "bad" (or, "socially desirable" and "socially undesirable") and respond in a generalized low-effort (but "consistent") manner on this basis.

Hamby and Taylor (2016) described this second dynamic in terms of "satisficing" questionnaire responses, as opposed to "optimizing" questionnaire responses through careful, effortful reflection (including the retrieval and weighing of relevant episodic memories). Their studies indicate that (a) M-Turk samples evidence higher rates of satisficing (assessed as non-differentiation between scale items) than college samples, and that (b) satisficing substantially predicted higher internal consistencies of measures but worsened discriminant validities[27].

The consequences of low-quality responding are different but especially severe in the presence of a mix of both positively-coded and negatively-coded items. Even a minor amount of careless or confused responding in a sample can cause spurious item coding method factors to

[27] As Hamby and Taylor (2016, p. 925) point out: this presents "the surprising, and unfortunate, implication that high scale reliabilities may sometimes signal that the scales have poor validity."

manifest in factor analyses (e.g., Woods, 2006), with unidimensional constructs splitting into two different factors founded in item coding direction in factor analyses (e.g., Murphy et al., 2018a; Weeks et al., 2005). Item coding factors are not necessarily completely artifactual, as item coding direction can relate to substantive domains such as self-esteem or defensiveness (e.g., DiStefano & Motl, 2009) or to item difficulty level (e.g., Ray, Frick, Thornton, Steinberg, & Cauffman, 2016). Nonetheless, if even a small proportion of respondents carelessly or confusedly mistake the directionality of query statements (e.g., mistakenly marking "Strongly Agree" rather than "Strongly Disagree"), this can seriously disturb factor analytic results for methodological rather than substantive reasons.

Although not (to my knowledge) yet discussed in terms of bifactor modeling, item coding direction artifacts may also undermine bifactor techniques. Specific factors may be generated based upon item coding direction, as with EFAs, of course. Moreover, though, if item coding direction effects are particularly strong, a single general factor may (often unbeknownst to the researcher) need to be split into two general factors, one for positively coded items and one for negatively coded items. One major indication of such a dynamic would be if a basic EFA extracting two factors (even if the expected number of factors is much larger) generates a factor of positive-coded items and a factor of negative-coded items.

In sum, it was imperative that I take strong steps both before and after protocol administration to limit low-quality responding in my datasets. First, I set very high standards for M-Turk reputation level of eligible participants (98% and 99% in my two studies, respectively, higher than typical recommendations, e.g., Buhrmester, Talaifar, & Gosling, 2018). Second, I used TurkPrime screening methods to exclude duplicate geolocations and suspicious geo-locations flagged as linked to server farms. Third, I limited the length of the study

administration, given that excessive study length is related to increasing negligent responding (e.g., Galesic & Bosnjak, 2009). Fourth, I randomized the order of questionnaire items for each participant, presumably decreasing the extent to which similar items, ordered immediately after one another, would promote non-differentiation between items. Fifth, I used a consistent 4-point Likert scale[28] for all items, to (a) decrease confusion and cognitive load resulting from constant switches in response options and (b) eliminate a "neutral" option, which can increase satisficing behavior (as recommended by Chyung, Roberts, Swanson, & Hankinson, 2017) and the accompanying "error of central tendency."

Nonetheless, I expected that a substantial number of respondents would still display unreliable response patterns. I used multiple methods to screen for such participants after data were collected. First, in each study, I included an open-ended question asking for a written text response, to potentially help identify bots or participants with limited English proficiency. Second, given that overly rapid completion times have been recommended as a filter with M-Turk samples (e.g., Aruguete et al., 2019), I excluded data from participants who completed the studies too quickly to have allowed for careful responding (in this case, I used a cutoff of 2 seconds or less per questionnaire item, as used by other researchers, e.g., Huang et al., 2012; Ward & Meade, 2018).

Third, and most importantly, I used a squared-difference procedure assessing deviation between pairs of antonym twinned items (positively coded items and versions of the same items written in the opposite coding direction), which has been observed to be particularly effective at identifying unreliable responders (e.g., Litman, Robinson, & Rosenzweig, 2015; Moss & Litman,

[28] Having a consistent response scale and randomizing item presentation also serves to prevent methodological differences (e.g., different response scale types, responding fatigue for later items, etc.) from competing with substantive item contents in factor analyses.

2018; for similar "consistent responding" approaches, see Aruguete et al., 2019, Lian et al., 2019, and Wood et al., 2019). I selected 10 simply worded items from the study item pool, and wrote facially opposite items for each of them (other researchers, however, have observed that a decent approximation of this method can be done simply by identifying psychometric antonyms within an existing set of variables, e.g., Huang et al., 2012). As described further below, in the Methods section of Study 1, I took an empirical approach to setting cutoff levels for participant exclusion, rather than choosing a cutoff level a priori. I ended up excluding approximately 10% of the participants in each study, a rate of unreliable responding similar to that observed in past M-Turk studies (e.g., Deetlefs, Chylinski, & Ortmann, 2015), despite my rigorous attempts to reduce unreliable responding through participant recruitment and study design.

It is likely that many questionnaire respondents, whether in M-Turk samples or any other types of samples, provide "reliable" response patterns even while "satisficing" rather than "optimizing." In other words, such participants may provide proper written text answers, attentively read items, reliably recognize coding direction shifts, and respond in a consistent manner, yet stop short of carefully considering the subtle differences between similar items. It is possible that, in almost any dataset of questionnaire responses, whether from M-Turkers or other study populations, only a minority of data points will actually be the result of "optimized" cognitive effort. Despite my various efforts to improve data quality, it is quite possible that my study data still result from comparatively elevated satisficing, as observed in other M-Turk samples (e.g., Hamby & Taylor, 2016).

Study 1: Initial Development of General and Specific Content Scales

The overarching aims of Study 1 were to (a) examine the heterogeneity within an extensive collection of existing, new, and revised empathy questionnaire items, (b) distill observed content dimensions into preliminary scales, and (c) investigate the distinguishability and incremental associative value of the preliminary scales under consideration.

Methods

Participant filtering. Via Amazon Mechanical Turk, 800 participants completed the study protocol for $1.50. The median time to completion was 28.4 minutes, which was longer than the expected 20-25 minutes indicated in the recruitment advertisement and consent form; of note, however, M-Turk participants are not required to complete tasks in a single uninterrupted block of time, and M-Turkers frequently multi-task or take breaks (e.g., Chandler et al., 2014).

Twenty-eight participants failed the speed filter, answering more quickly than 2 seconds per item on average. One participant did not answer the open-ended text question, "What is your favorite genre of music?" All other participants answered reasonably. I decided, though, that this question may have been too easy and might be completable by bots, and almost certainly by many LEP participants. The one participant who left the item blank clearly passed other filters, though, and so was not excluded from the dataset (participants were explicitly told they could choose to not answer certain questions, per IRB requirements).

For the 10 item pairs employed in the squared difference procedure of antonym pairs, I reverse-coded one item from each pair. I then calculated the deviation between items in each pair (e.g., if one was scored 4 and the other as 2, then the difference between them was 2), which ranged from 0 to 3 (the maximum). Then I squared the deviation for each pair. This means that small deviations (e.g., "Strongly Agree" when "Agree" is expected) are minimized, whereas

extreme deviations (e.g., “Strongly Agree” when “Strongly Disagree” is expected) are amplified. I then averaged the squared deviation across all 10 item pairs, which ranged from 0 to 4.8 (M = .72, median = .50, mode = .30, SD = .71). To empirically determine a proper cut-point for exclusion, I binned participants into sub-groups based upon their mean squared deviation score and then examined the internal consistency of the HEXACO general personality scales for each group. Above approximately 2 SDs higher than the median, alpha reliability of the scales dropped strongly below the other groups, so I chose 2 SDs as the threshold for exclusion (75 participants). At this level, with a cutoff score of approximately 2, all individuals above the cutoff point had substantial and facially troubling deviations in their responding for the item pairs. Ultimately, I excluded 88 subjects from analysis (fewer than 103 due to strong overlap between the two filter methods), leaving a total of 712 participants for analysis.

Participant characteristics. The 712 participants retained for analyses ranged in age from 19 to 81 (M = 41.3 years, SD = 12.3 years) and 52.0% identified as female. As with most M-Turk samples, participants were more highly educated than the general population; 38.3% had (only) a bachelor’s degree and 14.9% had graduate level degrees of some kind. Six percent of participants identified as mixed-race/ethnicity; of those who self-identified with a single race/ethnicity, 80.4 were “white,” while other race/ethnicities were much less well represented (black or African American, 8.2%; Hispanic/Latino, 3.7%; Asian, 5.8%; all others less than 1%).

Questionnaire items administered. The full collection of empathy items administered, both from extant measures and items I created for this project, is presented in Appendix A. The item key used for these items in study 1 is presented in Appendix D. I included the items from the Empathic Concern and Perspective-Taking scales of the IRI (Davis, 1983). I revised the IRI’s Personal Distress scale items to reference to the presence of other people or empathic encounters,

rather than simply "emergencies."[29] I also revised the IRI's Fantasy scale items to reference to face-to-face stories rather than books or movies.[30] Further, I revised the "affective empathy" items of the BES (Joliffe & Farrington, 2006) so that they referenced people in general, rather than all specifying "friends," given that the scale appears to inadvertently measure strength of quality of friendships rather than only general empathy.

In addition to the IRI and BES, I administered the item pools (excluding cognitive empathy ability and fantasy-equivalent scales) of the ER scale of the Empathy Quotient (EQ: Baron-Cohen & Wheelwright, 2004), the Toronto Empathy Questionnaire (TEQ: Spreng et al., 2009), the Questionnaire of Cognitive and Affective Empathy (QCAE: Reniers et al., 2011), the AR scale of the Affective and Cognitive Measure of Empathy (ACME: Vachon & Lynam, 2016), the Empathy Index (EI: Jordan, Amir, & Bloom, 2016; items retained by Murphy et al., 2018b), and the Empathy Components Questionnaire (ECQ: Batchelder, Brosnan, & Ashwin, 2017).

As described earlier, I theorized a number of dimensions that are not adequately captured in extant questionnaires, though some measures contain one or two such items: concern for happiness of others; empathic restraint; confidante listening; appetitive contagion; empathic approach-avoidance; empathic distress; empathic absorption; and empathy valuing. After examining items from existing measures that theoretically relate to these dimensions, I wrote new items to ensure that each potential domain had at least 10 items pertaining to it. I also wrote

[29] For example, I changed "In emergency situations, I feel apprehensive and ill-at-ease" to "When people share their deep emotions, I feel apprehensive and ill-at-ease."

[30] For example, I changed "Becoming extremely involved in a good book or movie is somewhat rare for me" to "Becoming emotionally involved in another person's story is very rare for me."

new items for well-covered domains, focusing on crafting "more difficult" items. Ultimately the empathy item pool consisted of 210 items.

Of all the extant empathy measures administered in this study, I only analyzed the TEQ (ordinal α = .95) as its own measure, with the intention of comparing it with the general factor of the comprehensive item pool. The TEQ is a general empathy scale created by analyzing 95 items from existing questionnaires, plus an additional 36 items generated by Spreng et al. (2009). To this point, it represents the most comprehensive attempt to aggregate empathy questionnaires to create a new emotional empathy questionnaire. In fashioning the TEQ, however, Spreng et al. (2009) forced all items into a unidimensional EFA, disregarding potentiality multidimensionality in the item pool. As a result, the TEQ offers an excellent "anchor" comparison measure for the general factor of emotional empathy I anticipated would be revealed in bifactor analyses of my dissertation item pool.

Finally, to assess general broadband personality traits, I administered the HEXACO Personality Inventory–60 (HEXACO–60; Ashton & Lee, 2009), a 60-item broadband questionnaire containing 24 facets that coalesce into six higher-order factors: Honesty/Humility (ordinal α = .85), Emotionality, Extraversion (ordinal α = .90), Agreeableness (ordinal α = .88), Conscientiousness (ordinal α = .86), and Openness (ordinal α = .86). The HEXACO Emotionality higher-order factor is atypical in its heterogeneity, consisting of three lower order dimensions conceptually related to negative emotionality (Fearfulness, Anxiety, and Dependence), but also a lower-order Sentimentality dimension containing items strongly related to empathic contagion (e.g., "I feel like crying when I see other people crying" and "I remain unemotional even in situations where most people get very sentimental"). The fact that emotional contagion items have been found in the creation of the HEXACO to be more related to negative

emotionality than to, say, trait agreeableness, provides a hint that at least some forms of emotional contagion may be more related to negative emotionality than to empathic caring. Because of the content overlap, I did not include the Sentimentality facet when scoring the HEXACO Emotionality dimension; as a result, I refer to HEXACO NSE ("non-sentimental emotionality"; ordinal $\alpha = .84$), as in Murphy et al. (2018b).

I did not make predictions regarding the HEXACO until after generating the preliminary scales through factor analysis. After scale dimensions emerged, I then made a priori predictions regarding their associations with HEXACO dimensions. These predictions are described in the Results section, immediately before the analyses are presented.

Initial item filtering for use in measure construction. I dropped psychometrically duplicative items from the empathy item pool, to reduce the likelihood that doublets or triplets of nearly identical items would manifest as separate factors. I identified duplicative items by examining the polychoric correlations between all items and noting any correlations at or above $r = |.8|$. After having identified 17 such correlations, I subjectively compared the two items for wording complexity and retained the member of the item pair that I judged to be less complex in wording interpretation. As a result, even though many items are facially very similar to and statistically strongly correlated with one another, none of the items considered in analyses can be understood as fully duplicative.

I also removed "fantasy" items referencing books or films or plays (e.g., "I usually stay emotionally detached when watching a film"), as well as unvalenced contagion items (e.g., "I get caught up in other people's feelings easily").[31] Finally, I removed multiple items that had content

[31] Though I excluded fantasy, CE ability, and other conceptually distal or problematic scales of existing measures from study administration entirely, some of the scales administered nonetheless had occasional items reflecting these content domains.

that appeared to be potentially more related to self-interested social monitoring than to empathy: QC5 (“I always try to consider the other fellow’s feelings before I do something”) and QC7 (“Before I do something, I try to consider how my friends will react to it”). This item filtering process left 184 items for further analysis.

Initial item cluster analyses. When a researcher theorizes that a bifactor structure may apply in a set of data, Mansolf and Reise (2016) recommend first preliminarily exploring the data with the hierarchical item clustering method ICLUST (Revelle, 1978, 1979), to confirm and ensure that subsequent bifactor analyses are conceptually and psychometrically reasonable. ICLUST is similar to factor analysis, but based upon a much simpler model (e.g., one which may not requires as large of a sample size for initial examination) in which “the correlations between variables reflect that each item loads on at most one cluster, and that items that load on those clusters correlate as a function of their respective loadings on that cluster, and items that define different clusters correlate as a function of their respective cluster loadings and the intercluster correlations” (Revelle, 2015, p. 210). The ICLUST function works by iteratively combining two items into a new cluster, or adding an item into an existing cluster, or combining two existing clusters. In each iteration, clustering only occurs if the alpha and beta reliabilities of a new cluster will improve upon the reliabilities of the components to be combined, according to specified decision rules (the specific criteria cutoffs can be set or changed by the researcher using the ICLUST program). When none of the clusters can be improved in reliability, according to the decision rules, the ICLUST analysis halts.

Importantly, Revelle (1979) proposed that the beta reliability (the worst split-half reliability, much less sensitive to test length than alpha) of a major cluster can serve as a useful index of the general factor saturation of the variables loading in the cluster. Although

McDonald's (1999) omega is now the dominant tool for assessing general factor saturation, beta can still provide some preliminary insight into general factor saturation.

If a pool of empathy questionnaire items is strongly saturated with an overall general factor, ICLUST will produce either 1 comprehensive cluster, a number of strongly correlated clusters, or a combination thereof. If the items load onto a number of different smaller clusters that are not strongly correlated with one another, this would indicate that a bifactor approach with that particular variable set may not be appropriate. If most of the items cluster in a manner consistent with general factor saturation, but some items load into one or more separate clusters that are only weakly correlated with the apparently general factor-saturated cluster/s, this would indicate that those items may not be appropriate for inclusion in a bifactor approach. Items that either do not cluster with other items, or do so only very weakly, should perhaps be dropped from the item pool under investigation (especially when measure construction is the goal).

A benefit of the ICLUST method is that it provides a step-by-step clustering table and also a graphical representation of each step in the clustering process (e.g., Figure 1).[32] The iterative process begins with the strongest possible clustering step and proceeds in order of descending strength. As a result, the stage at which clusters are combined can provide *heuristic* clues as to the structure of the data; earlier combinations are stronger than later combinations. As a result, though ICLUST lacks the precision of proper EFA methods, it can still be helpful as an initial supplementary examination of an item pool, even if bifactor modeling is not a goal.

ICLUST of the 184 items generated three final clusters. The first cluster consisted of 144 items, broadly covering the construct of empathic caring. The second cluster consisted of 27 items, all from the hypothesized domains of aversive contagion and empathic distress. The third

[32] I have included a diagram for ICLUST with only the 34 negative affect items, discussed below. Diagrams for the 184 and 140 item pools were far too large to feasibly include in this document, however.

cluster consisted of 13 items from the hypothesized domain of empathic restraint. The general empathic caring cluster and empathic restraint cluster were strongly correlated (r = .64), indicating that empathic restraint is probably appropriate for inclusion in a bifactor model of the empathic caring construct. The second cluster, characterized by the observer's negative emotionality in empathic processes, was only weakly correlated with the general empathic caring cluster (r = -.21) and the empathic restraint cluster (r = .17). Consistent with prior research (e.g., Murphy et al., 2018b), and with my study hypotheses, this finding suggests that susceptibility to negative emotionality in empathic encounters should not be viewed as part of the broader empathic caring construct. Instead, this "empathy-related negative emotionality" cluster should be regarded as a separate broader construct domain.

As a follow-up confirmation of ICLUST, I next conducted a unidimensional EFA of the 184 items (polychoric correlations, pairwise deletion), which indicated that the vast majority of variables loaded above .5 on the general factor, which could be interpreted as a "general empathic caring" factor. Aversive contagion and empathic distress items, however, typically did not reliably relate to the general factor (most did not load above .32).

To not only reduce the number of items (improving my ability to reliably factor analyse the data), but also to generate a model that fits conceptually with the theoretical bi-factor model of a general empathic caring factor that includes narrower, but distinguishable content domains within, I excluded all items that: (a) did not load above .32 on the general factor or (b) were intentionally written to represent aversive contagion, empathic distress, or empathic restraint. The remaining 140 items, in a new unidimensional EFA, loaded from .40 to .87. The lowest loading items tended to be empathic restraint items, but other empathic restraint items loaded much more highly (e.g., multiple items above .6). I kept the empathic restraint items in the total

item pool but noted that further analyses might suggest that they should not be included in the bifactor model.

A new ICLUST analysis of the remaining 140 indicated a cluster of 134 items and a smaller cluster of 6 items (6 of the 11 potential "empathic restraint" items). The two clusters were, however, strongly correlated (r = .71). Although the alpha estimate for the reliability of the largest cluster was excellent (.99), the beta value was .8, which indicates that the largest cluster has strong general factor saturation but also a substantial degree of heterogeneity within it ("lumpiness," as referred to by Revelle, 1979). Moreover, three noticeable sub-clusters (12 items, 18 items, and 8 items, from latest to earliest) were collapsed into the largest cluster at a comparatively late stage (when the largest cluster was at least 78 items), while three noticeable sub-clusters (8 items, 9 items, 6 items, from latest to earliest) were collapsed into the largest cluster at middle stages of clustering (when the largest cluster was between 50 and 69 items). Another notable cluster (13 items) was collapsed into the largest cluster early, when it was only 27 items.

This iterative clustering process led me to hypothesize that 6-8 factors would emerge in a EFA (and also as specific factors in a bifactor EFA). Even if 8 or more factors emerged, though, they may not be adequately distinguishable from the general factor to be meaningful for measurement purposes, given that the 8th cluster observable within the ICLUST procedure was combined with the overall cluster at an early stage of the iterative process.

Three items in particular did not load on any cluster until the final few steps when they were loaded onto the largest cluster (Q189, Q212, and Q172). Examining these item contents, I concluded they were all clearly multiply determined, with explicit or implicit "double barrelled" content that goes beyond the broader empathy construct. Other items in the pool, though, have

elements of double barrelled content, but still appear to perform better than these three items. I removed these three items moving forward. Thus, this combination of ICLUST and the follow-up unidimensional EFA reduced the main item pool down to 137 items which seem to be appropriate for further investigation of potential heterogeneity within an overall empathic caring construct, and the subsequent creation of domain subscales.

I next conducted an ICLUST analysis of all the items theorized/created for the aversive contagion and empathic distress dimensions (all of which had been removed prior to analysing the 140 items in the prior ICLUST analysis of the 140 empathic caring items). These 34 items formed two clusters, one of 24 items characterized mostly by aversive contagion (but also with some empathic distress items) and one of 10 items characterized entirely by empathic distress items. The two clusters only weakly correlated (r = .22). Interestingly, items related to contagion for sadness and contagion for anxiety, two prominent domains within aversive contagion, clustered separately until being combined late in the process. This finding suggests that aversive contagion may be best parsed into contagion for specific aversive emotions, particularly sadness and anxiety. At the same time, though, this ICLUST analysis suggests that it would not be appropriate to model aversive contagion and empathic distress together in a bifactor model, despite their conceptual similarities and common relations to negative emotionality.

After examining the 34 items, I dropped the following seven items, all of which loaded late as lone items in the clustering process and had apparent content that seemed to explain their difficulties in loading on a cluster (e.g., relation to interest in TV/movies) and also seemed likely to extend beyond the theorized content domains: Q255, Q85, Q82, DC1, DC3, BE8, BE7. I also dropped multiple items (BE11, EQ1, QC12) from the aversive contagion cluster because they do not necessarily indicate that the empathic witness experiences the *same* emotion as the

empathized person's negative emotion (a common definitional requirement for researchers who treat aversive contagion as a core aspect of the empathy construct, e.g., Bloom, 2017b).

In sum, preliminary ICLUST analyses (along with a unidimensional EFA of empathic caring items) indicate that (a) a pool of 137 items appear to be strongly saturated with a general factor of empathic caring and that (b) items referencing negative emotionality in empathic processes appear to be much less related to general empathic caring. Moving forward, I handled the 137-item empathic caring item pool separately from the 24-item empathy-related negative emotionality item pool.

Estimating the number of factors in the 137-item general empathic caring item pool. Examination of the ICLUST results led me to expect approximately 6-8 substantial and distinguishable factor dimensions within the overall empathic caring item pool. The scree plot (see Figure 2) of the item pool indicated 1 very strong factor (indicative of a prominent general factor) and 4-6 much smaller factors, with no graphically obvious cut-off point for setting the number of factors. Horn's parallel analysis indicated 9 factors. Velicer's MAP procedure indicated 13 factors. Improvement in fit statistics began to minimize between 5 and 8 factors (empirical BIC indicated 8 factors). Considering the evidence as a whole, I concluded that the maximum number of dimensions from which reliable, interpretable scales could be fashioned was likely to be no more than 8, and that 6-7 dimensions would probably be most appropriate for measure construction. Nonetheless, I examined factor analytic results for models with as many as nine extracted factors.

EFAs extracting 9 factors. I started by examining EFAs extracting 9 factors (not including a general factor). I conducted (a) a basic EFA and (b) a basic SL EBFA. Polychoric correlations were used in all factor analyses. When examining the EFA model, I used a loading

cutoff of .5 to examine the most salient loading items, ones which could potentially serve as reliable indicators. When using the SL bifactor model, I used a loading cutoff of .3.

In the EFA model, using minimum residual factor method and an oblimin rotation, all factors were substantially inter-correlated (*r*s from .21 to .52) and 6 factors had at least 5 items loading above .5. These six factors were highly interpretable. In order of extracted eigenvalues, they were: a perspective-taking factor; confidante listening factor; appetitive contagion factor; empathic avoidance factor; empathic absorption factor; and empathic restraint factor. Only two items loaded above .5 on the 7th factor (VL1, VL2). These two items both relate to feeling good as a result of showing caring behaviour to others. For the 8th factor, 1 item loaded at .5 ("More than anything else, I want to be remembered as a compassionate person"), with multiple other "empathy valuing" items loading above .32. For the 9th factor, one item loaded at .4, with multiple other items loading above .32; these items were about general lack of interest in the lives/experiences of others. Importantly, the potential content domains for the 7th, 8th, and 9th factors all had a significant number of items written for them: the weakness of these factors was based on most of the items for these potential domains loading outside of these factors, not due to lack of potential items in the pool. This EFA model indicated only 6 factors from which reliable scales could potentially be formed.

In the basic SL bifactor model, using minimum residual factor method and an oblimin rotation (default for omega function in PSYCH package), the ωH for the general factor was .83 and the ECV was .64, indicating a moderately strong degree of unidimensionality throughout the item pool. Six specific factors, though, had at least 10 items above .3 (all with also at least 6 items above .4). These six factors were highly interpretable. In order of extracted eigenvalues, they were: a confidante listening factor (ω HS = .18); empathic avoidance factor (ω HS = .25);

appetitive contagion factor (ω HS = .33); perspective-taking factor (ω HS = .39); empathic absorption factor (ω HS = .26); and an empathic restraint factor (ω HS = .29).

The 7th specific factor had 3 items loading above .3, but one item cross-loaded on the confidante listening dimension. These were three items about feeling positive emotions when helping others, but the ω HS was only .16. It does not seem that this factor is capable of producing a strong scale above and beyond the general factor.

The 8th specific factor had 5 items loading above .3. The content was superficially interpretable, as the items were primarily written for the hypothetical "empathy valuing dimension," but the ω HS was only .16, indicating that this factor is particularly difficult to distinguish from the general factor. I decided that the evidence, overall, did not support this factor as a psychometrically useful specific dimension.

The 9th specific factor had 5 items loading above .3, none of which loaded above .4. These items related to general disinterest in the lives/experiences of others, but the ω HS was only .12. I decided that this factor was not capable of producing a strong scale and is especially difficult to distinguish from the general factor. In sum, I concluded that an SL EBFA with 9 factors was likely an over-extraction, pragmatically speaking. Comparing the bifactor results with the results from the correlated factors model corroborated this view.

EFAs extracting 8 factors. In the EFA model, using minimum residual factor method and an oblimin rotation, all factors were substantially inter-correlated (*r*s from .21 to .52) and 6 factors had at least 5 items loading above .5. These six factors were highly interpretable. In order of extracted eigenvalues, they were: a confidante listening factor; perspective-taking factor; empathic avoidance factor; appetitive contagion factor; empathic absorption factor; and empathic restraint factor. No items loaded above .5 on the 7th factor. Only two items loaded, and only

barely so, above .5 on the 8th factor (VL1, VL2). These two items both relate to feeling good as a result of showing caring behaviour to others. This EFA model indicated only 6 potentially reliable factors, with 8-factors likely representing an over-extraction of the substantive dimensions within the data.

In the basic SL bifactor model, the ω H for the general factor was .83 and the ECV was .64, indicating a moderately strong degree of unidimensionality throughout the item pool. Six specific factors, though, had at least 10 items above .3 (all with also at least 6 items above .4). These six factors were highly interpretable. In order of extracted eigenvalues, they were: a confidante listening factor (ω HS = .20); empathic avoidance factor (ω HS = .24); appetitive contagion factor (ω HS = .32); empathic absorption factor (ω HS = .26); perspective-taking factor (ω HS = .37) and an empathic restraint factor (ω HS = .30).

The 7th specific factor had 6 items loading above .3, but none were above .4 and 3 cross-loaded nearly as highly on the previously mentioned interpretable factors. The content was superficially interpretable, as the items were primarily written for the hypothetical "empathy valuing dimension," but the ω HS was only .13, indicating that this factor is particularly difficult to distinguish from the general factor. I decided that the evidence did not support this factor as a psychometrically useful specific dimension.

The 8th specific factor had only 3 items loading above .3, two of which loaded above .4, but one which cross-loaded nearly as highly on the 7th specific factor. These three items (VL1, VL2, VL11) referenced feeling good as a result of empathically caring for others. The ω HS was only .17. I decided that this factor was not capable of producing a scale and is also difficult to distinguish from the general factor. In sum, I concluded that an SL EBFA with 8 factors was

bordering on over-extraction. Comparing the bifactor results with the results from the EFA model corroborated this view.

EFAs extracting 7 factors. In the EFA model, extracting 7 factors, the factors were all substantially intercorrelated (*r*s from .26 to .52) and 6 factors had at least 6 items above .5. These six factors were clearly interpretable and were essentially the same 6 strongest factors as in the 8-factor model. The 7th extracted factor had no items loading above .5. In sum, a 7-factor solution appears to be an over-extracted solution, in terms of substantive dimensions from which to fashion scales.

In the basic SL bifactor model, the ω H for the general factor was .83 and the ECV was .65, indicating a moderately strong degree of unidimensionality throughout the item pool. Six specific factors, though, had at least 10 items above .3 (all with also at least 6 items above .4). These six factors were highly interpretable and were essentially same six strong factors observed in the 8-factor SL model. In order of extracted eigenvalues, they were: a confidante listening factor (ω HS = .20); empathic avoidance factor (ω HS = .25); appetitive contagion factor (ω HS = .31); perspective-taking factor (ω HS = .38); empathic absorption factor (ω HS = .25); and an empathic restraint factor (ω HS = .29).

The 7th factor had 11 items above .3, but the highest loading was only .36 and the ω HS was only .14. Most of the items had salient cross-loadings on other factors; the few items that did not were items from the hypothesized domain of "Concern for the Happiness of Others." I concluded that this factor was essentially only a minor variant of the general factor, not a psychometrically meaningfully distinguishable specific factor.

EFAs extracting 6 factors. In the EFA model, extracting 6 factors, the factors were all substantially correlated with one another (*r*s from .28 to .52) and 6 factors had at least 7 items

above .5. These six factors were clearly interpretable and were essentially the same 6 strongest factors as in the 7-factor and 8-factor models. This indicates that a six-factor model does not under-extract any reliable substantive dimensions.

In the basic SL bifactor model, the ωH for the general factor was .83 and the ECV was .66, indicating a moderately strong degree of unidimensionality throughout the item pool. Six specific factors, though, had at least 12 items above .3 (all with also at least 6 items above .4). These six factors were highly interpretable and were essentially same six strong factors observed in the 9-factor, 8-factor, and 7-factor SL models. In order of extracted eigenvalues, they were: a confidante listening factor (ω HS = .19); empathic avoidance factor (ω HS = .26); empathic absorption factor (ω HS = .25); appetitive contagion factor (ω HS = .30); perspective-taking factor (ω HS = .38); and an empathic restraint factor (ω HS = .20). The basic SL six-factor model appears to neither over-extract nor under-extract reliable substantive specific dimensions.

Nonetheless, it is worth noting that, whether extracting 6, 7, 8, or 9 factors, the confidante listening factor had the lowest ω HS of the interpretable scales. This finding indicates that, at least prior to item winnowing, this dimension is the least distinguishable from the general factor of empathic caring. In contrast, this factor also routinely had a comparatively large number of loading items, enhancing the possibility that it could be improved through item selection. I expected the item selection process would increase the distinguishability of this scale; it is possible, though, that this factor may still end up being inadequately distinguishable from the general factor to form a high-value scale.

JB EBFA with 6 correlated specific factors. Having concluded that a 6-factor extraction was most appropriate, neither over-extracting nor under-extracting reliable substantive dimensions, I also conducted a JB EBFA (Jennrich & Bentler, 2012) with oblique rotation for

comparison. Specifically, I wanted to examine whether specific factors may continue to be substantially correlated even after partitioning out variance attributed to the general factor.

The oblique bifactor rotation produced essentially the same 6 interpretable specific factors, and nearly the same item loadings in terms of rank, as the SL model.[33] All inter-factor correlations were less than |.10|, except the correlation between empathic avoidance and appetitive contagion (r = -.18). This finding indicates that an orthogonal approach is likely to be adequately justified, but that there may be minor discrepancies in factor loadings between orthogonal and oblique methods. There was, however, one extremely important difference between the two different bifactor models. In the basic SL model, whether extracting 6, 7, or 8 factors, only one item (TQ11) had salient loading solely on the general factor. This is roughly expected, given the assumptions relied upon in the SL method; SL transformations appear to generally fail to recognize pure indicators, even when other methods clearly reveal them (see results in Abad et al., 2017). In the JB EBFA with oblique rotation, however, most of the items loading highest on the general factor had no salient loadings on the six specific factors.

Although the basic SL method will generally be reasonably accurate and is the most accurate method if data conforms to a perfect independent clusters structure with no pure indicators, simulation studies by Abad et al. (2017) indicate that the accuracy of its parameter estimates decreases dramatically in the combined presence of both pure indicators and cross-loadings. In such cases, the SLi method (Abad et al., 2017), however, is highly accurate (but this superiority has only been tested in orthogonal situations). Although the superiority of the SLi method in this regard has been based on using a specification cutoff for specific factor loadings of .2, in some cases, a cutoff of .05, which will tend to require more automatic iterations, will

[33] Extracting 6 specific factors, the JB EBFA with oblique rotation failed to converge after 1000 iterations (the default setting). I increased the maximum iterations to 2,000 and the rotation converged.

provide more accurate estimates (Garcia-Garzon, Abad, & Garrido, under review). For the fine-grained item selection process for each specific factor scale, I relied upon a combination of these two SLi method variants, with JB EBFA examined in cases where the two variants disagreed.

Selecting potential items for scales. For broad personality self-report scales, Soto and John (2019) have demonstrated that a "sweet spot" for efficiency is generally found between 6-9 items for a scale, with external validity decreasingly rapid with scales shorter than 6 items and increasingly only marginally for each item above 9 items. Nonetheless, "in general,…broader content areas should be represented by more items than narrower ones" (Clark & Watson, 2019, p. 4). For all specific content scales, I aimed to have between 6-9 final items, provided that enough items were available to serve as strong indicators. Nonetheless, given that these specific content scales will presumably be narrow in substantive content, it may be acceptable to have scales with fewer items, so long as the items are strongly inter-correlated and are robust markers of the construct in question (in other words, moderately redundant, e.g., McGrath, 2005). For the general factor of empathic caring (GFEC) scale, I opted to maximize reliability and validity by aiming for more than 9 items, even though each additional item would only marginally improve the scale. This is because (a) the scale, by its nature, is the opposite of narrow in substantive content and (b) my goal was for it to function as a robust substitute for the general factor of the 137-item pool.

For the general factor scale, my aim was to select extremely high loading items that did not exhibit salient cross-loadings (.20 or higher) on specific factors; I hereafter refer to such items as "quasi-pure indicators" of the general factor. In devising the general factor scale, only the bifactor models were utilized.

For the bifactor models, I relied upon the SLi method variants. I conducted the SLi method, extracting 6 specific factors, using both a .20 loading specification cutoff and a .05 loading specification cutoff. Although the two methods produced highly similar results, there were occasional discrepancies regarding salient cross-loadings on specific factors. When such disagreements occurred between the two methods, I examined loadings in the JB EBFA with oblique rotation; although the specific factors were largely orthogonal, the minor correlations between some of them could influence cross-loadings when forced into orthogonality.

Preliminary General Factor of Empathic Caring (GFEC) Scale. For both the .05 cutoff and .20 cutoff SLi models, I selected the top 20 quasi-pure indicators in each model, all of which had general factor loadings of at least .76. Excluding items in which the two cutoff methods disagreed (i.e., a quasi-pure indicator identified by one cutoff method had a salient specific factor loading in the other cutoff method model), 19 items were identified for potential inclusion. In cases where there were disagreements between the two methods, I included 5 additional items where (a) the potential salient cross-loading in the disagreeing model was only barely above .20 and (b) the JB EBFA with oblique rotation indicated no salient cross-loadings. I then examined the 24 items under consideration, both psychometrically and conceptually, to select final scale items.

I dropped Q270, as it had nearly the same item content as Q269 but with slightly more complicated wording. I dropped Q173, as it had nearly the same item content as Q170, but with slightly more complicated wording. I dropped Q174 as it had nearly the same content as Q175; there was no clear difference in wording difficulty, and their loadings were nearly largely indistinguishable, so I selected one based on subjective estimation of its relation to the overall

construct. The remaining items had an ordinal α of .98, which would not have been increased by removing any items (all items correlated with the scale, corrected for overlap, above .7).

I then conducted two different IRT procedures on the items, to help winnow the scale down. Based on prior research (Murphy et al., 2018a), I expected that most empathic caring items would have good measurement precision at low and average range levels of the latent trait, but would struggle to provide information at high levels of the trait.

I conducted both a unidimensional Grade Response Model (GRM; Samejima, 1969) IRT analysis of the 20 items by themselves and also a multidimensional 2PL IRT of the full 137 item pool (using the JB EBFA oblique rotation). I then examined the item information of each of the 20 items on the general factor at high levels of the latent trait. Almost all items struggled to provide information at very high levels of the latent trait in both methods, but the item pool provided a strong degree of information from very low to modestly high trait levels. Both methods indicated that three items in particular (VL4, CQ1, and Q267) had low item difficulties and provided extremely little information at higher levels of the latent trait. I dropped both items, given that the lowest ranges of the latent trait were effectively measured by the remaining items as a whole. Two other items had the lowest discrimination values of remaining items and did not add noticeable information at higher levels of the trait, so I also dropped them.

The preliminary GFEC scale (15 items) had an ordinal α of .97, which would not be improved (or noticeably worsened) by dropping any of the items. All items correlated with the scale above .72 (corrected for item overlap). Although Parallel Analysis indicated 2 factors, Velicer's MAP, VSS1, and empirical BIC indicated only one factor. Only one factor was above an eigenvalue of 1. The preliminary GFEC scale, in a unidimensional GRM IRT analysis, had good reliability (see Appendix C) from very low (-3 SDs) to moderate ranges (up to 1 SD) of the

latent trait (from .88 to .92), and acceptable reliability at 2SDs above the mean (.73), but almost no reliability at 3SDs above the mean of the trait (.06). In other words, although this scale appears capable of adequately reliably measuring empathic caring up to relatively high levels of the latent trait, it cannot be expected to reliably discriminate at the most extremely high levels. For comparison, the test information function for the TEQ is also presented in Appendix C; the GFEC scale has stronger measurement precision overall, but only minimally improves beyond the TEQ at higher trait levels. Nonetheless, this investigation constitutes biased "double-dipping" given that the properties of the scale were examined in the same dataset used to create the scale; Study 2 will provide a more reliable, unbiased comparison of the two similar scales.

The total score for the GFEC scale was highly correlated (r = .94) with the general factor score from the 137-item pool (saved from the earlier JB EBFA extracting 6 specific factors). Both conceptually and psychometrically, it appears to be a strong, brief approximation of the general factor from the overall item pool. As a result, it can likely be used with confidence as such in future research studies, alleviating much of the need of administering such an extensive item pool in future studies.

Furthermore, the total score for the GFEC scale was highly correlated (r = .89) with the TEQ total score. This result indicates that this scale is highly similar to a unidimensional scale fashioned from a substantially smaller item pool, one which included many items not included in this current dissertation study.

Importantly, though, this scale differs from the TEQ in that it is designed to *intentionally* minimize specific factor content. For comparison, for the TEQ, Horn's parallel analysis indicated 5 factors (only one above an eigenvalue of 1), whereas Velicer's MAP indicated 1 factor and the empirical BIC indicated 3 factors. Furthermore, in the JB EBFA of the 137-item pool, 2 TEQ

items loaded on the appetitive contagion factor and 2 TEQ items loaded on the empathic avoidance factor. In other words, the TEQ, which was derived from a basic unidimensional EFA, appears to be modestly but meaningfully heterogeneous, in comparison to the General Factor Empathic Caring scale. Nonetheless, the other 12 items of the TEQ had no salient loadings on specific factors, though all loaded strongly on the general factor. This illustrates the similarities between the two approaches to scale construction: both will tend to value and retain quasi-pure indicators, valuable items that multidimensional EFA approaches will often discard.

Selecting potential items for specific content scales. For specific factor scales, my aim was to select strongly loading items that did not have salient cross-loadings on other specific factors. In devising the specific factor scales, I first chose items in the 6-factor EFA model that loaded above .5 on a factor, but did not load above .32 on another factor.

After culling items based upon EFA, I then relied upon bifactor models to further reduce the item options. If an item was identified as a strong indicator (loading above .3, no cross-loadings above .2) in both SLi cutoff methods, it was automatically retained. If the SLi methods disagreed, I still included the item if the JB EBFA with oblique rotation supported it as a strong indicator.

Preliminary "Perspective-Taking" (PT) scale. Of the 12 items selected in the EFA model, the two SLi cutoff methods did not disagree on any items and supported the selection of 10 of the items. I removed the three items with noticeably weaker correlations with the scale than the other items, all of which had similar content that appeared to be more related to rational open-mindedness rather than empathic functioning per se (PT2, PT5, QC6). Two other items appeared to be nearly identical in content, even though they were not identified as psychometric duplicates in the initial screening; I removed the item with the slightly more complicated

wording (Q163). The remaining 6 items all correlated with the scale above .71, producing an ordinal α of .93.

Preliminary "Confidante Listening" (CL) scale. Of the 30 strong indicators observed in the EFA model, the two SLi methods agreed on 8 items. One item (Q192) upon which they disagreed was initially included based upon the JB EBFA with oblique rotation. Nonetheless, I dropped the item because it had a noticeable lower correlation (corrected for item overlap) with the scale, in addition to not being agreed upon between the SLi methods. This item inherently requires the respondent to judge how others viewed them, in comparison to others; this content, as a result, is much more complicated and relies on a very heightened degree of social awareness, compared to other items. I then dropped Q198, which had the next lowest correlation with the scale and was the lowest or second lowest loading identified indicator item in both of the SLi methods. This item related to "asking gentle questions," whereas the other items in the scale did not specify particular listening techniques. The remaining 7 items had an ordinal α of .94 and all were correlated (correcting for overlap) above .70 with the scale.

Preliminary "Appetitive Contagion" (AC) scale. Of the 12 items selected from the EFA model, the two SLi method agreed upon 4 of them. Another six items were also included, despite SLI disagreement, because the JB EBFA with oblique rotation supported them. I first examined the item contents and removed three items (TQ3, Q237, Q244) whose content might be considered by some researchers to not match with the narrow definition of isomorphic empathy proposed by some researchers (e.g., Jordan, Amir, & Bloom, 2016), as the items did not specify an emotional mirroring of observer and observed. All three were items upon which the SLi methods disagreed.

The remaining 7 items produced an ordinal α of .92, which would be decreased if any items were removed. One item (PN3), though, had a noticeably lower correlation with the scale (corrected for overlap); examining its item content it seemed that the item might be related not just to appetitive contagion but also to fondness for babies. I removed the item. The remaining 6 items had an ordinal α of .91, which would be decreased if any items were removed, and all items correlated with the scale (corrected for item overlap), above .66.

Preliminary Empathic Absorption (EAB) Scale. Of the 6 items selected in the EFA model, the two SLi cutoff methods agreed upon all 6 of them. One item (PN2) had substantially different item content than the others and appeared to be a stray measure of unvalenced general contagion. This is a potential sign that this scale might closely resemble the unvalenced contagion content type that I attempted to remove from the initial item pool based on prior research indicating appetitive and aversive contagion should not be conflated (Murphy et al., 2018b). I removed this item. The remaining 5 items had an ordinal α of .89, which would be very slightly improved by dropping one item (Q78), which poorly correlated with the scale (.6), so I dropped it. The remaining 4 items had an ordinal α of .89 and correlated with the scale from .68 to .78. This scale needed serious improvement in a second phase of data collection, unless it is dropped altogether.

Preliminary Empathic Avoidance Scale. Of the 11 items selected in the EFA model, the two SLi cutoff methods agreed upon 6 items, and the JB EBFA with oblique rotation indicated that three additional items could potentially be included. One item (EQ4) had content that seemed to resemble self-report CE or lack of sensitivity more than the empathic avoidance content domain represented in the other items, so I removed it. With the remaining 8 items, 1 item (TQ9) had a noticeably lower correlation with the scale (.67) than the other items and its

content was also noticeably different, referencing being irritated at people who cry rather than empathic avoidance. I removed the item. The remaining 7 items had an ordinal α of .94, and all were correlated with the scale (corrected for item overlap) above .71.

Preliminary Empathic Restraint Scale. Of the six items selected in the EFA model, the two SLi cutoff methods agreed upon only 2 items, but the JB EBFA with oblique rotation indicated that two additional items could potentially be included. Of the two items from the EFA model not supported by the bifactor models, one (Q268) had content that seemed very unrelated, while the other (Q185) had directly related content. I decided to examine the degree to which it would assist with scale reliability. It correlated with the scale well (.71) and the ordinal α would be substantially impaired if it were dropped, so I provisionally kept in in spite of its lack of support in the bifactor models. If it performs poorly in the second phase of data collection, it will be dropped. Together, the five items produced an ordinal α of .89, and all were correlated with the scale at least at $r = .64$. It is clear that further effort is needed in phase 2 of data collection to improve this scale.

Factor analyzing the preliminary 35 specific content scale items. In reducing from 137 items to only 35 items for specific content scales, while also removing quasi pure indicators of the general factor (e.g., the GFEC scale items), it was possible that the factor structure would shift. As a result, I conducted additional factor analyses with just the 35 items selected for preliminary scales. Horn's parallel analysis, Velicer's MAP, and the empirical BIC all suggested 6 factors.

I first conducted a EFA and a basic SL bifactor model, to investigate whether all preliminary scale items (a) loaded reliably on their intended factors and (b) did not have salient cross-loadings on other factors. In the EFA model, extracting 6 factors, all items loaded above .5

on their intended factors and no items cross-loaded above .3 on another factor. The closest to a salient cross-loading was the confidante listening item Q199, which cross-loaded at .26 on the empathic absorption factor. In a basic SL bifactor model, all items loaded above .3 on their intended factors, but the confidante listening item Q199 cross-loaded at .20 on the empathic absorption factor. I removed Q199 from further scale analyses in this study, but I hypothesized that the cross-loading concern would be alleviated in phase two once additional items were added to help clarify and solidify the empathic absorption scale.

I conducted both factor analytic methods again, excluding Q199. In the EFA model, all items loaded above .5 on their intended factors, with no salient cross-loadings. Inter-correlations between the factors ranged from $r = |.37|$ to $r = |.61|$.

For a graphical illustration of the SL bifactor model of these 35 specific content items, see Figure 3. In the basic SL bifactor model, all items loaded above .04 on their intended factors, except for Q167 which loaded at .38 on its intended factor. There were no salient cross-loadings. The ECV of the general factor was .60, with a PUC of .85 and an average relative parameter bias of .07. These indices, in conjunction with the high ωH of the general factor, indicate that a total score of all the items might be useable as a unidimensional measure.

For bifactor indices, see Table 4. In comparison with the average indices of the 50 "well-fitting" multidimensional bifactor models examined by Rodriguez, Reise, and Haviland (2016b), these specific factor scales appear to: (a) have slightly higher overall reliability, both as raw scores and as latent factor estimates; and (b) provide slightly more specific factor reliability based on non-general content. Nonetheless, most of the variance in each of the scales is accounted for by the general factor, and researchers will need to be careful in utilizing and

interpreting them. Moreover, given that these indices represent biased "double-dipping," indices may shift substantially in Study 2.

Generating "empathy-related negative emotionality" specific content scales. The initial 34 item cluster of "empathy-related negative emotionality items" had been reduced to only 24 items, as described in the earlier section detailing the ICLUST analyses. With the remaining 24 items, Velicer's MAP, parallel analysis, and the empirical BIC all indicated 3 factors. In a EFA, three clearly interpretable factors were obtained: an empathic distress/discomfort factor, a contagion for anxiety/fear factor, and a contagion for sadness factor. Although the two contagion factors were strongly correlated (r = .70), neither was strongly correlated with the empathic distress/discomfort factor (both below absolute value of r = .30).

The empathic distress/discomfort factor contained 9 items loading above .5, none of which had cross-loadings above .3. I kept the 7 items with loadings above .7. The resulting scale had an ordinal α of .92, and all items correlated with the scale (corrected for overlap) above .62. The contagion for sadness factor contained 4 items loading above .5, none of which had cross-loadings above .3. These 4 items, which formed the scale, had an ordinal α of .89, with all items correlated with the scale above .61.

The contagion for anxiety/fear factor contained five items loading above .5, none of which had cross-loadings above .3. These five items formed the scale (ordinal α = .91), with all items correlated with the scale above .64.

Nomological Networks Results

Inter-relationships among preliminary scales. Full intercorrelations between various empathic caring and empathy-related negative emotionality scales, at both the zero-order level and at the unique variance level (residuals after regressing on the GFEC scale) are presented in

Table 5. The general factor (saved factors scores from JB EBFA) of the 137-item pool of empathic caring was extremely similar to the TEQ total score, indicating that it closely parallels the "common factor" of empathy analysed in the smaller item pool of Spreng et al. (2009). It was very strongly correlated with all six of the specific content scales constructed from the same 137-item pool (all *r*s above |.7|). It was also strongly negatively correlated with the preliminary EID scale (r = -.58) and strongly positively correlated with the preliminary SADCON scale (r = .54). It was only weakly positively correlated, however, with the preliminary ANXCON scale (r = .21). This pattern of results indicates that, although contagion for the sadness of others might be strongly related to the broader empathic caring construct, contagion for anxiety may be much more of a separate construct.

The GFEC scale and the TEQ scale exhibit essentially the same patterns of correlations with the 9 preliminary specific content scales as were observed with the general factor scores of the JB EBFA in the 137-item pool. There were only slight differences between the two scales, with the GFEC more positively correlated with CL (Steiger's z = 19.2, p < .001, df = 2) and the TEQ more strongly correlated with EAV, ER, EID, ANXCON, and SADCON (Steiger's *z*s from |4.7| to |20.5|, *p*s all < .001).

In terms of zero-order correlations, all six of the preliminary specific factor scales formed from the 137-item empathic caring pool were strongly correlated (absolute value of *r*s from .50 to .65), as should be expected given that all of them share extremely strong correlations with the measures of general empathic caring. When comparing only their specific factor contents, though, as operationalized through their residual variances unique from the GFEC, these six preliminary scales were non-significantly or minimally related, with only the relationship between AC and EAB stronger than r = |.2|.

ANXCON and SADCON were strongly positively correlated (*r* = .64), but, surprisingly, neither was substantially correlated with EID at the zero-order level (*r*s = .25 and -.04, respectively). Interestingly, though, their residualized variances with EID were stronger than seen in the zero-order correlations (*r*s = .41 and .30, respectively). In other words, the general factor of empathy appears to operate as a suppressor variable, obscuring the strength of the relationship of the two forms of aversive contagion with empathic intimacy discomfort. Consistent with this observation, the GFEC also suppressed their relationships with the CL scale. At the zero-order level, anxiety contagion was not related to confidante listening (which involves supportively listening to the needs and feelings of others) and sadness contagion was positively associated with it; at the residualized level, though, both were similarly negatively correlated with the CL scale (*r*s from -.18 to -.21).

Two areas of findings were particularly concerning. First, at the residual level, the EAB scale was not correlated with any of the other specific content scales, except that it was modestly to strongly correlated with the AC scale (*r* = .27), the ANXCON scale (*r* = .39), and the SADCON scale (*r* = .52). This finding raises the suspicion that the empathic absorption construct, which is not emotionally valenced in its item contents, may be (a) difficult to distinguish from emotional contagion in general and (b) may, therefore, fail as a construct measure by inadvertently conflating appetitive and aversive absorption components into a single dimension.

Second, the strong correlations between EAV and EID at both the zero-order level (*r* = .76) and at the residualized variable level (*r* = .65) suggest that the two constructs may struggle to differentiate themselves from one another. In other words, the tendency to avoid empathic interactions may be so strongly linked to feeling discomfort or distress in empathic interactions

that the two constructs, avoidance behaviors and distressed feelings, cannot be soundly parsed apart in a self-report measure. Nonetheless, it is worth noting that, at the residualized level, EID is substantially associated with aversive contagion, whereas EAV is not, which is consistent with the conceptual differences between the two constructs and the separate analytical processes used to generate the two scales (EAV in the 137-item empathic caring pool and EID in the 35-item negative affect related pool).

Relationships with HEXACO broadband personality dimensions. In Murphy et al. (2018a), we observed that IRI EC and ACME AR, both of which are similar to the general factor of empathic caring, were positively correlated with all HEXACO dimensions except for HEXACO NSE. As a result, I predicted that the GFEC and TEQ would display similar associations. I predicted, however, that the empathy-related negative emotionality scales (EID, ANXCON, and SADCON) would relate substantially more strongly to HEXACO NSE than to any other broadband personality dimensions (consistent with associations observed for "distress contagion" in Murphy et al., 2018b).

Given that they conceptually represent the strongest links to disagreeable behavior, I predicted that the PT and ER scales would demonstrate particularly meaningful incremental value in associating with HEXACO Agreeableness (consistent with associations for the IRI PT scale observed by Murphy et al., 2018a). I predicted that the EAB scale, like the IRI FN scale which was its conceptual origin, would relate more strongly to HEXACO Openness than to any other domain, but would also relate positively to NSE (based on correlational patterns for the IRI FN observed by Murphy et al., 2018a). I predicted that the AC scale would add particular meaningful incremental validity in relating to extraversion, but perhaps also to NSE (the two strongest associations observed for "positive/neutral contagion" in Murphy et al., 2018b). Given

that the CL and EAV scales are particularly novel for empathy scales, I did not have strong grounds for making predictions about the HEXACO associations for their narrow contents.

All correlations between empathy scales and HEXACO dimensions, at both the zero-order and residualized variable levels, are presented in Table 6. The general factor of the 137-item empathic caring pool was positively correlated with all six HEXACO dimensions (*r*s from .24 to .49). Both the GFEC and the TEQ demonstrated nearly the same HEXACO correlations as the general factor, further supporting the general equivalence of the three measures. The only notable differences between the GFEC and the TEQ were that the GFEC was slightly more strongly correlated with extraversion (Steiger's $z = 5.5$, $p < .001$, $df = 2$), while the TEQ was slightly more strongly correlated with Honesty-Humility, NSE, and Openness (Steiger's *z*s from 2.3 to 3.0 , all *p*s < .01).

In examining the nomological networks of the specific content scales, I focused on the residualized variables, as they (a) better represent the narrow content of the scales and (b) indicate the extent to which the specific content domains display predictive value above and beyond the general factor. The PT residual offers substantial incremental value in relation to Agreeableness ($r = .36$), consistent with my predictions. It's residual also has a small negative correlation ($r = -.14$) with NSE, modestly consistent with my predictions; this relationship, though, is suppressed by the general factor content (β = .23, 99% CI [.14, .32]).

The CL residual has only small correlations with the HEXACO dimensions (all *r*s < .2). Like the PT residual, it has a small negative correlation with NSE ($r = -.16$) which is suppressed by the general factor content (β = .33, 99% CI [.21, .45]).

The EAV residual has only weak correlations with the HEXACO dimensions, with its strongest relationship being with Honesty/Humility ($r = -.21$). The AC residual is non-

significantly or negligibly correlated with all HEXACO dimensions except for Extraversion (r = .24). The ER residual has meaningful positive correlations with both Honesty/Humility (r = .22) and Agreeableness (r = .24), modestly consistent with my predictions. Interestingly, its small negative correlation with Extraversion (r = -.18) is suppressed by the general factor (β = .41, 99% CI [.31, .51]), indicating a hidden weak relationship between being introverted and being non-aggressive.

Consistent with concerns raised about the EAB scale in regard to its associations with the other empathy-related scales, the EAB residual was only meaningfully positively associated with NSE (r = .29). Contrary to my predictions, it was not associated with Openness. Furthermore, the EAB residual demonstrated negative correlations with Honesty/Humility, Extraversion, Agreeableness, and Conscientiousness (rs from -.10 to -.15). In other words, the specific content of the EAB scale appears to relate primarily to negative emotionality, with small relationships with other broadband personality dimensions that are in the opposite direction of that found with general empathic caring. Importantly, its negative associations with Honesty/Humility, Extraversion, Agreeableness, and Conscientiousness were all significantly suppressed by the GFEC (all βs above .33).

For most HEXACO dimensions, the EID residual demonstrated roughly the same pattern of associations demonstrated by the EAV residual. Whereas the EAV residual was not associated with NSE or Extraversion, the EID residual was meaningfully positively associated NSE (r = .28) and meaningfully negatively associated with Extraversion (r = -.21). EID's positive association with NSE was suppressed by the GFEC (β = -.21, 99% CI [-.28, -.14]).

ANXCON residual and SADCON residual both demonstrated substantial positive associations with NSE (rs were .65 and .49, respectively), indicating that both specific contents

relate far more to broadband negative emotionality than to other aspects of personality; this pattern was also clearly demonstrated at the zero-order level. Neither was associated with Openness. ANXCON residual was negatively associated with Honesty/Humility, Extraversion, Agreeableness, and Conscientiousness (*r*s from -.17 to -.31). Similarly, SADCON residual was negatively associated with Extraversion, Agreeableness, and Conscientiousness (rs from -.10 to -.23). In other words, the specific contents of both forms of aversive contagion were generally weakly or modestly associated with interpersonal disconnection, while being substantially associated with negative emotionality.

The GFEC significantly suppressed ANXCON's associations with all HEXACO dimensions except NSE and Openness (all significant βs above .07). The GFEC significantly suppressed SADCON's associations with Extraversion, Agreeableness, and Conscientiousness (all significant βs above .20).

Study 1 Discussion

In Study 1, I used a range of techniques (ICLUST, EFAs, EBFAS, etc.) to distill a collection of more than 200 empathy questionnaire items into one general factor of empathic caring scale (the GFEC scale) and nine specific content scales. Subsequent investigation indicated that one specific content scale ("empathic absorption") was ill-conceived and should be dropped, whereas multiple others scales have potentially serious issues in regard to their distinguishability from other scales. Nonetheless, nomological network analyses are encouraging; newly created scales may offer substantial value in measuring meaningful heterogeneity in the broad empathy construct. Moreover, initial results from Study 1 have important broader conceptual implications for empathy research.

Relationships between empathic caring and emotional contagion. The initial ICLUST analysis of the initial 184-item pool indicated that most items could be clustered into a broad, overarching cluster representing general empathic caring (or a small cluster which was strongly correlated with this overarching cluster). A substantial minority of items, however, all related to experiencing aversive emotion in empathic encounters, clustered separately, with this cluster only weakly correlated with the dominant general empathic caring cluster. These ICLUST results corroborated and extended prior research (e.g., Murphy et al., 2018b) indicating that (a) appetitive contagion and aversive contagion should not be conflated into a single dimension, (b) aversive contagion is highly distinguishable from general empathic caring, and (c) appetitive contagion is strongly related to general empathic caring.

These ICLUST results, in conjunction with prior research, led me to analyze most of the item pool together in a bifactor-minded approach, with specific content dimensions understood as narrow facets of a broad general empathic caring construct. I analysed the items referencing aversive emotion in empathic encounters separately, using only basic correlated factors methods. The two aversive contagion scales (one for sadness contagion and another for anxiety contagion) resulting from this smaller, separate item pool subsequently demonstrated stronger associations with trait negative emotionality than with general empathic caring.

Evaluating the General Factor of Empathic Caring (GFEC) scale. I formed a scale of 15 “quasi-pure” indicators of the general empathic caring construct, employing multiple bifactor methods as well as IRT analyses to select the final items. This GFEC scale measures the “core” of the general factor of the broader empathic caring construct, while intentionally excluding items that loaded substantially on specific factor dimensions. As a result, though facially broad and heterogeneous in its item contents, it should typically be more unidimensional than similar

general empathy scales (e.g., the TEQ) that have been created without bifactor modelling and the identification of pure indicators. Moreover, because this scale intentionally avoids specific factor content items and therefore will tend to have less built-in overlap with specific content domains, it is conceptually better equipped for use in incremental validity analyses of narrower or alternative empathy measures, whether the specific content scales in these dissertation studies or other questionnaires that might be developed in future research endeavors.

Nonetheless, this GFEC scale is extremely strongly correlated with both the general factor scores obtained in the 137-item pool of empathic caring items as well as with the TEQ scale. Despite minor dissimilarities, the three different methods of measuring "general empathic caring" appear to be roughly equivalent. Combined with the very strong correlation between the GFEC and the TEQ, their nearly identical patterns of associations with the specific content scales indicate that they may be mostly indistinguishable in their overall substantive content. One potential implication of this finding is that unidimensional treatments of empathy item pools may tend to generate nearly the same substantive content interpretations as interpretations of general factors in bifactor treatments of empathy item pools, even when working with very different pools of items. This is, of course, simply a demonstration of what has been repeatedly argued in regard to "multidimensional" item pools with high omega hierarchical values (e.g., Rodriguez, Reise, & Haviland, 2016).

Although I intentionally wrote items to be higher in difficulty and also employed IRT analyses to aid in item selection, the GFEC only modestly improves upon the measurement precision of the TEQ at higher trait levels. As observed with other empathy scales in prior research (Murphy et al., 2018a), the GFEC scale has poor measurement precision at high levels of the latent trait.

Evaluating the nine specific content scales. The PT scale measures self-reported tendency to sympathetically take the perspective of others with whom he or she has conflicts or frustrations. It is narrow in content compared with the IRI Perspective-Taking scale and other similar scales, given that those scales do not primarily focus on conflict/frustration situations. Though strongly related to general empathic caring, the PT scale is psychometrically distinguishable from the general factor of empathic caring and provides substantial incremental value in predicting trait agreeableness.

The CL scale, which measures a self-reported tendency to be a supportive, comforting, non-judgmental listener, may provide insufficient incremental value above and beyond the general factor of empathic caring. In the initial SL EBFAs extracting 6-9 specific factors, in comparison with the other 5 specific factor scale domains, it was consistently the least distinguishable in terms of ω HS. Although it demonstrated adequate distinguishability in the SL EBFA of the final 30 items comprising the 6 specific factor scales, the CL scale was not only particularly highly correlated with the GFEC (r = .78) but also demonstrated only minimal correlations with HEXACO dimensions at the residual level. I decided to include it in Study 2, to make a final determination as to whether it should be offered as a valuable specific factor scale.

Interestingly, the CL content domain has been strongly overlooked in past empathy questionnaire construction attempts. Of the 30 items identified as strong indicators of this dimension in the EFA of the 137-item pool, only two items (both from the ACME AR scale) were from existing measures. Regardless of whether the CL scale is ultimately retained as a specific factor scale, Study 1 indicates that confidante listening should be understood as an important aspect of the broader empathic caring construct.

The AC scale, which measures self-reported emotional contagion for the joy and pleasure of others, should be understood as a strong part of the general empathic caring construct. It is, however, psychometrically distinguishable from the general factor of empathic caring and also demonstrates incremental value in relation to trait extraversion.

The ER scale measures self-reported desire to avoid causing emotional pain to others. Although it should be understood as a strong part of the general empathic caring construct, it is psychometrically distinguishable as a specific content scale and offers meaningful incremental value in relation to trait honesty/humility and trait agreeableness. Whereas other specific aspects of empathic caring are positively associated with extraversion, the specific content of empathic restraint appears to be negatively associated with it.

The EAB scale, comprised of revised items from the IRI Fantasy scale and new items generated to align with them, appears to function as an unvalenced emotional contagion scale. Its content is primarily correlated with aversive forms of emotional contagion, but also meaningfully with appetitive contagion. In terms of broadband personality dimensions, it primarily relates to heightened negative emotionality, with other smaller relationships with personality that are in the opposite direction of those exhibited by general empathic caring.

In retrospect, given prior work demonstrating the need to parse appetitive from aversive contagion, my item revision and creation attempts should have specific appetitive and aversive emotions in the items created for this potential domain. Though it is possible that appetitive empathic absorption and aversive empathic absorption scales could theoretically be of value, the overall evidence from this study indicates that they would likely be nearly indistinguishable from the appetitive and aversive contagion scales. In other words, rather than assessing other-oriented "einfühlung empathy," these items may primarily relate to projection and, as a result, be little

more than an alternative rephrasing of contagion scales. I removed the empathic absorption construct from consideration in Study 2.

The EAV scale and EID scale, which are very strongly correlated, are difficult to distinguish from one another. If they had been formed from the same overall item pool, rather than split following the ICLUST analyses, they likely would have struggled to manifest as psychometrically separate dimensions. The EAV scale measures self-reported tendencies to avoid empathic encounters with others who are trying to express their emotional concerns. The EID scale measures self-reported tendencies to feel discomfort or distressed when in empathic encounters with others who are trying to express their emotional concerns. Although empathic intimacy discomfort can be reasonably assumed to be a major cause of empathic avoidance, it is not, at least from a theoretical standpoint, the only potential cause. Individuals may avoid engaging in empathy: (a) because it can be time and/or energy consuming (e.g., Zaki, 2014), (b) they are not very interested in the emotions of others, (c) they disfavour sad or struggling people, (d) they are too self-centered to give support to others, or (e) any number of other reasons. Nonetheless, the very strong correlations between EAV and EID indicate that discomfort/distress is perhaps the most prominent reason why individuals might avoid empathic encounters.

EAV and EID, though demonstrating similar relationships with other variables, differ from one another in terms of their nomological networks in multiple important ways. First, EAV is much more strongly associated with general empathic caring than is EID. Second, the specific content of EAV is not meaningfully associated with negative emotionality, anxiety contagion, sadness contagion, or trait extraversion; the specific content of EID, however, is meaningfully associated with all these variables. In other words, EID is fundamentally tied to negative emotionality and general interpersonal reticence, whereas EAV is not. The results of Study 1 do

not clearly indicate the correct approach to handling these two scales; Study 2 aimed to better elucidate the similarities and distinctions.

Aversive contagion manifested in Study 1 as two separate domains: anxiety contagion, measured by the ANXCON scale, and sadness contagion, measured by the SADCON scale. These two scales were strongly correlated, both at the zero-order level and also in terms of their residualized specific contents. Both were strongly tied to trait negative emotionality. Their residualized specific contents exhibited similar nomological networks, though ANXCON typically demonstrated stronger relationships with other variables. Conceptually, treating contagion for anxiety and sadness as separate constructs aligns with broader distinctions between anxiety and sadness (e.g., Clark & Watson, 1991). Psychometrically, the most striking difference between the two scales was that SADCON, at the zero-order level, was robustly positively associated with general empathic caring and its specific facets, whereas ANXCON was only modestly or negligibly associated with such constructs. In sum, despite their strong associations with one another, ANXCON and SADCON are importantly distinguishable.

Study 2: Refinement of Scales and Exploration of Nomological Networks

The overarching aims of Study 2 were to (a) confirm the general factor structure of the new scales, (b) refine the scales through consideration of new items generated to address potential weaknesses, and (c) investigate the distinguishability and incremental predictive value of the refined scales under consideration.

Methods

Participant filtering. Through Amazon Mechanical Turk, 371 participants completed the study protocol, for $2.20. The median time to completion was 27.7 minutes, which was less than the expected 30-35 minutes indicated in the recruitment advertisement and consent form.

Eight participants failed the speed filter, answering more quickly than 2 seconds per item on average. I included one open-ended query, “Write a sentence about the last movie you remember watching.” I flagged 8 participants as providing a questionable response: 3 left the item blank; 5 provided answers that seemed overly vague or minimal (e.g., “horror movie interesting”). Three of the 5 providing questionable non-blank responses also failed the squared difference procedure check, as did one of the blank-responding participants. The other four flagged participants, however, otherwise appeared to provide acceptable responding patterns and were retained in the sample given that (a) participants were not required to answer all items or (b) their responses were at least related to movies.

I used the same 10 item pairs employed in the squared difference procedure of antonym pairs in Study 1, and I scored their deviation scores in the same manner as before. I used the cutoff point chosen in Study 1 (in the deviation score itself, not in terms of sample SDs). Thirty participants failed this reliable responding filter. Ultimately, I excluded 36 subjects from analysis, leaving a total of 335 participants for analysis.

Participant characteristics. The 335 participants retained in analyses ranged in age from 19 to 71 (M = 39.2 years, SD = 12.1 years) and 52.4% identified as female. As with most M-Turk samples, participants were more highly educated than the general population; 38.5% had (only) a bachelor’s degree and 15.8% had graduate level degrees of some kind. 5.4% percent of participants identified as mixed-race/ethnicity; of those who self-identified with a single race/ethnicity, 82.7 were “white,” while other race/ethnicities were much less well represented (black or African American, 5.3%; Hispanic/Latino, 4.7%; Asian, 6.3%; all others less than 1%).

Measures administered. In addition to the preliminary empathy scales developed in Study 1, I included a small number of new items that I generated as potential candidates to

improve the psychometric properties of the specific content scales (but not the GFEC, which had no obvious limitations in Study 1).

I also included the TEQ, as in Study 1, to serve as an anchor comparison scale. For the TEQ (ordinal $\alpha = .95$), Horn's parallel analysis indicated 3 factors, only 1 of which was above an eigenvalue of 1, whereas Velicer's MAP indicated 1 factor. For the GFEC (ordinal $\alpha = .97$), Horn's parallel analysis and Velicer's MAP both indicated only 1 factor. In other words, as in Study 1, the GFEC is somewhat more unidimensional than the TEQ.

In addition to empathy scales, I also included self-report measures of: emotional distress (or emotional wellbeing); antagonism-related constructs; and schizotypal personality traits. These measures were selected to provide more fine-grained examinations of nomological networks than in Study 1, but with particular motivation based on the broadband personality dimensions most associated with empathy scales in Study 1. The emotional distress variables were motivated by the sharp contrasts between the empathy scales in relating to negative emotionality in Study 1. The antagonism-related variables were motivated by the strong relationships between some empathy scales and trait agreeableness in Study 1; at a general level, antagonism and agreeableness operate as bipolar ends of a shared dimension (e.g., Lynam & Miller, in press). Finally, I examined the multidimensional relationships between the empathy scales and schizotypal personality, given that it is (a) a relatively under-explored research area and (b) the interpersonal deficits in schizotypal personality appear to sharply differ from those in more studied personality-empathy links, such as psychopathy and autism.

As in Study 1, I analysed zero-order correlations between empathy scales and external correlates, but I focused on incremental associations between the residualized unique variances

of specific content scales, on the one hand, and the external correlates, on the other. All associations are fully presented in Tables 9, 11-13.

Emotional distress (emotional wellbeing) measures. As a measure of general anxiety, I administered the 16-item Penn State Worry Questionnaire (PSWQ: Meyer, Miller, Metzger, & Borkovec, 1990), which yields a total score (ordinal $\alpha = .97$). The relationship between general trait empathic caring and PSWQ scores is unclear due to lack of prior investigation, with one study reporting a positive association (Knight, Stoica, Fogleman, & Depue, 2019), but other studies, utilizing general trait anxiety measures similar to the PSWQ, often reporting no significant association (e.g., Fernandez, Dufey, & Kramp, 2011; Grynberg et al., 2010). Limited evidence, however, indicates that aversive contagion is positively associated with PSWQ total scores, but appetitive contagion is not significantly associated with it (Murphy et al., 2018b).

Based on prior findings, I predicted that GFEC and TEQ would be minimally related to PSWQ, but that the empathy-related negative emotionality scales would all be substantially positively related to it, especially ANXCON given that it relates directly to anxiety. Given the weak or null associations between the other specific content scales and NSE in Study 1, I predicted they would not be meaningfully associated with PSWQ.

The 5-item Satisfaction with Life scale (SLS; Diener, Emmons, Larsen, & Griffin, 1985) assesses a respondent's satisfaction with his or her life as a whole (e.g., "In most ways my life is close to my ideal"), producing a total score (ordinal $\alpha = .94$). It tends to be moderately to highly strongly correlated with measures of depression (e.g., Abdallah, 1998; Arrindell et al., 1991; Durak, Senol-Durak, & Gencoz, 2010; Swami et al., 2007). Although prior investigation has been limited, the available evidence indicates that SLS is modestly positively associated with general empathic caring (Gadermann, Schonert-Reichl, & Zumbo, 2010) or not significantly

associated with it (Hebert & Weaver, 2014). Assuming that depression is closely related to satisfaction with life, the results of Murphy et al. (2018b) tentatively suggest that appetitive contagion will be positively associated with SLS scores, while aversive forms of contagion will be negatively associated with it.

Based on prior studies, I predicted that GFEC and TEQ would be positively associated with SLS, though perhaps only weakly or moderately. I predicted that all three empathy-related negative emotionality scales (EID, ANXCON, and SADCON) would be negatively associated with it, but that AC would be positively associated with it. I also predicted that CL would be positively associated, given the link between emotionally intimate friendships and emotional wellbeing (e.g., Heinrich & Gullone, 2006).

The Forms of Self-Criticising/Attacking and Self-reassuring scale (FSCRS; Gilbert et al., 2004) measures the degree to which an individual is harshly critical or forgivingly supportive towards himself or herself. Some prior studies (e.g., Castilho, Pinto-Gouveia, & Duarte, 2015) have observed that the FSCRS conforms to a 3-factor structure: Inadequate Self (e.g., "I feel beaten down by my own self-critical thoughts"), Hated Self (e.g., "I have a sense of disgust with myself"), and Reassuring Self (I find it easy to forgive myself"). Reporting on the results of bifactor models of the FSCRS in 13 different samples, however, Halamova et al. (2018) demonstrated that the general factor of the FSCRS had an ωH above .8 in all but 4 of them (with those 4 samples with ωH between .71 and .79). In other words, it may be perfectly appropriate to report the FSCRS as a unidimensional total score.

To examine this possibility, I conducted a basic unidimensional CFA and also a bifactor CFA with uncorrelated factors, using the standard 3-factor model of the FSCRS, employing the WLSMV estimator. The single-factor model demonstrated inadequate fit (CFI = .89, TLI = .88,

SRMR = .05, RMSEA = .08, 95% CI [.07, .09]; $\chi 2 = 445.63$, df = 152, $p = .37$). The bifactor model, on the other hand, demonstrated good fit (CFI = .96, TLI = .95, SRMR = .03, RMSEA = .05, 95% CI [.04, .06]; $\chi 2 = 230.15$, df = 133, $p = .30$). However, the three scales had very weak distinguishability from the general factor (ω HS values from .07 to .2) and all items had higher loadings on the general factor than on their specific factors. The ECV was .85 and the ω H was .92, indicating that the FSCRS items are very strongly unidimensional. All but two of the items loaded above .6 on the general factor; I removed the two items that did not (FSCRS7 and FSCRS18) and created a total "Self-Compassion" score from the remaining 17 items (ordinal α = .97).

Past findings regarding the links between self-compassion and prosociality, however, are surprisingly mixed. Though some studies have observed a positive association between self-compassion and compassion or empathic caring (e.g., Neff & Beretvas, 2013; Neff & Pommier, 2013; Welp & Brown, 2014), others have reported no significant associations (e.g., Gerber, Tolmacz, & Doron, 2015; Lopez, Sanderman, Ranchor, & Schroevers, 2018; Neff, 2003). Interestingly, Marshall, Ciarrochi, Parker, and Sahdra (2019) observed that the BES Affective Empathy scale, which relates primarily to aversive contagion, is robustly negatively related to self-compassion (consistent with previous research with the IRI PD scale, e.g., Neff & Pommier, 2013). As Marshall et al. (2019) proposed, self-compassion may be more related to reduced empathic distress, and accompanying interpersonal avoidance and rejection sensitivity, than to empathic caring or compassion.

Based on prior findings, I predicted that GFEC and TEQ would be minimally or weakly positively related to self-compassion, but that the empathy-related negative emotionality scales would be negatively associated with it. Given the proposed negative associations between self-

compassion and interpersonal avoidance and distress, I predicted that AC and CL would be positively associated with self-compassion, whereas EAV (like EID) would be negatively associated with it.

Antagonism-related measures. The 29-item Buss-Perry Aggression Questionnaire (BPAQ; Buss & Perry, 1992) yields four scale scores: Physical Aggression (ordinal α = .91; e.g., "Once in a while, I can't control the urge to strike another person"); Verbal Aggression ordinal (ordinal α = .77; e.g., "My friends say that I'm somewhat argumentative"); Anger (ordinal α = .90; e.g., "I have trouble controlling my temper"); and Hostility (ordinal α = .88). The contents of the Physical, Verbal, and Anger scales appear to largely align with their scale labels, but the Hostility scale is slightly more difficult to interpret. It includes items relating to suspicion of others (e.g., "When people are especially nice to me, I wonder what they want"), bitterness (e.g., "I wonder why sometimes I feel so bitter about things"), and jealousy ("I am sometimes eaten up with jealousy").

Although one meta-analysis has reported that empathy is only negligibly related to aggression (Vachon, Lynam, & Johnson, 2014), that conclusion obscures the fact that widely used self-report empathy and aggression questionnaires have stronger negative relationships with one another (e.g., Vachon & Lynam, 2016). Nonetheless, it appears that only empathic caring is negatively related to aggression, whereas aversive contagion is either not related to it (e.g., Vossen, Piotrowski, & Valkenburg, 2015) or less strongly negatively associated with it (Vachon & Lynam, 2016).

Based on prior findings, I predicted that GFEC and TEQ would be negatively associated with all BPAQ scales, but that the empathy-related negative emotionality scales would not be substantially related to the BPAQ except for the Hostility scale, which reflects a negative

emotionality-infused hostile attribution tendency and could conceptually be positively associated, via negative emotionality, with EID, ANXCON, and SADCON. I predicted that ER and PT would both add meaningful incremental validity in negatively correlating with all BPAQ scales, given their obvious facial links to aggression.

Psychopathy, often centrally defined by relative absence of conscience and moral emotions, is perhaps the prototypical psychopathology construct associated with severe deficits in empathic caring, with most characterizations emphasizing coldheartedness or a lack of love for others (e.g., Cleckley, 1941/1988; Hare, 1993; Lilienfeld & Widows, 2005; McCord & McCord, 1964). The Triarchic Psychopathy Measure (TriPM; Patrick, 2010) is a 58-item measure of psychopathic traits that yields 3 factors: Boldness, which primarily measures agentic extraversion; Disinhibition (ordinal α = .93), which measures a heterogeneous range of impulsivity, irresponsibility, antagonism, and antisociality (e.g., "I get in trouble for not considering the consequences of my actions"); and Meanness (ordinal α = .93), which measures core psychopathic traits of callousness, cruelty, and low social closeness. Many of the items on the TriPM Meanness scale resemble items in empathy questionnaires (e.g., "I don't have much sympathy for other people" and "It's easy for me to relate to other people's emotions"). I elected not to include the TriPM Boldness scale, given that prior research (e.g., meta-analysis by Sleep, Weiss, Lynam, & Miller, in press) has generally indicated that it is negligibly related to empathy.

A meta-analytic review (Sleep et al., in press) of the TriPM's association with the IRI indicates that TriPM meanness is strongly negatively associated with empathic concern (effect size estimate, r = -.62) and moderately negatively associated with perspective-taking (effect size estimate, r = -.44). Very limited evidence indicates that appetitive contagion is strongly

negatively related to TriPM Meanness, but aversive contagion is not significantly related (TriPM Disinhibition not analysed, however; Murphy et al., 2018b).

I predicted that GFEC and TEQ would be strongly negatively correlated with TriPM Meanness and moderately negatively correlated with TriPM Disinhibition. Given its strong association with Agreeableness in Study 1, a construct strongly negatively related to psychopathy (e.g., Sleep et al., in press), I predicted that the PT scale would add meaningful incremental value in (negatively) statistically predicting both TriPM dimensions. Given that TriPM Disinhibition is substantially associated with negative emotionality, but Meanness is not (Sleep et al., 2019), I predicted that the empathy-related negative emotionality scales would add meaningful incremental value in statistically predicting TriPM Disinhibition. Given that the TriPM Meanness scale includes substantial content relating to harming others, I predicted that the ER scale would add meaningful incremental value in statistically predicting it. Based upon findings in Murphy et al. (2018b), I predicted that AC would add meaningful incremental value in (negatively) statistically predicting Meanness. I did not make specific predictions regarding the CL and EAV scales.

Machiavellianism significantly overlap with various features related to psychopathy (e.g., Paulhus & Williams, 2002); it relates primarily to (a) a willingness to harm others and engage in immoral behavior, (b) in the pursuit of power or other goals, and (c) a worldview that assumes that others are also ruthlessly willing to engage in similar behaviors (Dahling, Whitaker, & Levy, 2009). The 16-item Machiavellian Personality Scale (MPS; Dahling, Whitaker, & Levy, 2009) is composed of 4 subscales (4-factor structure confirmed in Miller, Smart, & Rechner, 2015). MPS Amorality (AMO; ordinal $\alpha = .87$) relates to the core of Machiavellianism, the willingness to cheat and hurt others in order to succeed (e.g., “I am willing to sabotage the efforts of other

people if they threaten my own goals"). MPS Distrust of Others (DO; ordinal α = .83) relates to perceiving that others are Machiavellian in their motives and behaviors (e.g., "If I show any weakness at work, other people will take advantage of it"). MPS Desire for Status (DS; ordinal α = .83) measures one's desire for wealth, power, and status, but does not specify using immoral or amoral means to achieve it (e.g, "I want to be rich and powerful someday"). MPS Desire for Control (DC; ordinal α = .82) measures one's enjoyment of controlling others (e.g., "I enjoy having control over other people").

As a unidimensional construct, Machiavellianism has generally been observed to be associated with reduced empathic caring or compassion (e.g., Andrew, Cooke, & Muncer, 2008; de Corte et al., 2007), but positively associated with trait neuroticism or negative emotionality (e.g., Stead et al., 2012; Vernon et al., 2008). At a facet level, MPS AMO has been observed to be strongly negatively correlated with a generalized form of trait empathy, whereas MPS DO has been observed to weakly negative correlate with it and MPS DC and DS appear to be unrelated to general trait empathy (Miller, Smart, & Rechner, 2015). Importantly, the cynical/hostile views of others associated with Machiavellianism have been found to be particularly strongly related to internalizing symptoms and psychological dysfunction (Monaghan, Bizumic, & Sellbom, 2016).

I predicted that GFEC and TEQ would, consistent with prior findings, be negatively correlated with MPS AMO and DO, but not exhibit substantial correlations with MPS DC and DS. I predicted that ER, which is conceptually related to avoiding manipulating or cheating others, would add meaningful incremental validity in negatively correlating with MPS AMO and DO, whereas AC would add meaningful incremental validity in negative correlating with MPS AMO and DO, given its associations with extraversion and positive emotionality. I predicted that both CL and PT would add incremental validity in negatively associating with MPS DO, given

that hostile views of others could reasonably inhibit supportive listening and sympathetic perspective-taking. I predicted that the three empathy-related negative emotionality scales, however, would positively relate to both DO and AMO, given the shared associations with negative emotionality.

Schizotopyal personality. Schizotypal personality is a heterogeneous construct, relating to deficits in social functioning, suspiciousness, social anxiety, eccentric/odd behavior, and cognitive-perceptual distortions (e.g., supernatural beliefs and unusual perceptions) (Raine, 2006; Ryan, McDonald, & Walker, 2013). It is conceptualized as a constellation of traits conferring increased liability of developing schizophrenia, with both genetic and environmental factors contributing to clinical risk (Debbane & Barrantes-Vidal, 2014; Raine, 2006). Like empathy, schizotypy has been conceptually and operationally defined in many different ways and it is difficult to integrate findings across studies (Kwapil & Barrantes-Vidal, 2015).

Although schizotypy can be understood as a dimensional construct within the general population, at high levels it can manifest as clinically diagnosable schizotypal personality disorder, or as schizophrenia (Chemerinski, Triebwasser, Roussos, & Siever, 2013; Raine, 2006). The taxonicity proposed for schizotypy (e.g., Meehl, 1962) has, however, been challenged and remains unclear (Kwapil & Barrantes-Vidal, 2015). Schizotypal personality is particularly strongly correlated with borderline personality disorder; the differences between the two appear to be most highlighted in the cognitive/perceptual distortions associated with schizotypy (Raine, 2006).

The substantial interpersonal deficits associated with schizotypal personality, particularly the lack of close friendships and constricted affect, have been observed to be related to individual differences in empathy (e.g., Bedwell et al., 2014; Kallai et al., 2019). Conceptually, however,

schizotypal personality is characterized by social anxiety, social anhedonia, social avoidance, and suspiciousness about the intentions of others, rather than by the antagonism/callousness associated with psychopathy. Moreover, limited evidence presented by Aaron, Benson, and Park (2015) suggests that the negative association between schizotypal personality and empathic caring may be heavily mediated (statistically at least) by alexithymia, a deficit in the conscious ability to identify and describe one's own feelings. Perhaps related to heightened alexithymia, Cohen et al. (2015) noted that studies regarding cognitive empathy abilities in schizotypy have been decidedly mixed; they speculate, however, that social cognition in schizotypy may be supported by compensatory mechanisms, including increased processing in the right temporoparietal junction (TPJ).

Also intriguingly, a number of researchers have proposed that schizotypy is associated with abnormalities in self-other boundaries (as reviewed by Cohen et al., 2015). Theoretically, it may be that weakened awareness of the distinction between self and others may confer elevated risk of empathic distress for people with elevated schizotypal traits. In sum, schizotypy represents a substantial difference vantage point than psychopathy by which to examine the fine-grained relations between emotional empathy and personality psychopathology.

The 32-item Schizotypal Personality Questionnaire Brief-Revised (SPQ-BR; Cohen, Matthews, Najolia, & Brown, 2010) is modeled on diagnostic criteria for schizotypal personality disorder and consists of 7 scales, each of which can be understood as a facet of either the positive, negative, or disorganized symptoms. The Social Anxiety (SA; ordinal $\alpha = .94$) facet measures anxiety in the presence of unfamiliar people (e.g., "I feel very uncomfortable in social situations involving unfamiliar people"). The Constricted Affect/No Close Friends (CFCA; ordinal $\alpha = .88$) facet measures lack of interpersonal intimacy (e.g., "I find it hard to be

emotionally close to other people") alongside constricted emotional expression (e.g., "I rarely laugh and smile"). The Eccentric Behavior (EB; ordinal α= .92) facet measures "oddness" of behaviour (e.g., "Other people see me as slightly eccentric (odd)"). The Odd Speech (OS; ordinal α= .86) facet measures disorganized speech tendencies (e.g., "I tend to wander off the topic when having a conversation"). The Suspiciousness/Ideas of Reference facet (SUIR; ordinal α= .90) measures a tendency to believe that others are watching you (e.g., "When shopping, I get the feeling that other people are taking notice of me") or out to harm you (e.g., "I often feel that others have it in for me"). The Unusual Perceptions (UP; ordinal α= .79) facet measures unusual perceptual experiences (e.g., "I often hear a voice speaking my thoughts aloud"). The Magical Thinking (MT; ordinal α= .91) facet measures beliefs in telepathy, clairvoyance, and other supernatural phenomena (e.g., "I have had experiences with astrology, seeing the future, UFO's, ESP, or a sixth sense"). Of these scales, only the MT and SUIR have been observed to be significantly associated with family history of schizophrenia; the CACF and SA scales have been observed to trend nearly to significance at the $p = .05$ level (in a sample with only 49 participants with family history of schizophrenia, Callaway, Cohen, Matthews, Dinzeo, 2014).

The relationships between schizotypal personality and emotional empathy have not yet been extensively explored, especially in any fine-grained manner. The cognitive-perceptual symptoms of schizotypy have generally been observed to be unrelated to empathic caring (e.g., Henry, Bailey, & Rendell, 2008; Ripoli et al., 2013; Wang et al., 2013; but see Melchers, Montag, Markett, & Reuter, 2015). The negative symptoms and disorganized symptoms of schizotypy, however, have been observed to be either (a) weakly to moderately negatively (e.g., Henry, Bailey, & Rendell, 2008; Thakkar & Park, 2010; Wang et al., 2013) or (b) non-significantly related (e.g., Asai, Mao, Sugimori, & Tanno, 2011; Ripoli et al., 2013) to empathic

caring. More specifically, in regard to SPQ specific scales, IRI EC and IRI PT have been observed to be negatively associated with CFCA, whereas IRI PD (which is largely a measure of trait neuroticism) has been found to be positively associated with SUIR, CFCA, SA, EB, and OS (Kallai et al., 2019; consistent with other findings that the negative symptoms of schizotypy are positively associated with IRI PD, Melchers et al., 2015).

Based upon these prior findings, I predicted that GFEC and TEQ would be negatively associated with CFCA, which is to be expected given the conceptual link between lacking empathic caring for others and having limited emotional intimacy with others. I predicted that all three of the empathy-related negative emotionality scales (EID, ANXCON, and SADCON) would demonstrate similar associations as previously observed with the IRI PD scale, being positively related to all SPQ-BR scales except UP and MT. I predicted that the EAV scale would demonstrate similar associations to those between EID and SPQ-BR, but that the strengths of the relationships would be substantially weaker, given that EID appears to be more strongly related to negative emotionality and dysfunction.

Given that AC is the most strongly associated with extraversion, I predicted that it would add meaningful incremental validity in negatively correlating with SA, CFCA, and SUIR. Given that CL is strongly conceptually related to emotional intimacy, I predicted that it would add meaningful incremental validity in negatively correlating with CFCA. Given that PT and ER are most strongly associated with agreeableness, I predicted that both would add meaningful incremental validity in negatively associating with SUIR; being suspicious/paranoid about the intentions of others is likely to inhibit empathic restraint and perspective-taking in conflict situations.

Refining specific content scales. I used scale reliability analyses to determine whether any additional items improved the individual scales, or whether any existing items should be dropped. In addition to examining the internal consistency of the scales, measured by ordinal α reliability, I examined their homogeneity (i.e., unidimensionality), as assessed by Horn's Parallel Analysis and Velicer's MAP test. Finally, I examined the measurement precision at varying latent trait levels of each scale using unidimensional GRM models.

For the PT scale, the 7 items had an ordinal α of .93. All 7 items were correlated with the scale above .7. No items were dropped. Horn's Parallel Analysis and Velicer's MAP both indicated 1 factor.

For the EAV scale, the 8 items had an ordinal α of .95. All 8 items were correlated with the scale above .7. No items were dropped. Horn's Parallel Analysis and Velicer's MAP both indicated 1 factor.

For the ER scale, the 7 items had an ordinal alpha reliability of .86. Two new items (ER6 and ER7) performed poorly, with scale correlations of .51 or lower. The scale's ordinal alpha and the MIC would improve by dropping them. The scale of the remaining 5 items had an ordinal α of .90; all items were correlated with the scale at .7 or higher. Horn's Parallel Analysis and Velicer's MAP both 1 factor.

For the AC scale, the 7 items had an ordinal α of .92. The weakest item (AC5) was correlated with the scale at .67, but I was concerned about its content. This item, which queries contagion for laughter, might be weaker given that some participants could reflexively associate laughter with not only appetitive emotional moments, but also aversive moments such as someone being laughed at unfairly or hurtfully, or people laughing at offensive jokes. I decided to drop this item, and the resulting ordinal α was .91. All 6 remaining items were correlated with

the scale at or above .7. Horn's Parallel Analysis indicated 2 factors (only 1 component, however), whereas Velicer's MAP indicated 1 factor. Given that the 2nd factor in the parallel analysis was barely significant and was well under an eigenvalue of 1, I concluded the scale was approximately unidimensional. This scale should consequently be understood as contagion for happiness, joy, or excitement, not for all possible appetitive emotions.

For the CL scale, the 8 items had an ordinal α of .94. All items were correlated with the scale at or above .7. Examining the weakest item (CL8), I noticed that its content was also very similar to an item on the General Empathic Caring Factor scale, and thus could interfere with its distinguishability as a specific factor. I removed that item and the remaining 7 items had an ordinal α of .94; all items were polychorically correlated with the scale (corrected for overlap) at .7 or above. Horn's Parallel Analysis and Velicer's MAP both indicated 1 factor.

For the EID scale, the 7 items had an ordinal α of .94. All items were correlated with the scale (corrected for overlap) at or above .67. Horn's Parallel Analysis and Velicer's MAP both indicated 1 factor.

For the SADCON scale, the 7 items had an ordinal α of .92. One item (SADCON2) was correlated with the scale at only .59, and the scale reliability would not decrease if it were dropped, so I removed it. Horn's Parallel Analysis and Velicer's MAP both indicated 1 factor.

For the ANXCON scale, the 7 items had an ordinal α of .94. All items were correlated with the scale at or above .67. Horn's Parallel Analysis and Velicer's MAP both indicated 1 factor.

Test Information Function (TIF) charts for all scales are presented in Appendix C. The CL, PT, and AC scales all had good to very good measurement precision from -2 SDs to 1 SD of the latent trait, but struggled at higher latent trait levels. The ER scale exhibited the poorest

measurement precision of any of the specific content scales, with adequate precision from roughly -3 SDs to slightly under 1 SD; its poor measurement precision above average latent trait levels indicates that the items are all too "easy" for respondents. The EAV and EID scales demonstrated good to very good precision from -1 SD to 3 SDs of the latent trait level, but was inadequate at low latent trait levels. ANXCON and SADCON both demonstrated good measurement precision from -2 SDs to 2 SDs of their latent trait levels.

Confirming the factor structure of the new scales. I first conducted a basic correlated factors CFA of the 5 empathic caring specific content scales, using the robust variant of the diagonally weighted least squares estimator (WLSMV). The model demonstrated only minimally acceptable fit (CFI = .90, TLI = .89, SRMR = .05, RMSEA = .04, 95% CI [.04, .05]; $\chi 2$ = 734.49, df = 485, p < .001). Inter-factor correlations were strong, ranging from |.53| to |.70|.

To explore whether the model was poorly specified, I analysed the number of likely factors. Velicer's MAP and Horn's parallel analysis both indicated 5 factors. I subsequently conducted a basic EFA, with polychoric correlations and oblimin rotation, extracting 5 factors. All items loaded above .5 with their expected factor groupings, and there were no cross-loadings above .5. There was only one cross-loading above .3, with a reverse-coded item from the AC scale (AC1) loading at .46 with the items from the EAV scale. In other words, it appears that the 5-factor structure holds adequately, but it is possible that item coding direction might slightly impact loading patterns.

I next examined these 5 modified subscales using bifactor CFA, with uncorrelated specific factors. For the hypothesized 5 specific factor structure, the model demonstrated adequate fit (CFI = .92, TLI = .91, SRMR = .05, RMSEA = .04, 95% CI [.03, .04]; $\chi 2$ = 647.29, df = 462, p = < .001). However, as shown in Table 7, the CL scale demonstrated very weak

distinguishability from the general factor (ω HS = .21), while the other 4 scales demonstrated much stronger distinguishability (ω HS values from .35 to .42).

As a complementary analysis, I conducted a second bifactor CFA, including the 15 General Empathic Caring Factor scale items (specified as pure indicators) and allowing the specific factors to correlate. For this hypothesized structure, the model demonstrated adequate fit (CFI = .92, TLI = .91, SRMR = .04, RMSEA = .03, 95% CI [.03, .04]; $\chi 2$ = 1,259.85, df = 991, $p < .001$). As shown in Table 8, the confidante listening scale, however, did not adequately distinguish itself from the general factor in this model (ω HS = .23), whereas the other 4 scales had much stronger distinguishability (ω HS values from .31 to .42).

To further explore the distinguishability of the CL scale, I ran a JB EBFA with correlated specific factors of the 5 subscale items and the general scale items (see Table 10 for factor loadings). Although it is often unacceptable to perform an EFA (or EBFA) on the same data used for a CFA, in this case the intention was merely to better elucidate the same data, not to confirm a prior EFA. For the PT, AC, EA, and ER scales, all subscale items loaded together as expected, with all loadings above .3. The only item from these subscales with a salient cross-loading was the negatively-coded item AC1, which loaded at .27 with the EA items (but .55 on its expected factor). The 5^{th} specific factor, though, only contained one item barely above .3, an item from the GFEC scale. Except for one item (CL2), no CL items loaded above .2 on a specific factor. In other words, in study 2, the CL items appear to operate generally as quasi-pure indicators of the general factor (as assessed with the abbreviated item pool in study 2, as opposed to the comprehensive item pool in study 1).

The weight of the factor analytic evidence indicates that the CL scale is not adequately distinguishable from the general factor of empathic caring. Instead, as perhaps tentatively

indicated by the results in Study 1, confidante listening appears to be a core aspect of broader empathic caring rather than a distinguishable non-core facet. Nonetheless, I analysed its correlations with external variables to evaluate its substantive differences from general empathic caring.

To examine the factor structure of the three empathy-related negative emotionality scales (EID, ANXCON, and SADCON), I first conducted a basic CF CFA of the 3 scales, using the robust variant of the diagonally weighted least squares estimator (WLSMV). The model demonstrated mediocre but acceptable fit (CFI = .91, TLI = .90, SRMR = .06, RMSEA = .06, 95% CI [.05, .07]; $\chi 2 = 357.25$, df = 167, $p < .001$). All items loaded above .6 on their scales, except for SADCON6, which loaded at .56. SADCON and EID were not significantly correlated, but ANXCON was correlated with EID ($r = .29$). ANXCON and SADCON were very strongly correlated ($r = .80$).

Horn's parallel analysis and Velicer's MAP both indicated 3 factors. I subsequently conducted a basic EFA, with polychoric correlations and oblimin rotation, extracting 3 factors. All items loaded above .5 with their expected factor groupings, except for SADCON5 (which loaded at .49), and there were no cross-loadings above .3. I concluded that the three scales adequately conformed to their expected factor structure.

Nomological Networks Results

Inter-relationships between empathy scales. All inter-correlations are presented in Table 9. The GFEC and TEQ were very strongly correlated with one another ($r = .88$) and with all five of the specific content scales (all *r*s above |.65|). The GFEC and TEQ were also strongly negatively correlated with the EID scale (*r*s were -.58 to -.63, respectively) and strongly positively correlated with the preliminary SADCON scale (*r*s were .39 and .41, respectively).

They were only weakly positively correlated, however, with the preliminary ANXCON scale (both *r*s = .14). These results corroborate the findings in Study 1, indicating that although contagion for the sadness of others appears to be strongly related to the broader empathic caring construct, contagion for anxiety may be much more of a separate construct. There were only slight differences between the two scales, with the GFEC more positively correlated with PT (Steiger's $z = 4.4$, $p < .001$, $df = 2$) and CL (Steiger's $z = 5.5$, $p < .001$, $df = 2$) and the TEQ more strongly correlated with EAV, ER, and EID (Steiger's *z*s all > |2.3|, *p*s all < .01).

In terms of zero-order correlations, all five of the empathic caring specific content scales were strongly correlated (absolute value of *r*s from .48 to .63). When comparing only their specific factor contents, though, as operationalized through their residual variances unique from the GFEC, these five scales were not significantly correlated.

ANXCON and SADCON were strongly positively correlated ($r = .74$), but only ANXCON was significantly correlated with EID at the zero-order level ($r = .26$). As observed in Study 1 as well, though, their residualized variable associations with EID were stronger than seen in the zero-order correlations (*r*s = .42 and .38, respectively). As similarly observed in Study 1, the GFEC also suppressed their relationships with the CL scale. At the zero-order level, anxiety contagion was not related to confidante listening and sadness contagion was positively associated with it; at the residualized level, though, both were similarly negatively correlated with the CL scale (*r*s from -.16 to - .17).

Study 2 indicated even stronger correlations between EAV and EID than observed in Study 1, at both the zero-order level ($r = .86$) and the residualized variable level ($r = .78$). More troublingly, unlike in Study 1, both EAV residual and EID residual were meaningfully correlated with the two aversive contagion scales (for EAV and ANXCON, $r = .25$; for EAV and

SADCON, r = .20). In other words, unlike in Study 1, where only EID was related to aversive contagion, EAV was also related to aversive contagion in Study 2.

To further examine the distinguishability of the EAV and EID scales, I combined them and examined them for undimensionality. Horn's parallel analysis indicated 2 factors, but only 1 was above an eigenvalue of 1. Velicer's MAP test indicated only 1 factor. In a EFA extracting 2 factors, all EAV and EID items loaded on the 1st factor, with the EAV items loading between .82 and .93 and the EID items loading between .68 and .91. Only 3 items had loadings at or above .3 on the 2nd factor (largest was .36); all the EAV items had negligible negative loadings and all the EID items had positive loadings between .01 and .36. In other words, from these statistical perspectives, the EAV and EID items are virtually indistinguishable factor-analytically, despite their facial conceptual differences.

Relationships with antagonism-related variables. All relationships with antagonism-related variables are presented in Tables 11 and 12. Both GFEC and TEQ were exceptionally strongly related to TriPM Meanness (*r*s were -.71 and -.81, respectively). They both also demonstrated meaningful correlations with TriPM Disinhibition (*r*s were -.29 and -.39, respectively), as well as with all four of the BPAQ scales (*r*s from -.24 to -.38). Both GFEC and TEQ were strongly to moderately negatively associated MPS Distrust of Others (DO) and MPS Amorality (AMO) (*r*s from -.36 to -.42) and weakly negatively associated with MPS Desire for Status (DS) (*r*s were -.19 and -.24, respectively). Neither scale was associated with MPS Desire for Control (DC).

As in Study 1, I focused on the nomological networks of the specific content empathy scales at the level of their residualized variables after regressing on GFEC, not on their zero-order correlations. At the residual level, TriPM Meanness was significantly associated with the

residuals of EAV (r = .26), ER (r = .39) and EID (r = .15). It is worth noting that EAV residual exhibited a significantly stronger association with trait meanness than did EID residual (Steiger's z = 3.8, p < .001, df = 2).

TriPM Disinhibition was significantly associated with the residuals of EAV (r = .32), ER (r = .27), EID (r = .33), ANXCON (r = .35), and SADCON (r = .29). SADCON's association with Disinhibition were suppressed by the GFEC (β = -.16, 99% CI [-.24, -.10]). In other words, heightened susceptibility to aversive contagion is positively correlated with the disinhibition traits associated with psychopathy (consistent with Disinhibition's association with negative emotionality, Sleep et al., 2019), as are susceptibility to discomfort in empathic encounters and tendencies to avoid empathic encounters.

All of the residuals of the specific content scales, except for AC, were associated with BPAQ Anger (rs from |.14| to |.34|). Importantly, ANXCON residual and SADCON residual exhibited the strongest correlations with trait anger of any of the scales (rs were .34 and .30). BPAQ Hostility, which represents bitterness and suspicion of others, was associated with the residuals of EAV, AC, EID, ANXCON, and SADCON (rs from |.28| to |.52|). The residuals of the three scales most strongly associated with negative emotionality (EID, ANXCON, and SADCON) were the most strongly associated (positively) with trait hostility. The GFEC suppressed SADCON's associations with both Anger (β = -.15, 99% CI [-.26, -.06]) and Hostility (β = -.17, 99% CI [-.27, -.08]). Of note, EID residual was substantially more strongly related to Hostility than EAV residual (Steiger's z = 4.4, p < .001, df = 2).

Only ER's residual was significantly associated with BPAQ verbal aggression (r = -.34). It was also perhaps the only residualized scale to meaningfully be associated with BPAQ

physical aggression (r = -.24), though EAV residual and AC residual had very small significant associations with BPAQ physical aggression.

AC residual and ER residual were negatively correlated with MPS DO (rs were -.23 and -.19, respectively), whereas the residuals of EAV, EID, ANXCON, and SADCON were all positively associated with MPS DO (rs from .26 to .37). SADCON's positive association with it was suppressed by the GFEC ((β = -.18, 99% CI [-.31, -.08]). EAV residual and EID residual did not significantly differ in their associations with distrust of others.

ER residual was negatively correlated with MPS AMO (r = -.29), whereas EAV residual and EID residual were positively correlated with it (rs were .31 and .23, respectively); EAV residual was significantly more strongly related to amorality than EID residual (Steiger's z = 2.8, p < .001, df =2).

Only CL residual was associated with MPS DS; surprisingly, the relationship was positive (r = .17), indicating that there may be some small association between confidante listening and a desire for elevated social status. This relationship, however, was suppressed by the GFEC (β = -.33, 99% CI [-.52, -.15]). Only ER residual was associated with MPS DC (r = -.23), indicating that empathic restraint is particularly related to reduced desire for controlling the behaviors of others.

Relationships with emotional distress. Relationships with emotional distress variables are presented in Table 12. The GFEC and TEQ were both positively associated with Self-Compassion (rs were .27 and .26, respectively) and Satisfaction with Life (rs were .32 and .25, respectively). Neither scale, however, was significantly associated with General Worry (PSWQ).

At the level of residualized variables, PT, CL, EAV, and AC were significantly associated with Self-Compassion (absolute value of rs from .18 to .31), whereas ER residual was

not significantly correlated with it. The residuals of EID, ANXCON, and SADCON were moderately to strongly negatively associated with Self-Compassion (*r*s from -.36 to -.51). SADCON's negative association with Self-Compassion was suppressed by the GFEC (β = .18, 99% CI [.09, .28]).

The residuals of PT, CL, and AC were significantly negatively associated with General Worry (*r*s from -.16 to -.17), whereas the residuals of EAV and ER were not significantly associated with it. The residuals of EID, ANXCON, and SADCON were moderately to strongly positively associated with General Worry (*r*s from .30 to .64).

Of the residuals of the empathic caring specific content scales, only AC was significantly (but only weakly, $r = .14$) correlated with Satisfaction for Life. The residuals of EID, ANXCON, and SADCON, however, were all weakly to moderately negatively associated with it (*r*s from -.20 to -.30). SADCON's association with SLS was suppressed by the GFEC (β = .17, 99% CI [.08, .28]).

Relationships with schizotypal personality. Relationships with schizotypal personality features are presented in Table 13. The GFEC and TEQ scales were both moderately to strongly negatively associated with SPQ-BR CFCA (*r*s were -.47 and -.51), and weakly negatively associated with both SPQ-BR SA (*r*s were -.18 and -.15) and SPQ-BR EB (*r*s were -.17 and -.18). Both the GFEC ($r = -.18$) and the TEQ ($r = -.28$) were negatively correlated with SUIR, but neither was significantly associated with MT, UP, or OS. In other words, general empathic caring was associated with substantially reduced constriction of affect and social isolation, as well as with somewhat reduced social anxiety, suspiciousness/ideas of reference, and eccentric behavior. It does not, however, appear to be related to magical thinking, odd speech, or unusual perceptions.

At the residual variable level, PT was not significantly associated with any of the SPQ-BR scales. CL residual was negatively correlated with SA, CFCA, and OS (*r*s from -.17 to -.23). AC residual was negatively correlated with SA, CFCA, and SUIR (*r*s from -.20 to -.28).
ER residual was positively correlated with SA (r = .18), but negatively correlated with both UP (r = -.18) and MT (r = -.23). Its association with SA was suppressed by the GFEC (β =- .23, 99% CI [-.38, -.11]), as was its relationship to MT (β = .19, 99% CI [.07, .33]).

EAV residual was positively associated with all seven SPQ-BR scales, with its strongest association with SUIR (r = .37) and its weakest association with MT (r = .19). Its association with MT was significantly suppressed by the GFEC (β = -.16, 99% CI [-.31, -.02]) as was its association with UP (β = .16, 99% CI [-.30, -.03]).
EID residual was positively associated with all SPQ-BR scales except for MT, with its strongest association with CFCA (r = .52) and its weakest significant association with UP (r = .27).

ANXCON residual was moderately to strongly positively associated with all SPQ-BR scales except for MT, with its strongest association with SA (r = .58) and its weakest significant association with EB (r = .25). SADCON residual was weakly to moderately associated with all SPQ-BR scales except for MT, with its strongest association with SA (r = .48) and its weakest significant association with EB (r = .17). SADCON's associations with SA, CFCA, SUIR, EB, OS, and SUIR were all significantly suppressed by the GFEC (all significant βs stronger than -.09).

Importantly, EAV and EID exhibited substantial differences in their relations to emotional distress and the dimensions of schizotypal personality. Though EAV and EID are highly similar in many ways, at the residual variable level, EID is much more strongly associated with general worry (Steiger's z = 6.9, p < .001, df = 2), social anxiety (Steiger's z = 8.0, p <

.001, *df* = 2), general dissatisfaction with life (Steiger's z = 6.4, $p < .001$, *df* = 2), and constricted affect/lack of close friendships (Steiger's z = 6.4, $p < .001$, *df* = 2), etc.

Discussion of these Study 2 results is collapsed into the General Discussion section.

General Discussion

The "hard core" of the empathy construct, which I refer to here as "einfühlung empathy," is imaginatively experiencing the subjective consciousness of another person, sensing, understanding, and structuring the world *as if* one were that person. It involves experiencing the qualia of another person to at least some limited degree, entering into "what is like to be" that person. As many or most scholars have argued (e.g., Stein, 1917; Rogers, 1975), this also requires maintaining some degree of awareness of the "otherness" of the other's consciousness and modulating the degree to which one unknowingly attributes one's own feelings/perspectives to the other's consciousness. As with the general sweep of prior studies of empathy, the studies in this dissertation did not *directly* measure this heart of the construct. Rather, the two studies only aimed to measure trait-level *theorized* close correlates of the experiential core of the empathy construct.

Studying einfühlung empathy, perhaps more so than many other psychological constructs, is glancingly analogous to studying a black hole, albeit immensely less precisely and with far more threats to validity. A black hole has dramatic effects on objects close to it, while its effects ripple far beyond in a diminishing fashion. Yet, it requires herculean efforts to capture even a superficial picture of a black hole, while whatever specifically is inside a black hole is unseeable. To study a black hole, one studies the perceivable matter/energy around it and, in turn, makes indirect inferences about the black hole itself.

Similarly, einfühlung empathy is thought to have important effects on the person experiencing it and those with whom he/she interacts in this manner, with rippling implications extending outward into broader social dynamics (e.g., political tolerance). Yet, for various reasons discussed in earlier sections of this dissertation, empirical measurement of this process, which manifests within the subjective consciousness of a person and is difficult to distinguish from other processes, such as projection and emotional contagion, thus far has presented a stubbornly intractable problem. As a result, studies of empathy have focused on measuring the more observable or self-reportable theorized close correlates of this core construct. Unlike black hole physicists, though, many empathy researchers appear to have subtly, and perhaps gradually over time, lost sight of the core of the construct, and have (at least in presenting findings) seemed to define various theorized correlates as "empathy" itself, while failing to acknowledge the murky core of the construct. Some researchers, proposing restrictive "isomorphic emotion matching" definitions, have gone so far as to nearly exclude einfühlung empathy itself from being called "empathy" (cf. Zahavi & Rochat, 2015).

Open acknowledgement of our inability, at least to this date, to specifically measure core einfühlung empathy is necessary. To that end, the studies in this dissertation should be understood only as attempting to better describe and measure theorized close correlates of einfühlung empathy and to better understand their relations to one another and to other conceptually related variables. More specifically, the two studies indicate that self-reported "empathy" constructs can be perhaps be broadly understood as falling into two general domains: a broad domain of "empathic caring" and a domain of "aversive emotion matching," with a construct of "empathic intimacy discomfort" substantially overlapping with both domains.

Heterogeneity within these domains is parsed and then the inter-relations between sub-domains are explored, as are relations with more external variables.

The arguments for considering general caring for others as a close correlate of einfühlung empathy can be briefly summarized. First, imaginatively stepping into another's person consciousness may often, but certainly not always, tend to "humanize" the other person, increasing one's *valuing* of the feelings of that person (cf. Smith, 1759; Schutte & Stilinovic, 2017; Wang et al., 2019). In other words, einfühlung empathy makes the feelings of others more experientially "real" and intelligible to the observer, rather than remaining distantly abstract and alien, and this *allows* them to be more readily emotionally valued by the perceiver. Put differently, in more Rogerian terms, empathy is intimately related to building and maintaining authentic, validating, comforting relationships between people; feeling into the consciousness of another person may involve, even in passing instances, a flicker of "relatedness" to that person.

Second, in the opposite causal direction, valuing the feelings of others may be *the* primary motivator for einfühlung empathy (cf. Zaki, 2014). A self-centered, solipsistic person may be less motivated to contemplate the feelings of others, much less feel himself/herself into them, except perhaps in situations when doing so has a strategic self-seeking purpose. A more generous, caring person, however, may tend to be more motivated to approach empathic encounters, expend effort in imagining the perspectives of others, and de-prioritize his or her own feelings and judgments in considering situations (cf. Rogers, 1975).

The arguments for considering isomorphic emotion-matching a close correlate of einfühlung empathy appear weaker in comparison. As discussed extensively in the introductory sections of this dissertation, although isomorphic emotion matching and einfühlung empathy may have some unique connection in terms of spontaneous mimicry and/or contagion, this

connection may be relatively weak in comparison to the role of evolved functionality (e.g., Keltner & Haidt, 1999) and/or situation appraisal (e.g., Smith, 1759; Wondra & Wellsworth, 2015). Moreover, as indicated by a number of past theorists (e.g., Cuff et al., 2016; Murphy et al., 2018b; Smith, 1759), the nature of such connections may differ radically depending upon the particular emotion being spontaneously matched. Nonetheless, regardless of whether this isomorphic matching domain is a critical trait-like aspect of the empathy construct, it is imperative that it be measured as well as possible.

The findings and implications of these studies are presented in the following section in terms of (a) overarching implications for the conceptualization of "empathy," (b) incremental advances of the self-report measurement of general and specific "empathy" constructs, and (c) tentative fine-grained associations between these constructs and other external correlates.

Overarching Implications for Empathy Conceptualization

Empathic contagion is not a unidimensional construct. The two studies in this dissertation corroborated and extended the findings of Murphy et al. (2018b) indicating that appetitive contagion and aversive contagion should not be collapsed into the same measurement model. In Study 1, in the initial ICLUST analysis of the total item pool, appetitive contagion items loaded strongly on the 140-item cluster interpreted as "general empathic caring" whereas the aversive contagion items loaded on their own separate cluster, which was minimally correlated with the "general empathic caring" cluster. The sharp divergence between appetitive contagion and aversive contagion was manifested on many levels throughout both studies.

For example, whereas appetitive contagion was strongly associated with emotional well-being, aversive contagion was robustly associated with emotional distress. Similarly, whereas appetitive contagion was primarily associated with extraversion and agreeableness in regard to

broadband personality dimensions, aversive contagion was overwhelmingly primarily associated with trait negative emotionality. Tending to "catch" the positive emotions of others is associated with a happier, more harmonious life, whereas tending to "catch" the negative emotions of others is associated with an anxious, unsatisfying life, with friction in relation to other people.

Moreover, although Jordan, Amir, and Bloom (2016) argued that emotion-matching is not substantially related to compassion, these studies indicate that appetitive contagion is strongly correlated with general empathic caring, whereas aversive forms of contagion are only negligibly to modestly associated with empathic caring. This strongly corroborates similar findings by Murphy et al. (2018b).

Interestingly, both in terms of factor analytical results and nomological network results, it appears that aversive contagion itself should also not be understood as a unidimensional construct. Although strongly correlated (e.g., $r = .74$ in Study 2), anxiety contagion and sadness contagion manifested strikingly diverging correlations at the zero-order level. For instance, across both studies, sadness contagion was robustly positively associated with both the general factor of empathic caring and with many of the specific content empathic caring domains; anxiety contagion, however, was generally minimally related to empathic caring. Moreover, at the zero-order level, their nomological networks differed substantially in regard to many external variables. After accounting for shared variance with general empathic caring, however, their nomological networks were mostly indistinguishable. In other words, it appears that the main difference between anxiety contagion and sadness contagion is that the latter bears some association with caring for others (while being associated with reduced self-compassion).

Moving forward, empathy researchers should take care not to conflate matching the negative emotions of others with matching the positive emotions of others. Moreover, it seems

that even a dichotomy of aversive and appetitive matching may be overly simplistic. These studies only revealed the distinction between sadness matching and anxiety matching, but that is probably due to my failure to include a sufficient number of items for each specific emotional domain (e.g., surprise, disgust).

These results may tend to further undermine the emerging "standard definition" (Coll et al., 2017) of empathy as simply "Identify the Emotion → Feel Roughly the Same Emotion." If isomorphic matching has radically different correlates depending upon the particular emotion that is being matched, this indicates that mimicry, contagion, or whatever unique process might be thought to generate matching in general, is at best only a weak unifying mechanism relative to functionality of spontaneous emotion transmission (e.g., Keltner & Haidt, 1999) and/or cognitive appraisal (Smith, 1759; Wondra & Wellsworth, 2015). If so, then the case for sharply demarcating matching from non-matching vicarious emotional responses is weak.

Empathic avoidance is critical. As suggested by Weisz and Zaki (2018), empathy researchers should focus on measuring approach-avoidance tendencies (and the motivations behind them). In this dissertation, this domain was assessed with a range of mostly novel items, such as "When someone starts to talk to me about his or her emotional struggles, I usually change the subject to something less serious" and "I try to avoid having to listen to people talk about their deep emotions." Although empathic avoidance, in both studies, was extremely strongly (negatively) correlated with general empathic caring, its incremental value in predicting other external variables generally far surpassed other specific content domains of empathic caring. In comparison to general empathic caring, empathic avoidance was particularly valuable in incrementally predicting hostility, amorality, distrust of others, disinhibition, suspiciousness, and lack of friends and/or constricted affect.

Asking people if they care about others, feel bad if they hurt others' feelings, try to put themselves in other people's shoes, and catch the happiness of others are all helpful lines of questioning. It appears, though, that asking people if they tend to avoid empathy (through situation selection, attentional deployment, etc.) might be an even more useful domain of questioning. It is also worth noting that, at least facially, most of these items, while having some degree of evaluative content, do not have the same high degree of obviously strong evaluative content as do the other scale item contents. Though I did not quantitatively assess this issue in my studies, my hypothesis is that these items are slightly less vulnerable to social desirability and related biases than are other empathic caring questionnaire domains.

Interestingly, in these two studies, it appears that the underlying motivations to avoid empathy may stem overwhelmingly from empathic distress/discomfort. In crafting the items for different proposed domains, I assumed that empathic approach-avoidance and empathic distress/discomfort would be quite distinguishable. Much as theorized in Weisz and Zaki (2018), I assumed that a wide variety of motivations generate empathic avoidance, such as a desire to avoid expending time and effort on others, a desire to associate with people who appear more successful rather than struggling, a general lack of interest in the emotions of others, and so on. In both Study 1 and Study 2, however, empathic avoidance and empathic intimacy discomfort were so strongly associated that they were nearly indistinguishable from some perspectives (e.g., combined in Study 2, they constituted essentially a single dimension). In other words, contrary to my expectations, it appears that empathic avoidance may be *overwhelmingly* driven by empathic intimacy discomfort (e.g., "When people share their deep emotions, I feel apprehensive and ill-at-ease" and "I tend to get anxious when people start talking about their emotional trauma").

It is possible, however, that my studies may exaggerate the near-isomorphism between this specific motivation (distress/discomfort) and this specific behavioral tendency (empathic avoidance). First, it is possible that the items I wrote for these domains are not as cleanly specified as they should have been. Although none of the empathic avoidance items mention negative emotionality on the part of the respondent, and all of the empathic intimacy discomfort items do just that, the items nonetheless relate to much of the same domain: listening to others talk about emotions. This shared aspect may tend to pull the two constructs closer together. Second, it is possible that satisficing tendencies by respondents, which are a particular concern in M-Turk samples (Hamby & Taylor, 2016), may have exaggerated the convergence of these two domains. Nonetheless, it is clear that (a) empathic avoidance is a critical domain to be measured in empathy research, which has been overlooked by many researchers (but, again, see Shaw, Batson, & Todd, 1994) and (b) feelings of discomfort or distress may be the primary motivating cause of such avoidance.

Confidante listening may be an under-appreciated central aspect of empathic caring. Although Rogerian empathy is closely associated with attentive, validating listening, self-report measures of empathy have typically neglected to include any items related to being a supportive confidante listener (with very few exceptions, e.g., “Friends usually talk to me about their problems as they say that I am very understanding” in the EQ). As a result, of all the potential domains I hypothesized, “confidante listening” was a domain that necessitated particularly heavy attention in writing new items.

In Study 1, this domain emerged as a large specific content dimension, but appeared to be only modestly distinguishable from the general factor of empathic caring, specifically in terms of bifactor indices. In Study 2, this distinguishability, in terms of bifactor modeling, collapsed

entirely; all confidante listening items manifested as “pure indicators” of the general factor of empathic caring. Furthermore, after accounting for shared variance with the general factor of empathic caring, confidante listening failed to substantially predict any external variables, further indicating its lack of distinguishability. This indistinguishability may not replicate in future studies; if my concerns about satisficing in M-Turk samples are justified, it is possible that confidante listening may be more distinguishable in samples where optimizing responses are more likely. Nonetheless, it appears that confidante listening is, indeed, a central aspect of being a broadly caring person toward others.

In other words, although empathy research appears to have largely moved away from thinking of empathy in Rogerian (or what we might call proto-Rogerian) terms, at least outside of psychotherapy research, empathic listening tendencies might nonetheless be a core aspect of the construct in daily life (as suggested by Rogers, 1975). If empathy is understood as evolving out of mammalian caregiving, reflected in attachment (e.g., Stern & Cassidy, 2018), and involved in relationship bonding, then this link would appear to be even stronger. If so, empathy research might be particularly furthered by investigating attentive, validating listening skills in non-therapists, especially in behavioral paradigms. In some ways, this would be a “back to the future” approach, returning to some of the kinds of lines of research conducted by Rogerians decades ago in psychotherapy research (e.g., the Accurate Empathy Scale, for evaluating psychotherapists, Truax, 1967; critiques of the Accurate Empathy Scale, e.g., Bergin & Jasper, 1969).

Measurement Advances of New Self-Report Empathy Scales

These dissertation studies aimed to improve upon self-report empathy measurement methods by (a) administering a much larger, more comprehensive item pool, specifically crafted

to adequately cover potentially overlooked domains; (b) employing bifactor modeling as a complement to Thurstonian correlated factors analysis; and (c) attempting to improve the measurement precision of empathy scales at particularly high levels of latent traits. The first two methodological approaches generated substantial value; attempts to improve the measurement precision of the scales at higher latent trait levels, though, were only minimally successful. It should be emphasized that these scales, although an improvement over existing empathy questionnaires, are far from perfect and should be revised or replaced in future studies.

In Table 14, I have summarized the *prominent* correlates of each scale. The scales themselves are presented in Appendix B. Some scales, such as the Empathic Avoidance (EAV) scale, demonstrate strong levels of incremental value above and beyond the general factor of empathic caring; others, however, such as the confidante listening (CL), perspective-taking (PT), and appetitive contagion (AC) scales, do not.

The General Factor of Empathic Caring (GFEC) scale. The General Factor of Empathic Caring scale, which was constructed using multiple bifactor exploratory factor analysis (EBFA) methods that were compared for convergence, consists of 15 "quasi-pure" indicators of the general empathic caring construct. It was distilled from a large pool of 137 items, all of which broadly clustered together as "empathic caring." The GFEC scale measures the "core" of the general factor of the broader empathic caring construct, while intentionally excluding items that strongly measure specific content domains. As a result, though facially broad and heterogeneous in its item contents, both Study 1 and Study 2 revealed that it is slightly more unidimensional than a similar general empathy scale (the TEQ; Spreng et al., 2009) which was created by selecting top loadings from a unidimensional EFA of a heterogeneous body of empathy questionnaire items.

Nonetheless, the GFEC scale is strong correlated with the TEQ (r = .88 in Study 2) and exhibits an extremely similar nomological network. On the one hand, this confirms that the GFEC scale (and the general factor of the item pool from which it was formed) strongly converges with the common empathy factor from a substantially different empathy item pool, which corroborates its convergent validity. On the other hand, though, it suggests that its improvement over past methods is only modestly incremental.

Because I intentionally wrote new items intended to be of higher "difficulty" (informed by findings in Murphy et al., 2018a), and the GFEC was refined slightly through IRT analyses, it has slightly better measurement precision than the TEQ at higher levels of trait empathy. Despite these efforts, it still has poor measurement precision at the highest levels of the latent trait; both the GFEC and the TEQ had a substantial number of respondents indicating the highest possible score. It is possible that, perhaps due to socially desirable responding or related biases, it may be inherently difficult to create a self-report measure of empathic caring that is capable of strong measurement precision above 2 SDs of the latent trait.

Because of its heightened unidimensionality and comparative lack of specific factor content, the GFEC, as a measure specifically of the "core" of the empathic caring construct, is conceptually better equipped for use in incremental validity analyses of narrower or alternative empathy measures. Furthermore, it can provide a strong "anchor" for future attempts to devise or refine self-report empathy questionnaires, as well as perhaps closely-related constructs such as compassion (from which empathic caring may be pragmatically indistinguishable).

The GFEC is significantly correlated with all domains of general personality as assessed by the HEXACO-60, ranging from a weak association with trait negative emotionality to a moderate/strong association with trait agreeableness. These association patterns are highly

similar to those demonstrated by the TEQ, as well as the general factor of the entire empathic caring item pool.

I recommend that any empathy research study utilizing self-report questionnaires consider administering the GFEC scale. It can serve as a robust proxy for the general factor of empathic caring one would find when examining a comprehensively large item pool. Because of its particularly unidimensional, non-specific design, it can also serve as a valuable standard against which specific content empathy scales or alternative empathy measures can be investigated for incremental value.

Specific Content Scales. The two studies produced a final collection of eight specific content scales. Six scales (Perspective-Taking, Confidante Listening, Appetitive Contagion, Empathic Avoidance, Empathic Restraint, and Empathic Absorption) were formed as specific content domains from the large initial 137-item cluster, identified via ICLUST analysis, which I characterized as "general empathic caring." These scales were constructed through EFAs, supplemented by EBFAs. CFAs were employed to subsequently confirm factor structures. The Empathic Absorption scale was dropped after Study 1, for reasons summarized further below. The Confidante Listening scale is highly suspect as a distinguishable specific content domain, also as summarized below.

Given that the initial ICLUST analysis indicated that items referencing negative emotionality on the part of the respondent tended to cluster separately from the "general empathic caring" items, and with the goal of setting up bifactor analyses in a conceptually justified manner, I analysed all items related to experiencing negative emotionality in interpersonal encounters separately from the empathic caring items. I only used EFAs and CFAS, without supplementing with bifactor models, in constructing the three resulting specific

content domain scales (Empathic Intimacy Discomfort, Sadness Contagion, and Anxiety Contagion).

None of these specific content scales possess excellent measurement precision at both very high and very low trait levels (see Appendix C for TIF curves). All of them demonstrate good overall internal consistency (all ordinal αs above .89); all of them also appear to be essentially unidimensional, as assessed by Horn's parallel analysis and Velicer's MAP test.

Perspective-Taking (PT) scale. The PT scale measures self-reported tendency to perceive situations from the perspective of others, especially those with whom one has conflicts. It is narrower in content than the IRI Perspective-Taking scale, which encompasses a broader range of agreeableness and reasonableness. The PT scale is psychometrically distinguishable from the general factor of empathic caring, demonstrating modest ωHS values in Study 2 (.37 in orthogonal bifactor model; .32 in oblique bifactor model), but only seems to provide substantial incremental value in predicting trait agreeableness. It appears to offer little if any incremental value above and beyond the general factor of empathic caring in relating to antagonism-related variables, emotional distress, and schizotypy. Researchers may wish to use the PT scale if they are particularly interested in the relationships between facets of empathy and trait agreeableness, or if they are specifically interested in the PT scale's content domain. In most cases, though, the PT scale may have little if any added value when included in study designs alongside the GFEC.

Confidante Listening (CL) scale. The CL scale measures the self-reported tendency to be a supportive, comforting, non-judgmental listener. It does not appear to have been measured as a specific domain in prior self-report empathy questionnaires. As already discussed, it seems to be pragmatically indistinguishable from the general factor of empathic caring. It offered no substantial incremental value in associating with external variables. In self-report empathy

research, it seems unlikely to offer meaningful added value in study designs alongside the GFEC. It would, however, be valuable to know whether or not it predicts confidante listening as assessed behaviorally and, if so, whether it does so above and beyond that provided by the general factor of empathic caring. Although the CL scale adds little predictive value above and beyond the GFEC scale, it may be particularly of interest to researchers exploring the specific domain of emotionally supportive listening.

Appetitive Contagion (AC) Scale. The AC scale measures self-reported emotional contagion for the joy and pleasure of others but does not significantly encompass other appetitive emotions (e.g., laughing, pride, contentment). It is psychometrically distinguishable from the general factor of empathic caring, demonstrating modest ωHS values in Study 2 (.42 in orthogonal bifactor model; .39 in oblique bifactor model). It offers modest incremental value in relating to constructs like extraversion, self-compassion, and reduced hostility. The AC scale may be at least modestly valuable in most self-report investigations of empathy, but it is essential in any study that aims to focus on isomorphic emotion matching.

Empathic Restraint (ER) scale. The ER scale measures self-reported desire to avoid causing emotional pain to others. Although similar items are contained with the ACME AR scale, no extant empathy questionnaire has specifically measured this domain as its own separate scale. It is psychometrically distinguishable from the general factor of empathic caring, demonstrating modest ωHS values in Study 2 (.35 in orthogonal bifactor model; .31 in oblique bifactor model). Attempts to increase the number of reliable indicator items in Study 2 were largely unsuccessful, and the final scale only has 5 items. As most clearly presented in its TIF curve in Appendix C, the measurement precision of this scale is particularly lacking. Further efforts are needed to improve its psychometric properties. Nonetheless, it demonstrates

meaningful incremental value in associating with important external variables, especially trait meanness and aggressive behavior. I recommend that researchers who choose to employ this scale in future studies do so while adding and investigating potential items to improve its functioning.

Empathic Absorption (EAB) scale, *subsequently eliminated after Study 1*. The EAB scale, composed of revised items from the IRI Fantasy scale and new items generated to align with them, manifested as a distinguishable dimension in factor analyses in Study 1. Nevertheless, my concerns that it would simply function as an alternative form of a contagion (or projection) scale appeared to be founded. Its content was primarily correlated with aversive forms of emotional contagion, but also meaningfully with appetitive contagion. In terms of broadband personality dimensions, it related primarily to heightened negative emotionality, with other smaller relationships with personality that were in the opposite direction of those exhibited by general empathic caring. Particularly surprisingly, its unique variance was not associated with trait openness (which is strongly related to trait absorption).

In retrospect, I should have written items to distinctly reflect absorption for positive and negative emotions, rather than only administering items that had no particular valence. I concluded that the scale was merely a restatement of an unvalenced contagion scale, likely with more relation to projection than to einfühlung empathy. I dropped this scale after Study 1. I worry, however, that the decision to drop this scale may have been premature; it is possible that a stronger item generation process could have improved the scale's value.

Empathic Avoidance (EAV) and Empathic Intimacy Discomfort (EID) scales. The EAV scale facially measures self-reported tendencies to avoid empathic encounters with others. The EID scale facially measures self-reported tendencies to feel discomfort or distressed when in

empathic encounters with others. Unexpectedly, the EAV scale and EID scale, which are strongly correlated (e.g., r = .86 in Study 2), were difficult to distinguish. In a factor analysis of just the items from these two scales, in Study 2, they manifested as a unidimensional construct. As discussed earlier, although I assumed empathic intimacy discomfort would be a substantial motivation for empathic avoidance, I expected avoidance tendencies would stem from a range of other motivations as well. Nonetheless, the strong correlations between EAV and EID indicate that discomfort/distress may be the primary motivation for individuals to tend to avoid empathic encounters.

Despite their general appearance of indistinguishability, there is some evidence that the two scales differ, in ways that are consistent with their theoretical differences. First, EAV is more strongly associated with general empathic caring than is EID. This finding indicates that lack of caring for others is somewhat more related to avoidance than discomfort. Second, EID is substantially more strongly related to anxiety contagion, trait negative emotionality, social anxiety, general worry, reduced self-compassion, and lower life satisfaction than EAV. In other words, despite their factor analytic indistinguishability, EAV appears to be more related to lack of caring for others, whereas EID is more strongly associated with anxiety and other aspects of negative emotionality.

Both scales have strong incremental value, above and beyond the general factor of empathic caring, and it would be useful to include them both in future studies. Though they are closely related, they do appear to relate differently to external variables in ways that are consistent with their item contents and labels. Researchers should examine them closely before deciding whether and how they might like to use them in particular research projects.

Sadness Contagion (SADCON) and Anxiety Contagion (ANXCON) scales. Aversive contagion items split into two strongly correlated scales. The SADCON scale measures a respondent's tendency to isomorphically contagiously match the sadness or depression of other people. The ANXCON scale measures the same tendency, but for anxiety or fear. Both scales relate far more strongly to trait negatively emotionality (*r*s were .54 and .67, respectively), as assessed by the HEXACO-60, than to the other broadband personality dimensions. As discussed earlier, their residualized specific contents exhibited similar nomological networks, though ANXCON typically demonstrated stronger relationships with other variables.

SADCON's relationships with a wide range of variables (e.g., extraversion, conscientiousness) were suppressed by the general factor of empathic caring. In future research, effective fine-grained interpretations of sadness contagion's associations with external variables should examine the construct both at the zero-order level as well as in terms of its unique variance after accounting for the GFEC scale.

Fine-Grained Associations with External Variables

In discussing associations with the specific content scales, I focus here (as elsewhere throughout this manuscript) on associations with the residualized empathy variables (their unique variances after accounting for shared variance with the GFEC scale). All references to associations are for the unique variances of the specific content scales, never for their zero-order relationships.

My extensive a priori predictions for these associations, presented in the Methods section of Study 2, are not recapitulated here; I do, however, note particularly surprising findings. Rather than discuss all statistically significant associations, I only focus on those that appear to be

meaningful, setting an admittedly arbitrary threshold of $r > |.20|$. All associations, however, are presented in Tables 11-13.

Relationships with antagonism-related variables. Given the strong theoretical associations between empathy and antagonism (or reversed agreeableness), associations that have been frequently observed in prior studies (e.g., Murphy et al., 2018a; Vachon & Lynam, 2016), I sought to conduct a more fine-grained examination with the nine different empathy-related scales generated in these two studies.

Except for the "Desire for Control" scale of the Machiavellian Personality Scales, all antagonism related variables (e.g., psychopathic meanness, various aspects of aggression) were significantly negatively associated with both the GFEC and TEQ scales, indicating a robust (but often only modest) relationship between general empathic caring and antagonism. In particular, TriPM Meanness was strongly negatively associated with empathic caring (*r*s from -.71 to -.81), which is consistent with the long-standing view that lack of caring for others is a core trait of psychopathy.

Psychopathy (TriPM) associations. TriPM Meanness was most strongly associated (negatively) with empathic restraint, but also meaningfully positively associated with empathic avoidance. TriPM Disinhibition was meaningfully positively associated with empathic avoidance, empathic intimacy discomfort, anxiety contagion, and sadness contagion; it was positively associated with empathic restraint. In other words, consistent with prior examination of the nomological networks of these facets of psychopathy (e.g., Sleep et al., 2019), the antagonism associated with Disinhibition appears to be possibly partly driven by empathy-related negative emotionality, such as feeling distress or discomfort as a result of interpersonal interactions.

Aggression (BPAQ) associations. The bitter, resentful hostility associated with aggression (BPAQ Hostility) was specifically characterized by moderate to strong positive associations with empathic intimacy discomfort, anxiety contagion, and sadness contagion. It was also, to a lesser degree, characterized by empathic avoidance and reduced contagion for the positive emotions of others.

Anger-proneness was meaningfully positively associated with both forms of aversive contagion, as well as with empathic intimacy discomfort and empathic avoidance. It also displayed a small negative association with empathic restraint. Interestingly, verbal and physical forms of aggression were only meaningfully associated (negatively) with empathic restraint.

Machiavellianism (MPS) associations. The amorality at the core of Machiavellianism was positively associated with empathic avoidance and empathic intimacy discomfort, and negatively associated with empathic restraint. Machiavellian distrust of others (i.e., perceiving *others* to be amoral and threatening) was positively associated with empathic avoidance, empathic intimacy discomfort, anxiety contagion, and sadness contagion. Appetitive contagion, however, was meaningfully associated with having a more trusting view of others. In general, neither desire for status nor desire for control was meaningfully associated with the unique variances of the specific content scales; empathic restraint, however, was negatively associated with desire for control.

The most surprising finding in the domain of antagonism was that contrary to my predictions and to its distinctively strong association with trait agreeableness, the PT scale added no incremental value above the GFEC scale in statistically predicting any antagonism-related variables. This is especially surprising given that many of the PT scale items explicitly describe a willingness to take the perspectives of others *when in conflict with them*. This result potentially

indicates that this kind of restraining perspective-taking, though facially theoretically different from general caring towards others, may be practically indistinguishable from it. On the other hand, empathic restraint, despite the weak reliability of the scale, was relatively robustly negatively associated almost all antagonism-related variables. Whereas perspective-taking is an effortful process, empathic restraint may be a more automatic, reflexive tendency; perhaps antagonism inhibition is more dependent upon reflexive empathic tendencies than on deliberative empathic tendencies.

Relationships with emotional distress variables. Both the GFEC and the TEQ were positively (though only modestly) associated with self-compassion (FSCRS) and with life satisfaction (SLS). Neither, however, was significantly associated with general worry tendencies (PSWQ). In other words, although trait empathic caring may be associated with being kinder to oneself and with being satisfied with one's life, these wellbeing benefits of being a caring person may be independent of an individual's level of ruminative anxiety.

In terms of the unique variances of the specific content scales, self-compassion was positively associated with appetitive contagion and confidante listening, but negatively associated with all three empathy-related negative emotionality scales (EID, ANXCON, SADCON; *r*s from -.36 to - .51). In other words, the relationship between self-compassion and empathy is complicated, with some evidence in support of Nietzsche's claim that sympathy is associated with lack of self-love, and other evidence indicating the opposite. The relationship very much depends upon how one defines empathy; "matching" for aversive emotions is associated with substantially reduced self-compassion, but "matching" for appetitive emotions is associated with it in the opposite direction.

General worry was not meaningfully associated with any of the empathic caring specific content scales, much as it was not associated with the GFEC scale. General worry was, however, markedly positively associated with the empathy-related negative emotionality scales (EID, ANXCON, SADCON; *r*s from .30 to .64); in particular, anxiety contagion is far more strongly associated with general worry tendencies than with any other external correlate.

Satisfaction with life was not meaningfully associated with any of the empathic caring specific content scales. It was, however, meaningfully associated with all three empathy-related negative emotionality scales (*r*s from -.20 to -.30).

Overall, the unique variances of most specific content aspects of empathic caring appear to be minimally associated with emotional distress, except for appetitive contagion, which is meaningfully associated with being kinder to oneself. The three empathy-related negative emotionality scales, however, were robustly positively associated with general worry and negatively associated with self-compassion and satisfaction for life. These findings are consistent with the marked associations between the empathy-related negative emotionality scales and the HEXACO measure of negative emotionality. What is particularly interesting is that, although most researchers who subscribe to the isomorphic "matching" definition of empathy readily note that empathic distress is a maladaptive phenomenon, the self-reports of matching for aversive emotions appear to be just as clear, or even clearer, predictors of emotional distress than of any domains related to empathic caring or agreeableness.

Relationships with schizotypal personality features. The GFEC and TEQ were both moderately to strongly associated with the constricted affect and/or lack of friendships (SPQ-BR CFCA) associated with schizotypal personality. The TEQ was also meaningfully but modestly associated with suspiciousness and/or ideas of reference (SPQ-BR SUIR). Associations with the

five other SPQ-BR scale scores of schizotypal features were non-significant or minimal. Holistically, it does not appear that a lack of overall empathic caring should be considered characteristic of schizotypy. Although schizotypy has been characterized by a lack of interest in others (e.g., Kwapil & Barrantes-Vidal, 2015), the *appearance* of such disinterest may be generated more by interpersonal discomfort than by an underlying lack of caring about others.

In terms of specific content domains of empathic caring and empathy-related negative emotionality, however, there are some clear and robust patterns. First, anxiety contagion and empathic intimacy discomfort, at both the zero-order level and at the residual variance level, are robustly positively associated with all aspects of schizotypal personality except for Magical Thinking (SPQ-BR MT), with both having moderate/strong associations with social anxiety, odd speech, and suspiciousness and/or ideas of reference.

Second, the only specific content domain associated with Magical Thinking was empathic restraint. Third, appetitive contagion was meaningfully negatively associated with the social anxiety (SPQ-BR SA), constricted affect and/or lack of friends (SPQ-BR CFCA), and suspiciousness and/or ideas of reference (SPQ-BR SUIR) features of schizotypy.

Importantly, although EAV had similarly robust patterns of association with schizotypy features as did EID, its relationships with social anxiety and constricted affect and/or lack of friends were substantially weaker. This finding may suggest that empathic avoidance in schizotypy is primarily associated with the motivation of discomfort/distress in interpersonal encounters, not with any kind of emotional coldheartedness (cf. findings that lack of affect in schizophrenia is not driven by lack of emotional experiencing, e.g., Kring & Neale, 1996).

In sum, schizotypal personality appears to holistically be characterized by heightened empathy avoidance, proneness to empathic intimacy discomfort, and contagion for the anxiety of

others. Although not as holistically, it also appears to be positively associated with sadness contagion and negatively associated with appetitive contagion. In other words, at least from this particular self-report perspective, any empathy abnormalities in schizotypy appear to be driven by a tendency to find empathic encounters aversively contagious, distressing, and comparatively unrewarding. As a result, empathic encounters tend to be avoided.

As a side note, relevant to the critique of the "matching" definition of empathy and its promotion by some neuroscientists, consider what these findings might mean in terms of predicting empathy in schizotypy in the typical fMRI "pain empathy" paradigm. Although (to my knowledge) there has not yet been an fMRI "pain empathy" study among individuals with schizotypal personality disorder, there has been one basic such study for schizophrenia (Horan et al., 2016). In that study, in the "empathy" condition, patients with schizophrenia had significantly higher levels of brain activation in the regions most associated with "pain empathy" (the anterior cingulate cortex and anterior insula) than did healthy controls; when passively observing the stimuli without instruction, though, the two groups had similar levels of brain activation. In other words, schizophrenia, from the general perspective of this paradigm (though not the interpretation of Horan et al., 2016), appears to be associated with heightened "empathy;" in truth, though, this brain activation is probably indicating less regulation of empathic distress.

Limitations and Future Directions

The most glaring limitation of these dissertation studies is that these new scales were only examined for convergent and discriminant validity in relationship to other self-report scales. This is an unfortunate limitation extending throughout the field, as it seems the validity evidence for new empathy questionnaires is largely founded on correlations with other empathy questionnaires and conceptually related self-reported constructs. Until validity evidence is

produced for these scales beyond relationships with other self-report measures, any theoretical conclusions drawn from their use (including conclusions drawn by me in this dissertation), should be regarded with substantial caution.

Although behavioral assessments of cognitive empathy ability are common (Murphy & Lilienfeld, 2019), behavioral measures of empathic caring have been sparsely investigated. The most frequent such method is to ask respondents to donate to a charity, and a number of studies have observed a positive association between general empathic caring and donation decisions within or beyond the laboratory context (e.g., Davis, 1983; Jordan, Amir, & Bloom, 2016; Verhaert & Van den Poel, 2011). Yet, this kind of task is extremely limited in what it measures, and could be confounded by any number of factors, such as personal distress, demand characteristics, and social desirability motives.

Ideally, I would like to examine the predictiveness of these new scales in a longitudinal examination of real-world interpersonal behavior. Following the perspective that empathy emerges out of mammalian caregiving (e.g., de Waal, 2008; Stern & Cassidy, 2018) and may be highly related to relationship emergence and maintenance, it would be valuable to investigate whether these new scales, administered to incoming freshman university students, predict characteristics of relationship emergence (both romantic relationships and friendships) and maintenance, such as feelings of authenticity, safety, love, and non-egoistic relationship motives (e.g., not based in desire for popularity or status).

Moreover, it would be useful to generate other-report versions of these new scales and investigate their functioning. Especially given the concerns with social desirability in self-report measures of empathy (e.g., Batson et al., 2016), it would be important to assess to what extent self-reports converge with other-reports in this domain. Prior investigations of the convergence

of parent-report and self-report for unidimensional total scores of empathy questionnaires have observed moderate or strong correlations (e.g., Dadds et al., 2008; Johnson, Filliter, & Murphy, 2009), but this remains an under-explored topic, especially at the specific content domain level. It is possible that the extent of convergence may be entirely due to a general factor of agreeableness or "kindness," with specific content domains demonstrating little or no such convergence (especially if examined at the residual level). It is also likely that, even if there is substantial self-other report convergence for specific content domains, it may heavily depend upon the particular domain; some domains, such as the emotional contagion domains, may relate to more "private" information that is not easily observable to others, whereas the confidante listening domain may be more directly observable.

These two studies are also strongly limited in that they both utilized online paid participant samples from M-Turk. The factor structure of the scales may be different in other samples, such as undergraduate or clinical samples. In general, I would predict that such samples may demonstrate a cleaner factor structure than observed in these two studies; given the heightened degree of satisficing found (and theoretically expected) in M-Turk samples (Hamby & Taylor, 2016), the various empathic caring specific content domain scales may be less strongly associated in other samples. It is also possible that the GFEC scale may fail to replicate as consisting of "pure indicators" of the construct in other samples. If, however, the GFEC is replicated in this fashion in other samples, that would strongly corroborate its theoretical and psychometric value.

Despite substantial limitations, these dissertation studies generated important overarching implications for the conceptualization of empathy, such as corroborating the highly distinguishable multidimensionality of empathic contagion observed by Murphy et al. (2018b);

providing tentative support for the critical need to measure and discuss empathy in terms of approach-avoidance behaviors and motives, as proposed by Weisz and Zaki (2018); and pointing to confidante listening, or perhaps other proto-Rogerian elements of empathy, as a potentially neglected but important aspect of the general construct of empathic caring. Moreover, these studies provided more fine-grained examinations of aspects of empathy in relation to external variables, such as further illuminating how aversive contagion and empathic discomfort, rather than coldheartedness, help to account for "empathy abnormalities" in schizotypal personality; and demonstrating that the relationship between empathy and self-compassion depends on which aspect of empathy is considered.

The eight new scales generated in this dissertation are likely to be a substantial improvement upon existing self-report measures, with heightened conceptual clarity and improved psychometric properties. Although they need to undergo further refinement and validation, I suspect that they could emerge as valuable tools for empathy's scientific community of inquiry.

References

Aaron, R. V., Benson, T. L., & Park, S. (2015). Investigating the role of alexithymia on the empathic deficits found in schizotypy and autism spectrum traits. *Personality and Individual Differences*, *77*, 215-220.

Abad, F. J., Garcia-Garzon, E., Garrido, L. E., & Barrada, J. R. (2017). Iteration of partially specified target matrices: Application to the bi-factor case. *Multivariate Behavioral Research*, *52*(4), 416-429.

Abdallah, T. (1998). The Satisfaction with Life Scale (SWLS): Psychometric properties in an Arabic-speaking sample. *International journal of Adolescence and Youth*, *7*(2), 113-119.

Allport, G. W. (1924). The study of the undivided personality. *The Journal of Abnormal Psychology and Social Psychology*, *19*(2), 132-141.

Allport, G. W. (1929). The study of personality by the intuitive method. An experiment in teaching from The Locomotive God. *The Journal of Abnormal and Social Psychology*, *24*(1), 14-27.

Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: Holt.

Allport, G. W. (1942). *The use of personal documents in psychological science.* (Social Science Research Council, Bulletin 49). Ann Arbor, MI: Edwards Brothers.

Allport, G.W. (1961) *Pattern and Growth in Personality*. New York: Holt, Rinehart, & Winston.

Allport, G. W. (1968). The historical background of modern social psychology. In G. Lindzey & E. Aronson (Eds.), *Handbook of Social Psychology* (pp. 3-56). Reading, MA: Addison-Wesley.

Alterman, A. I., McDermott, P. A., Cacciola, J. S., & Rutherford, M. J. (2003). Latent structure of the Davis Interpersonal Reactivity Index in methadone maintenance patients. *Journal of Psychopathology and Behavioral Assessment, 25*, 257–265.

Ames, D. R., & Kammrath, L. K. (2004). Mind-reading and metacognition: Narcissism, not actual competence, predicts self-estimated ability. *Journal of Nonverbal Behavior, 28,* 187–209.

Ammirati, R. J. (2013). *Self-Assessed Emotion Recognition Skill and Social Adjustment among College Students*. Doctoral dissertation, Emory University, Atlanta, GA.

Anderson, A. E., & Marcus, D. K. (2019). A bifactor model of meanness, coldheartedness, callousness, and sadism. *Personality and Individual Differences, 137,* 192-197.

Andrew, J., Cooke, M., & Muncer, S. J. (2008). The relationship between empathy and Machiavellianism: An alternative to empathizing–systemizing theory. *Personality and Individual Differences*, *44*(5), 1203-1211.

Arrindell, W. A., Meeuwesen, L., & Huyse, F. J. (1991). The Satisfaction With Life Scale (SWLS): Psychometric properties in a non-psychiatric medical outpatients sample. *Personality and Individual Differences*, *12*(2), 117-123.

Aruguete, M. S., Huynh, H., Browne, B. L., Jurs, B., Flint, E., & McCutcheon, L. E. (2019). How serious is the 'carelessness' problem on Mechanical Turk?. *International Journal of Social Research Methodology*, 1-9.

Asai, T., Mao, Z., Sugimori, E., & Tanno, Y. (2011). Rubber hand illusion, empathy, and schizotypal experiences in terms of self-other representations. *Consciousness and Cognition*, *20*(4), 1744-1750.

Ashar, Y. K., Andrews-Hanna, J. R., Dimidjian, S., & Wager, T. D. (2017). Empathic care and distress: predictive brain markers and dissociable brain systems. *Neuron, 94*(6), 1263-1273.

Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment, 91,* 340–345.

Aspy, D. N. (1975). Empathy: Let's get the hell on with it. *The Counseling Psychologist*, *5*(2), 10-14.

Bäckström, M., Björklund, F., & Larsson, M. R. (2009). Five-factor inventories have a major general factor related to social desirability which can be reduced by framing items neutrally. *Journal of Research in Personality*, *43*(3), 335-344.

Bandalos, D. L., & Finney, S. J. (2010). Factor analysis: Exploratory and confirmatory. *The reviewer's guide to quantitative methods in the social sciences*, 93-114.

Barends, A. J., & de Vries, R. E. (2019). Noncompliant responding: Comparing exclusion criteria in MTurk personality research to improve data quality. *Personality and Individual Differences*, *143*, 84-89.

Barnett, G., & Mann, R. E. (2013). Empathy deficits and sexual offending: A model of obstacles to empathy. *Aggression and Violent Behavior*, *18*, 228–239.

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. *The Journal of Child Psychology and Psychiatry and Allied Disciplines, 42*(2), 241-251.

Batchelder, L., Brosnan, M., & Ashwin, C. (2017). The Development and Validation of the Empathy Components Questionnaire (ECQ). *PLoS One, 12*. e0169185.

Batson, C. D. (1991). *The Altruism Question: Toward a social-psychological answer*. Hillsdale, NJ: Erlbaum.

Batson, CD. (2009). These things called empathy. In *The Social Neuroscience of Empathy*, ed. J Decety, W Ickes, pp. 16–31. Cambridge, MA: MIT Press.

Batson, C. D., Ahmad, N., Lishner, D. A., & Tsang, J. (2005). Empathy and altruism. In C. R. Snyder & S. J. Lopez (Eds.), *Handbook of Positive Psychology* (pp. 485–498). Oxford, UK: Oxford University Press.

Batson, C. D., Duncan, B. D., Ackerman, P., Buckley, T., & Birch, K. (1981). Is empathic emotion a source of altruistic motivation?. *Journal of Personality and Social Psychology*, *40*(2), 290.

Batson, C. D., Fultz, J., & Schoenrade, P. A. (1987). Distress and empathy: Two qualitatively distinct vicarious emotions with different motivational consequences. *Journal of Personality, 55*(1), 19-39.

Bedwell, J. S., Compton, M. T., Jentsch, F. G., Deptula, A. E., Goulding, S. M., & Tone, E. B. (2014). Latent factor modeling of four schizotypy dimensions with theory of mind and empathy. *PloS one*, *9*(11), e113853.

Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods, 43*(3), 800-813.

Bergin, A. E., & Jasper, L. G. (1969). Correlates of empathy in psychotherapy: A replication. *Journal of Abnormal Psychology*, *74*(4), 477-481.

Berlin, I. (2000). *The Proper Study of Mankind: An anthology of essays*. New York: Macmillan.

Berg, J. (2018). Lumping and splitting. *Science*, *359*, 1309-1309.

Bird, G., & Viding, E. (2014). The self to other model of empathy: providing a new framework for understanding empathy impairments in psychopathy, autism, and alexithymia. *Neuroscience & Biobehavioral Reviews*, *47*, 520-532.

Bloom, P. (2017a). *Against Empathy: The case for rational compassion*. New York: Random House.

Bloom, P. (2017b). Empathy and its discontents. *Trends in Cognitive Sciences*, *21*(1), 24-31.

Bonifay, W., & Cai, L. (2017). On the complexity of item response theory models. *Multivariate Behavioral Research*, *52*(4), 465-484.

Bonifay, W., Lane, S. P., & Reise, S. P. (2017). Three concerns with applying a bifactor model as a structure of psychopathology. *Clinical Psychological Science*, *5*, 184-186.

Bornovalova, M. A., Hicks, B. M., Patrick, C. J., Iacono, W. G., & McGue, M. (2011). Development and validation of the Minnesota Borderline Personality Disorder scale. *Assessment*, *18*(2), 234-252.

Bozarth, J. (1999). Empathy from the framework of client-centered theory and the rogerian hypothesis. In A. Bohart & L. Greenberg (Eds.), *Empathy reconsidered: New directions in psychotherapy* (2nd ed., pp. 81–102). Washington, DC: American Psychological Association.

Bozarth, J. (2009). Rogerian empathy in an organismic theory: A way of being. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (1st ed., pp. 101–112). Cambridge, MA: The MIT Press.

Briggs, N. E., & MacCallum, R. C. (2003). Recovery of weak common factors by maximum likelihood and ordinary least squares estimation. *Multivariate Behavioral Research, 38*(1), 25-56.

Brook, M., & Kosson, D. S. (2013). Impaired cognitive empathy in criminal psychopathy: Evidence from a laboratory measure of empathic accuracy. *Journal of Abnormal Psychology, 122*(1), 156-166.

Brown, H. C. (1915). The Beautiful. *The Journal of Philosophy, Psychology and Scientific Methods, 12*(23), 640-641.

Bryan, C. J., & Harris, J. A. (2019). The Structure of Suicidal Beliefs: A Bifactor Analysis of the Suicide Cognitions Scale. *Cognitive Therapy and Research, 43*(2), 335-344.

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

Buchheimer, A. (1963). The development of ideas about empathy. *Journal of Counseling Psychology, 10,* 61–70

Buhrmester, M. D., Talaifar, S., & Gosling, S. D. (2018). An evaluation of Amazon's Mechanical Turk, its rapid rise, and its effective use. *Perspectives on Psychological Science, 13*(2), 149-154.

Burgess, E. W., & Cottrell, L. S. (1939). *Predicting success or failure in marriage*. New York: Prentice Hall.

Buss, A. H., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology*, *63*(3), 452.

Cadwallader, T. C. (1974). Charles S. Peirce (1839-1914): The first American experimental psychologist. *Journal of the History of the Behavioral Sciences*, *10*(3), 291-298.

Callaway, D. A., Cohen, A. S., Matthews, R. A., & Dinzeo, T. (2014). Schizotypal Personality Questionnaire—Brief Revised: Psychometric replication and extension. *Personality Disorders: Theory, Research, and Treatment*, *5*(1), 32-38.

Carey, S., (1985). *Conceptual Change in Childhood*. MIT Press, Cambridge, MA.

Carré, A., Stefaniak, N., D'ambrosio, F., Bensalah, L., & Besche-Richard, C. (2013). The Basic Empathy Scale in Adults (BES-A): Factor structure of a revised form. *Psychological Assessment*, *25*(3), 679.

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, *29*(6), 2156-2160.

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., ... & Moffitt, T. E. (2014). The p factor: one general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science*, *2*, 119-137.

Castilho, P., Pinto-Gouveia, J., & Duarte, J. (2015). Exploring self-criticism: confirmatory factor analysis of the FSCRS in clinical and nonclinical samples. *Clinical Psychology & Psychotherapy*, *22*(2), 153-164.

Cattell, R. B. (1937). Measurement versus intuition in applied psychology. *Journal of Personality*, *6*, 114–131.

Cattell. R. B. *The Scientific Analysis ofPpersonality.* Baltimore: Penguin. 1965.

Catterson, A. D., Naumann, L. P., & John, O. P. (2015). Confidence of social judgments is not just error: Individual differences in the structure, stability, and social functions of perceptual confidence. *Journal of Research in Personality, 58,* 11–20.

Chan, D. (2009). So Why Ask Me? Are Self-Report Data Really That Bad?. In *Statistical and Methodological Myths and Urban Legends* (pp. 329-356). Routledge.

Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods*, *46*(1), 112-130.

Chemerinski, E., Triebwasser, J., Roussos, P., & Siever, L. J. (2013). Schizotypal personality disorder. *Journal of Personality Disorders*, *27*(5), 652-679.

Chen, J. I., Osman, A., Freedenthal, S. L., & Gutierrez, P. M. (2019). An examination of the psychometric properties of the reasons for living inventory within a male veteran clinical sample. *Archives of Suicide Research*.

Chin, E. G., Buchanan, E. M., Ebesutani, C., & Young, J. (2018). Depression, Anxiety, and Stress: How should clinicians interpret the total and subscale scores of the 21-item Depression, Anxiety, and Stress Scales?. *Psychological Reports*, 0033294118783508.

Christov-Moore, L. and Iacoboni, M. (2014) Forum: against empathy. *Boston Review* August

Chyung, S. Y., Roberts, K., Swanson, I., & Hankinson, A. (2017). Evidence-based survey design: The use of a midpoint on the Likert scale. *Performance Improvement*, *56*(10), 15-23.

Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology, 100*(3), 316-336.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309-319.

Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment*.

Cleckley, H. (1988). *The Mask of Sanity: An attempt to clarify some issues about the so called psychopathic personality* (5th ed.). St. Louis: Mosby.

Cohen, A. S., Matthews, R. A., Najolia, G. M., & Brown, L. A. (2010). Toward a more psychometrically sound brief measure of schizotypal traits: introducing the SPQ-Brief Revised. *Journal of Personality Disorders, 24*(4), 516-537.

Cohen, A. S., Mohr, C., Ettinger, U., Chan, R. C., & Park, S. (2015). Schizotypy as an organizing framework for social and affective sciences. *Schizophrenia Bulletin, 41*(Suppl 2), S427-35.

Coll, M. P., Viding, E., Rütgen, M., Silani, G., Lamm, C., Catmur, C., & Bird, G. (2017). Are we really measuring empathy? Proposal for a new measurement framework. *Neuroscience & Biobehavioral Reviews, 83*, 132-139.

Colman, A. M. (2009). *A dictionary of psychology*. Oxford, UK: Oxford University Press.

Coplan, A. (2011). Will the real empathy please stand up? A case for a narrow conceptualization. *The Southern Journal of Philosophy, 49*, 40-65.

Costello, A. B., & Osborne, J. W. (2005) Best Practices in Exploratory Factor Analysis: Four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation*, *10*(7), 2.

Cottrell, L. S. (1950). Some neglected problems in social psychology. *American Sociological Review*, *15*(6), 705-712.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334.

Cronbach, L. J. (1955). Processes affecting scores on" understanding of others" and" assumed similarity." *Psychological Bulletin*, *52*(3), 177-193.

Cuff, B. M., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: a review of the concept. *Emotion Review*, *8*(2), 144-153.

Dadds, M. R., Hunter, K., Hawes, D. J., Frost, A. D., Vassallo, S., Bunn, P., ... & El Masry, Y. (2008). A measure of cognitive and affective empathy in children using parent ratings. *Child Psychiatry and Human Development*, *39*(2), 111-122.

Dahling, J. J., Whitaker, B. G., & Levy, P. E. (2009). The development and validation of a new Machiavellianism scale. *Journal of Management*, *35*(2), 219-257.

Darwall, S. (1998). Empathy, sympathy, care. *Philosophical Studies*, *89*(2), 261-282.

Davies, S. E., Connelly, B. S., Ones, D. S., & Birkland, A. S. (2015). The General Factor of Personality: The “Big One,” a self-evaluative trait, or a methodological gnat that won’t go away?. *Personality and Individual Differences*, *81*, 13-22.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*(1), 113-126.

Davis, M. H., & Kraus, L. A. (1997). Personality and empathic accuracy. In W. Ickes (Ed.), *Empathic accuracy* (pp. 144–168). New York, NY: Guilford Press.

De Ayala, R. J. (2013). The IRT tradition and its applications. *Oxford Handbook of Quantitative Methods: Foundations, 1*, 144-169.

Debbané, M., & Barrantes-Vidal, N. (2014). Schizotypy from a developmental perspective. *Schizophrenia bulletin, 41*(2), S386-S395.

Debes, R. (2015). From Einfühlung to empathy: Sympathy in early phenomenology and psychology. In E.Schliesser (Ed.), *Sympathy: A history* (pp. 286–322). New York: Oxford University Press.

Decety, J. (2015). The neural pathways, development and functions of empathy. *Current Opinion in Behavioral Sciences, 3*, 1-6.

Decety, J., & Cowell, J. M. (2014). The complex relation between morality and empathy. *Trends in Cognitive Sciences, 18*(7), 337-339.

Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews, 3*(2), 71-100.

Decety, J., & Lamm, C. (2006). Human empathy through the lens of social neuroscience. *The Scientific World Journal, 6*, 1146–1163.

Decety, J., & Meyer, M. (2008). From emotion resonance to empathic understanding: A social developmental neuroscience account. *Development and Psychopathology, 20*(4), 1053-1080.

De Corte, K., Buysse, A., Verhofstadt, L. L., Roeyers, H., Ponnet, K., & Davis, M. H. (2007). Measuring empathic tendencies: Reliability and validity of the Dutch version of the Interpersonal Reactivity Index. *Psychologica Belgica*, *47*(4), 235-260.

Deetlefs, J., Chylinski, M., & Ortmann, A. (2015). MTurk 'Unscrubbed': Exploring the good, the 'Super', and the unreliable on Amazon's Mechanical Turk. *UNSW Business School Research Paper*, (2015-20A).

Dekeyser, M., Elliott, R., & Leijssen, M. (2009). Empathy in psychotherapy: Dialogue and embodied understanding. In J. Decety & W. Ickes (Eds.), *The Social Neuroscience of Empathy* (1st ed., pp. 113–124). Cambridge: The MIT Press.

Derntl, B., Finkelmeyer, A., Toygar, T. K., Hülsmann, A., Schneider, F., Falkenberg, D. I., & Habel, U. (2009). Generalized deficit in all core components of empathy in schizophrenia. *Schizophrenia Research*, *108*(1), 197-206.

Deveney, C. M., Stoddard, J., Evans, R. L., Chavez, G., Harney, M., & Wulff, R. A. (2019). On defining irritability and its relationship to affective traits and social interpretations. *Personality and Individual Differences*, *144*, 61-67.

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

de Waal, F. (2008). Putting the Altruism Back into Altruism: The evolution of empathy. *Annual Review of Psychology*, *59*, 279-300.

de Waal, F. B., & Preston, S. D. (2017). Mammalian empathy: behavioural manifestations and neural basis. *Nature Reviews Neuroscience*, *18*(8), 498-510.

Dick, P.K. (1968). *Do Androids Dream of Electric Sheep?*. New York: Ballantine Books.

Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49*(1), 71-75.

Dinsdale, N., & Crespi, B. J. (2013). The borderline empathy paradox: evidence and conceptual models for empathic enhancements in borderline personality disorder. *Journal of Personality Disorders*, *27*(2), 172-195.

DiStefano, C., & Motl, R. W. (2009). Personality correlates of method effects due to negatively worded items on the Rosenberg Self-Esteem scale. *Personality and Individual Differences*, *46*, 309-313.

Dombrowski, S. C., Beaujean, A. A., McGill, R. J., & Benson, N. F. (2018). The Woodcock-Johnson IV Tests of Achievement Provides Too Many Scores for Clinical Interpretation. *Journal of Psychoeducational Assessment*, 0734282918800745.

Ducasse, C. J. (1928). What has Beauty to do with Art?. *The Journal of Philosophy*, *25*(7), 181-186.

Durak, M., Senol-Durak, E., & Gencoz, T. (2010). Psychometric properties of the satisfaction with life scale among Turkish university students, correctional officers, and elderly adults. *Social Indicators Research*, *99*(3), 413-429.

Dyck, M. J. (2012). The ability to understand the experience of other people: development and validation of the emotion recognition scales. *Australian Psychologist*, *47*(1), 49-57.

Dymond, R. (1948). A preliminary investigation of the relation of insight and empathy. *Journal of Consulting Psychology*, *12*, 228–233.

Dymond, R. F. (1949). A scale for the measurement of empathic ability. *Journal of Consulting Psychology*, *13*(2), 127-133.

Dymond, R. F. (1950). Personality and empathy. *Journal of Consulting Psychology*, *14*(5), 343 350.

Dymond, R. (1953). Can clinicians predict individual behavior? *Journal of Personality*, *22*, 151–161.

Dymond, R., Hughes, A., & Raabe, V. (1952). Measurable changes in empathy with age. *Journal of Consulting Psychology*, *16*, 202–206.

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.

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.

Ebesutani, C., Reise, S. P., Chorpita, B. F., Ale, C., Regan, J., Young, J., ... & Weisz, J. R. (2012). The Revised Child Anxiety and Depression Scale-Short Version: Scale reduction via exploratory bifactor modeling of the broad anxiety factor. *Psychological Assessment*, *24*(4), 833.

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), 269-281.

Einstein, A. (1932) Prologue. In M. Planck, *Where is Science Going?* (ps. 7-12). New York: W.W. Norton.

Eisenberg, N., & Eggum, N. D. (2009). Empathic responding: Sympathy and personal distress. In J. Decety & W. Ickes (Eds.), *The Social Neuroscience of Empathy* (pp. 71–83). Cambridge, MA: MIT Press.

Eisenberg, N., & Fabes, R.A. (1992). Emotion, regulation, and the development of social competence. In M.S. Clark (Ed.), *Review of Personality and Social Psychology: Emotion and Social Behavior* (ps. 119–50). Newbury Park, CA: Sage.

Eisenberg, N., Fabes, R.A., & Spinrad, T.L. (2006). Prosocial orientation. In W. Damon & R.M. Lerner (Series Eds.), & N. Eisenberg (Vol. Ed.), *Handbook of Child Psychology: Vol. 3. Social, Emotional, and Personality development* (6th ed., pp. 646–718). New York: Wiley.

Eisenberg, N., & Strayer, J. (1987). Critical issues in the study of empathy. In N.Eisenberg & J.Strayer (Eds.), *Empathy and its Development* (pp. 3–13). New York: Cambridge University Press.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.

Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology, 37,* 1097–1126.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272-299.

Fernández, A. M., Dufey, M., & Kramp, U. (2011). Testing the psychometric properties of the Interpersonal Reactivity Index (IRI) in Chile: Empathy in a different cultural context. *European Journal of Psychological Assessment*, *27*(3), 179-185.

Ferrando, P. J., & Lorenzo-Seva, U. (2019). An External Validity Approach for Assessing Essential Unidimensionality in Correlated-Factor Models. *Educational and Psychological Measurement*, 0013164418824755.

Feshbach, N. D. (1975). Empathy in children: Some theoretical and empirical considerations. *The Counseling Psychologist*, *5*(2), 25-30.

Feshbach, N. D., & Feshbach, S. (1969). The relationship between empathy and aggression in two age groups. *Developmental Psychology*, *1*(2), 102-107.

Feshbach, N.D., & Feshbach, S. (2009). Empathy and education. In: Decety J, Ickes WJ, (Eds.). *The Social Neuroscience of Empathy* (pp. 85-98). Cambridge: The MIT Press.

Fisher, J. A. (2017, November). Empathic Actualities: Toward a Taxonomy of Empathy in Virtual Reality. In *Interactive Storytelling: 10th International Conference on Interactive Digital Storytelling, ICIDS 2017 Funchal, Madeira, Portugal, November 14–17, 2017, Proceedings* (Vol. 10690, p. 233-243).

Fisher, R. J., & Katz, J. E. (2000). Social-desirability bias and the validity of self-reported values. *Psychology & Marketing*, *17*(2), 105-120.

Foster, J. D., McCain, J. L., Hibberts, M. F., Brunell, A. B., & Johnson, R. B. (2015). The grandiose narcissism scale: a global and facet-level measure of grandiose narcissism. *Personality and Individual Differences*, *73*, 12-16.

Gadermann, A. M., Schonert-Reichl, K. A., & Zumbo, B. D. (2010). Investigating validity evidence of the satisfaction with life scale adapted for children. *Social Indicators Research*, *96*(2), 229-247.

Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly*, *73*(2), 349-360.

Gallese, V. (2005). Embodied simulation: From neurons to phenomenal experience. *Phenomenology and the Cognitive Sciences*, *4*(1), 23-48.

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

Garcia-Garzon, E., Abad, F. J., & Garrido, L. E. (2019). Improving Bi-Factor Exploratory Modeling. *Methodology*, 15, 45-55.

Garfield, C. F., Fisher, S. D., Barretto, D., Rutsohn, J., & Isacco, A. (2019). Development and validation of a father involvement in health measure. *Psychology of Men & Masculinities*, *20*(1), 148-160.

Gerber, Z., Tolmacz, R., & Doron, Y. (2015). Self-compassion and forms of concern for others. *Personality and Individual Differences*, *86*, 394-400.

Gerdes, K. E., Segal, E. A., & Lietz, C. A. (2010). Conceptualising and measuring empathy. *British Journal of Social Work*, *40*(7), 2326-2343.

Gilbert, P., Clarke, M., Hempel, S., Miles, J. N., & Irons, C. (2004). Criticizing and reassuring oneself: An exploration of forms, styles and reasons in female students. *British Journal of Clinical Psychology*, *43*(1), 31-50.

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of MTurk samples. *Journal of Behavioral Decision Making*, 26, 213–224.

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*, *138*(6), 1085.

Green, E. D. (2016). What are the most-cited publications in the social sciences (according to Google Scholar)?. *Impact of Social Sciences Blog*.

Gross, J. J. (1998). Antecedent-and response-focused emotion regulation: divergent consequences for experience, expression, and physiology. *Journal of Personality and Social Psychology*, *74*(1), 224.

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, *85*(2), 348.

Grosz, M. P., Emons, W. H., Wetzel, E., Leckelt, M., Chopik, W. J., Rose, N., & Back, M. D. (2019). A comparison of unidimensionality and measurement precision of the Narcissistic Personality Inventory and the Narcissistic Admiration and Rivalry Questionnaire. *Assessment*, *26*(2), 281-293.

Grynberg, D., Luminet, O., Corneille, O., Grèzes, J., & Berthoz, S. (2010). Alexithymia in the interpersonal domain: A general deficit of empathy?. *Personality and Individual Differences*, *49*(8), 845-850.

Haberman, S. J. (2008). When can subscores have value? *Journal of Educational and Behavioral Statistics*, *33*, 204-229.

Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, *133*(4), 55-66.

Halamová, J., Kanovský, M., Gilbert, P., Troop, N. A., Zuroff, D. C., Hermanto, N., ... & Krieger, T. (2018). The factor structure of the Forms of Self-Criticising/Attacking & Self-Reassuring Scale in thirteen distinct populations. *Journal of Psychopathology and Behavioral Assessment*, *40*(4), 736-751.

Hall, J. A., Andrzejewski, S. A., & Yopchick, J. E. (2009). Psychosocial correlates of interpersonal sensitivity: A meta-analysis. *Journal of Nonverbal Behavior, 33,* 149–180.

Hall, J. A., & Schwartz, R. (2019). Empathy present and future. *The Journal of Social Psychology, 159*(3), 225-243.

Hamby, T., & Taylor, W. (2016). Survey satisficing inflates reliability and validity measures: An experimental comparison of college and Amazon Mechanical Turk samples. *Educational and Psychological Measurement*, *76*(6), 912-932.

Hare, R. D. (1993). *Without Conscience: The disturbing world of the psychopaths among us.* New York, NY: Simon & Schuster.

Hartshorne, C., Weiss, P., & Burks, A. W. (Eds.). (1931). *Collected Papers of Charles Sanders Peirce* (Vol. 1). Cambridge: Harvard University Press.

Hatcher, S. L., Nadeau, M. S., Walsh, L. K., Reynolds, M., Galea, J., & Marz, K. (1994). The teaching of empathy for high school and college students: Testing Rogerian methods with the Interpersonal Reactivity Index. *Adolescence*, *29*(116), 961.

Hatfield, E., Rapson, R.L., & Le, Y.C. (2009). Emotional contagion and empathy. In J. Decety & W. Ickes (Eds.), *The Social Neuroscience of Empathy* (pp. 19-30). Cambridge, MA: MIT Press.

Hauser, D., Paolacci, G., & Chandler, J. J. (2019). Common concerns with MTurk as a participant pool: Evidence and solutions. (Chapter to appear in *Handbook in Research Methods in Consumer Psychology*, Eds: Kardes, F., Herr, P., and Schwarz, N)

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, *48*(1), 400-407.

Hawk, S. T., Keijsers, L., Branje, S. J., Graaff, J. V. D., Wied, M. D., & Meeus, W. (2013). Examining the interpersonal reactivity index (IRI) among early and late adolescents and their mothers. *Journal of Personality Assessment, 95*(1), 96-106.

Hayes, A., 2018. The PROCESS Macro for SPSS and SAS. http://www.processmacro. Org

Hébert, A., & Weaver, A. (2014). An examination of personality characteristics associated with BDSM orientations. *The Canadian Journal of Human Sexuality*, *23*(2), 106-115.

Heilbrun, A. B. (1982). Cognitive models of criminal violence based upon intelligence and psychopathy levels. *Journal of Consulting and Clinical Psychology, 50*(4), 546–557.

Heinrich, L. M., & Gullone, E. (2006). The clinical significance of loneliness: A literature review. *Clinical Psychology Review, 26*(6), 695-718.

Henry, J. D., Bailey, P. E., & Rendell, P. G. (2008). Empathy, social functioning and schizotypy. *Psychiatry Research, 160*(1), 15-22.

Hepper, E. G., Hart, C. M., Meek, R., Cisek, S., & Sedikides, C. (2014). Narcissism and empathy in young offenders and non-offenders. *European Journal of Personality, 28*(2), 201-210.

Herder, J. G. (2002). This too a philosophy of history for the formation of humanity (M. N.Forster, Trans.). In M. N.Forster (Ed.), *Herder: Philosophical Writings* (pp. 272–358). New York: Cambridge University Press.

Hodges, S. D., & Biswas-Diener, R. (2007). Balancing the empathy expense: Strategies for regulating empathic response. In T. F. D. Farrow & P. W. R. Woodruff (Eds.), *Empathy in mental illness* (pp. 389–407). Cambridge, UK: Cambridge University Press.

Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology*, *33*(3), 307-316.

Hogan, R. (1975). Empathy: A conceptual psychometric analysis. *The Counseling Psychologist, 5*(2), 14-18.

Holzinger, K.J., Swineford, F. (1937). The bi-factor method. *Psychometrika,* 2, 41–54.

Hooper, N., Erdogan, A., Keen, G., Lawton, K., & McHugh, L. (2015). Perspective taking reduces the fundamental attribution error. *Journal of Contextual Behavioral Science, 4*(2), 69-72.

Horan, W. P., Jimenez, A. M., Lee, J., Wynn, J. K., Eisenberger, N. I., & Green, M. F. (2016). Pain empathy in schizophrenia: an fMRI study. *Social Cognitive and Affective Neuroscience, 11*(5), 783-792.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179-185.

Howes, E. P. (1913). *Æ*sthetics. *Psychological Bulletin, 10*(5), 196-201.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55.

Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology, 27*(1), 99-114.

Hurlemann, R., Patin, A., Onur, O. A., Cohen, M. X., Baumgartner, T., Metzler, S., ... & Kendrick, K. M. (2010). Oxytocin enhances amygdala-dependent, socially reinforced learning and emotional empathy in humans. *Journal of Neuroscience, 30*(14), 4999-5007.

Iannotti, R. J. (1975). The nature and measurement of empathy in children. *The Counseling Psychologist, 5*(2), 21-25.

Ickes, W. (1993). Empathic accuracy. *Journal of Personality, 61*, 587–610.

Ickes, W., Stinson, L., Bissonnette, V., & Garcia, S. (1990). Naturalistic social cognition: Empathic accuracy in mixed-sex dyads. *Journal of Personality and Social Psychology*, *59*(4), 730.

Ilyes, I. (2017). Empathy in Hume and Smith. In *The Routledge Handbook of Philosophy of Empathy* (pp. 98-109). Routledge.

Jackson, P. L., Brunet, E., Meltzoff, A. N., & Decety, J. (2006). Empathy examined through the neural mechanisms involved in imagining how I feel versus how you feel pain. *Neuropsychologia, 44*, 752–761.

Jahoda, G. (2005). Theodor Lipps and the shift from "sympathy" to "empathy". *Journal of the History of the Behavioral Sciences, 41*(2), 151-163.

James, W. (2000). *Pragmatism and Other Writings*. New York: Penguin.

Jardine, J., & Szanto, T. (2017). Empathy in the phenomenological tradition. *Routledge Handbook of Philosophy of Empathy* (86-97). New York: Routledge.

Jennrich, R.I., Bentler, P.M. (2011). Exploratory bifactor analysis. *Psychometrika, 76*, 537-549.

Jennrich, R.I., Bentler, P.M. (2012). Exploratory bifactor analysis: the oblique case. *Psychometrika, 77*, 442-454

Johnson, J. A., Cheek, J. M., & Smither, R. (1983). The structure of empathy. *Journal of Personality and Social Psychology, 45*(6), 1299-1312.

Johnson, S. A., Filliter, J. H., & Murphy, R. R. (2009). Discrepancies between self-and parent-perceptions of autistic traits and empathy in high functioning children and adolescents on the autism spectrum. *Journal of Autism and Developmental Disorders, 39*(12), 1706-1714.

Joireman, J. (2004). Relationships Between Attributional Complexity and Empathy. *Individual Differences Research, 2*(3), 197-202.

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.

Jones, A. L. (1905). Die Einfühlung und das Symbol. *The Journal of Philosophy, Psychology and Scientific Methods, 2*(23), 639-641.

Jordan, M. R., Amir, D., & Bloom, P. (2016). Are empathy and concern psychologically distinct?. *Emotion, 16*(8), 1107.

Judge, T. A., Erez, A., Bono, J. E., & Thoresen, C. J. (2002). Are measures of self-esteem, neuroticism, locus of control, and generalized self-efficacy indicators of a common core construct?. *Journal of Personality and Social Psychology, 83*(3), 693-710.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141–151.

Kállai, J., Rózsa, S., Hupuczi, E., Hargitai, R., Birkás, B., Hartung, I., ... & Simon, M. (2019). Cognitive fusion and affective isolation: Blurred self-concept and empathy deficits in schizotypy. *Psychiatry Research, 271*, 178-186.

Kant, I. (1785). Groundwork of the Metaphysics of Morals.

Keefer, K. V. (2015). Self-report assessments of emotional competencies: A critical look at methods and meanings. *Journal of Psychoeducational Assessment, 33*(1), 3-23.

Keltner, D., & Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition & Emotion, 13*(5), 505-521.

Keltner, D., & Kring, A. M. (1998). Emotion, social function, and psychopathology. *Review of General Psychology*, *2*(3), 320-342.

Kim, S. J., & Kou, X. (2014). Not all empathy is equal: How dispositional empathy affects charitable giving. *Journal of Nonprofit & Public Sector Marketing*, *26*(4), 312-334.

Knight, L. K., Stoica, T., Fogleman, N. D., & Depue, B. E. (2019). Convergent Neural Correlates of Empathy and Anxiety During Socioemotional Processing. *Frontiers in Human Neuroscience*, *13*, 94.

Kohlberg, L. (1976) Moral stages and moralization. In Lickona, T (ed.), *Moral Development and Behavior: Theory, research and social issues*. New York: Holt, Rinehart and Winston.

Konstabel, K., Aavik, T., & Allik, J. (2006). Social desirability and consensual validity of personality traits. *European Journal of Personality*, *20*(7), 549-566.

Kring, A. M., & Neale, J. M. (1996). Do schizophrenic patients show a disjunctive relationship among expressive, experiential, and psychophysiological components of emotion?. *Journal of Abnormal Psychology*, *105*(2), 249-257.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*, 1121–1134.

Kwapil, T. R., & Barrantes-Vidal, N. (2014). Schizotypy: looking back and moving forward. *Schizophrenia Bulletin*, *41*(supplement 2), S366-S373.

Lamm, C., Rütgen, M., & Wagner, I. C. (2019). Imaging empathy and prosocial emotions. *Neuroscience Letters*, *693*, 49-53.

Lanzoni, S. (2018). *Empathy: A History*. New Haven: Yale University Press.

Lewis, C. I. (1929). *Mind and the World Order*. New York: C. Scribner's Sons.

Lian, B., Huynh, H., McCutcheon, L. E., Aruguete, M. S., & Murtagh, M. (2019). Is gambling addiction related to celebrity addiction*? Journal of Projective Psychology and Mental Health*

Lietz, C. A., Gerdes, K. E., Sun, F., Geiger, J. M., Wagaman, M. A., & Segal, E. A. (2011). The Empathy Assessment Index (EAI): A confirmatory factor analysis of a multidimensional model of empathy. *Journal of the Society for Social Work and Research, 2*(2), 104-124.

Lilienfeld, S.O., & Widows, M. (2005). *Professional Manual for the Psychopathic Personality Inventory-Revised (PPI-R).* Lutz, Florida: Psychological Assessment Resources.

Lindgren, H. C., & Robinson, J. (1953). An evaluation of Dymond's Test of Insight and Empathy. *Journal of Consulting Psychology, 17*(3), 172-176.

Lipps, T. (1903). Einfühlung, innere Nachahmung, und Organempfindungen. *Archiv für die gesammte Psychologie,1*, 185–204.

Litman, L., Robinson, J., & Rosenzweig, C. (2015). The relationship between motivation, monetary compensation, and data quality among US-and India-based workers on Mechanical Turk. *Behavior Research Methods, 47*(2), 519-528.

Lockwood, P. L. (2016). The anatomy of empathy: Vicarious experience and disorders of social cognition. *Behavioural brain research, 311,* 255-266.

Loevinger, J. (1954). The attenuation paradox in test theory. *Psychological Bulletin, 51*(5), 493-504.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3*(3), 635-694.

López, A., Sanderman, R., Ranchor, A. V., & Schroevers, M. J. (2018). Compassion for others and self-compassion: levels, correlates, and relationship with psychological well-being. *Mindfulness*, *9*(1), 325-331.

Lorenzo-Seva, U., & Ferrando, P. J. (2019). A General Approach for Fitting Pure Exploratory Bifactor Models. *Multivariate Behavioral Research*, *54*(1), 15-30.

Lydekker, R. (1897). Species or Subspecies?. *Nature*, *56*, 256-258.

Lykken, D. T. (1957). A study of anxiety in the sociopathic personality. *The Journal of Abnormal and Social Psychology*, *55*(1), 6-10.

Lynam, D. R., Hoyle, R. H., & Newman, J. P. (2006). The perils of partialling: Cautionary tales from aggression and psychopathy. *Assessment*, *13*(3), 328-341.

Lynam, D. R., & Miller, J. D. (2019). The basic trait of Antagonism: An unfortunately underappreciated construct. *Journal of Research in Personality*.

MacIntyre, A. C. (2007). *Edith Stein: a philosophical prologue, 1913-1922*. Lanham, MD: Rowman & Littlefield.

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science*, *1*(4), 173-181.

Main, A., Walle, E. A., Kho, C., & Halpern, J. (2017). The interpersonal functions of empathy: a relational perspective. *Emotion Review*, *9*(4), 358-366.

Manganaro, P. (2017). The Roots of Intersubjectivity–Empathy and Phenomenology according to Edith Stein. In *Empathy* (pp. 271-286). London: Palgrave Macmillan.

Mansolf, M., & Reise, S. P. (2016). Exploratory bifactor analysis: The Schmid-Leiman orthogonalization and Jennrich-Bentler analytic rotations. *Multivariate Behavioral Research*, *51*(5), 698-717.

Markon, K.E. (2019). Bifactor and Hierarchical Models: Specification, inference, and interpretation. *Annual Review of Clinical Psychology*.

Marsh, A. A. (2018). The neuroscience of empathy. *Current Opinion in Behavioral Sciences, 19*, 110-115.

Marshall, S. L., Ciarrochi, J., Parker, P. D., & Sahdra, B. K. (2019). Is Self-Compassion Selfish? The Development of Self-Compassion, Empathy, and Prosocial Behavior in Adolescence. *Journal of Research on Adolescence*.

McCord, W., & McCord, J. (1964). *The Psychopath: An essay on the criminal mind*. Princeton, NJ: Van Nostrand.

McCrae, R. R., & Costa Jr, P. T. (1983). Joint factors in self-reports and ratings: Neuroticism, extraversion and openness to experience. *Personality and Individual Differences, 4*(3), 245-255.

McCrae, R. R., Stone, S. V., Fagan, P. J., & Costa Jr, P. T. (1998). Identifying causes of disagreement between self-reports and spouse ratings of personality. *Journal of Personality, 66*(3), 285-313.

McCredie, M. N., & Morey, L. C. (2018). Who are the Turkers? A characterization of MTurk workers using the Personality Assessment Inventory. *Assessment*, 1073191118760709.

McDaniel, K. (2016). *Edith Stein*. In Schliesser, E. (Ed.), *Ten Neglected Classics of Philosophy*. Oxford University Press.

McDonald, R. P. (1999). *Test theory: A unified approach*. Mahwah, NJ: Erlbaum.

McGrath, R. E. (2005). Conceptual complexity and construct validity. *Journal of Personality Assessment, 85*(2), 112-124.

McGrath, R. E., Mitchell, M., Kim, B. H., & Hough, L. (2010). Evidence for response bias as a source of error variance in applied assessment. *Psychological Bulletin, 136*(3), 450-470.

Meehl, P.E. (1962). Schizotaxia, schizotypy, schizophrenia. *American Psychologist, 17*, 827-838.

Meehl, P. A. (1975). Genes and the unchangeable core. *Voices: The Art and Science of Psychotherapy, 1*, 173–181.

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

Mehrabian, A. (1996). Manual for the balanced emotional empathy scale (BEES). *Available from Albert Mehrabian*.

Melchers, M., Montag, C., Markett, S., & Reuter, M. (2015). Assessment of empathy via self-report and behavioural paradigms: data on convergent and discriminant validity. *Cognitive Neuropsychiatry*, *20*(2), 157-171.

Meneses, R. W., & Larkin, M. (2012). Edith Stein and the contemporary psychological study of empathy. *Journal of Phenomenological Psychology*, *43*(2), 151-184.

Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the Penn State Worry Questionnaire. *Behaviour Research and Therapy, 28*, 487.

Miller, B. K., Smart, D. L., & Rechner, P. L. (2015). Confirmatory factor analysis of the Machiavellian Personality Scale. *Personality and Individual Differences, 82*, 120-124.

Monaghan, C., Bizumic, B., & Sellbom, M. (2016). The role of Machiavellian views and tactics in psychopathology. *Personality and Individual Differences, 94*, 72-81.

Montgomery, C. B., Allison, C., Lai, M. C., Cassidy, S., Langdon, P. E., & Baron-Cohen, S. (2016). Do adults with high functioning autism or Asperger Syndrome differ in empathy and emotion recognition?. *Journal of Autism and Developmental Disorders*, *46*(6), 1931-1940.

Morin, A. J., Arens, A. K., & Marsh, H. W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(1), 116-139.

Moss, A. J., & Litman, L. (2018, Sept. 18). After the bot scare: Understanding what's been happening with data collection on MTurk and how to stop it [blog post]. Retrieved from https://blog.turkprime.com/after-the-bot-scare-understanding-whats-been-happening-with-data-collection-on-mturk-and-how-to-stop-it

Murphy, B. A. (2019a). The Griffith Empathy Measure does not validly distinguish between cognitive and affective empathy in children. *Australian Psychologist*, *54*(3), 159-164.

Murphy, B. A. (2019b). The Griffith empathy measure needs to be revised with attention to item coding direction: A reply to Dadds (2018). *Australian Psychologist*, *54*(3), 167-170.

Murphy, B. A., Costello, T. H., & Lilienfeld, S. O. (2018b). Is empathic contagion helpful or harmful? Overlooked heterogeneity in the Empathy Index. *Psychological Assessment*, *30*(12), 1703-1706.

Murphy, B. A., Costello, T. H., Watts, A. L., Cheong, Y. F., Berg, J. M., & Lilienfeld, S. O. (2018a). Strengths and weaknesses of two empathy measures: A comparison of the measurement precision, construct validity, and incremental validity of two multidimensional indices. *Assessment*, 1073191118777636.

Murphy, B. A., & Lilienfeld, S. O. (2019). Are self-report cognitive empathy ratings valid proxies for cognitive empathy ability? Negligible meta-analytic relations with behavioral task performance. *Psychological Assessment*.

Murphy B. A., Lilienfeld S. O., Watts A. L. (2017). "Psychopathy and heroism: unresolved questions and future directions," in *Handbook of Heroism and Heroic Leadership* eds Allison S. T., Goethals G. R., Kramer R. M., editors. (New York, NY: Routledge;) 525–546.

Murstein, B. I. (1957). Some comments on the measurement of projection and empathy. *Journal of Consulting Psychology*, *21*(1), 81-82.

Nagel, T. (1974). What is it like to be a bat?. *The Philosophical Review*, *83*(4), 435-450.

Neff, K. D. (2003). The development and validation of a scale to measure self-compassion. *Self and Identity*, *2*(3), 223-250.

Neff, K. D., & Beretvas, S. N. (2013). The role of self-compassion in romantic relationships. *Self and Identity*, *12*(1), 78-98.

Neff, K. D., & Pommier, E. (2013). The relationship between self-compassion and other-focused concern among college undergraduates, community adults, and practicing meditators. *Self and Identity*, *12*(2), 160-176.

Neff, K. D., Whittaker, T. A., & Karl, A. (2017). Examining the factor structure of the Self-Compassion Scale in four distinct populations: Is the use of a total scale score justified?. *Journal of Personality Assessment*, *99*(6), 596-607.

Neumann, D. L., Chan, R. C., Boyle, G. J., Wang, Y., & Westbury, H. R. (2015). Measures of empathy: Self-report, behavioral, and neuroscientific approaches. *Measures of Personality and Social Psychological Constructs*, 257-289.

Nietzsche, F. (1881/1997) *Daybreak: Thoughts on the Prejudices of Morality*, ed. Maudemarie Clark and Brian Leiter, trans. R. J. Hollingdale. Cambridge: Cambridge UP.

Nietzsche, F. (1891/1978). *Thus spoke Zarathustra (*W. Kaufmann, Trans.*)*, New York, NY: Penguin.

Nowicki, S., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The Diagnostic Analysis of Nonverbal Accuracy Scale. *Journal of Nonverbal Behavior*, *18*(1), 9-35.

Obama, B. (2006, December 4). Speech delivered at the Kids In Distressed Situations (K.I.D.S.)/Fashion Delivers gala. Retrieved from https://www.youtube.com/watch?v=4md_A059JRc

Olderbak, S., Sassenrath, C., Keller, J., & Wilhelm, O. (2014). An emotion-differentiated perspective on empathy with the emotion specific empathy questionnaire. *Frontiers in Psychology*, *5*, 653.

Oliveira-Silva, P., & Gonçalves, O. F. (2011). Responding empathically: A question of heart, not a question of skin. *Applied Psychophysiology and Biofeedback*, *36*, 201–207.

O'Neill, G. (1914). Empathy and the Beautiful. *Studies: An Irish Quarterly Review*, 268-279.

Passarelli, M., Masini, M., Bracco, F., Petrosino, M., & Chiorri, C. (2018). Development and validation of the Facial Expression Recognition Test (FERT). *Psychological Assessment*, *30*(11), 1479-1490.

Patrick C.J. (2010). *Triarchic Psychopathy Measure (TriPM)* (https://www.phenxtoolkit.org/index.php?pageLink=browse.protocoldetails&id=121601). Accessed 9 April 2015.

Paulhus, D. L., Robins, R. W., Trzesniewski, K. H., & Tracy, J. L. (2004). Two replicable suppressor situations in personality research. *Multivariate Behavioral Research, 39*(2), 303-328.

Paulhus, D. L., & Vazire, S. (2007). The self-report method. *Handbook of Research Methods in Personality Psychology, 1*, 224-239.

Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality, 36*(6), 556-563.

Peirce, C. S. (1868). Questions concerning certain faculties claimed for man. *Journal of speculative philosophy, 2*, 103-114.

Peirce, C. S. (1878). How to make our ideas clear. *Popular Science Monthly*. 286–302.

Persson, B. N., Kajonius, P. J., & Garcia, D. (2019). Revisiting the structure of the Short Dark Triad. *Assessment, 26*(1), 3-16.

Poincare, H. (1908/1952). *Science and Method*. Dover Publications.

Preacher, K. J., & Hayes, A. F. (2008). Contemporary approaches to assessing mediation in communication research. In A. F.Hayes, M. D.Slater, & L. B.Snyder (Eds.), *The Sage sourcebook of advanced data analysis methods for communication research* (pp. 13–54). Thousand Oaks, CA: Sage.

Preckel, K., Kanske, P., & Singer, T. (2018). On the interaction of social affect and cognition: empathy, compassion and theory of mind. *Current Opinion in Behavioral Sciences, 19*, 1-6.

Preston, S. D. (2013). The origins of altruism in offspring care. *Psychological Bulletin, 139*(6), 1305-1341.

Preston, S. D., & De Waal, F. B. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, *25*(1), 1-20.

Prinz, J. (2011). Against empathy. *The Southern Journal of Philosophy*, *49*, 214-233.

Pryor, R. (1983). *Richard Pryor--here and now*.

Raine, A. (2006). Schizotypal personality: neurodevelopmental and psychosocial trajectories. *Annual Review of Clinical Psychology*, *2*, 291-326.

Raine, A., & Benishay, D. (1995). The SPQ-B: a brief screening instrument for schizotypal personality disorder. *Journal of Personality Disorders*, *9*(4), 346-355.

Ravenscroft, I. (1998). What is it like to be someone else? Simulation and empathy. *Ratio*, *11*(2), 170-185.

Ray, J. V., Frick, P. J., Thornton, L. C., Steinberg, L., & Cauffman, E. (2016). Positive and negative item wording and its influence on the assessment of callous-unemotional traits. *Psychological Assessment*, *28*(4), 394-404.

Realo, A., Allik, J., Nõlvak, A., Valk, R., Ruus, T., Schmidt, M., & Eilola, T. (2003). Mind-reading ability: Beliefs and performance. *Journal of Research in Personality*, *37*(5), 420-445.

Regan, D. T., & Totten, J. (1975). Empathy and attribution: Turning observers into actors. *Journal of Personality and Social Psychology*, *32*(5), 850-856.

Reginster, B. (2015). Sympathy in Schopenhauer and Nietzsche. In E.Schliesser (Ed.), *Sympathy: A history* (pp. 254–285). New York: Oxford University Press.

Reik, T. (1936). *Surprise and the Psychoanalyst*. London: Kegan Paul.

Reise, S.P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavior Research*, 47, 667-696.

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment, 95*(2), 129-140.

Reise, S. P., Kim, D. S., Mansolf, M., & Widaman, K. F. (2016). Is the bifactor model a better model or is it just better at modeling implausible responses? Application of iteratively reweighted least squares to the Rosenberg Self-Esteem Scale. *Multivariate Behavioral Research, 51*, 818-838.

Reise, S., Moore, T., & Maydeu-Olivares, A. (2011). Target rotations and assessing the impact of model violations on the parameters of unidimensional item response theory models. *Educational and Psychological Measurement, 71*(4), 684-711.

Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement, 73*(1), 5-26.

Reise, S. P., & Waller, N. G. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology, 5*, 27-48.

Reniers, R. L., Corcoran, R., Drake, R., Shryane, N. M., & Völlm, B. A. (2011). The QCAE: A questionnaire of cognitive and affective empathy. *Journal of Personality Assessment, 93*(1), 84-95.

Revelle, W. (1978). ICLUST: A cluster analytic approach to exploratory and confirmatory scale construction. *Behavior Research Methods*, *10*(5), 739-742.

Revelle, W. (1979). Hierarchical cluster analysis and the internal structure of tests. *Multivariate Behavioral Research, 14*(1), 57-74.

Revelle, W. (2016). *PSYCH: Procedures for personality and psychological research.* (R package version 1.6.6). Retrieved from http://cran.r-project.org/package_psych

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, *74*(1), 145.

Ripoll, L. H., Zaki, J., Perez-Rodriguez, M. M., Snyder, R., Strike, K. S., Boussi, A., ... & New, A. S. (2013). Empathic accuracy and cognition in schizotypal personality disorder. *Psychiatry research*, *210*(1), 232-241.

Ripoll, L. H., Snyder, R., Steele, H., & Siever, L. J. (2013). The neurobiology of empathy in borderline personality disorder. *Current Psychiatry Reports*, *15*(3), 344.

Rochat, P. (2002). Various kinds of empathy as revealed by the developing child, not the monkey's brain. *Behavioral and Brain Sciences*, *25*(1), 45-46.

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016a). Evaluating bifactor models: calculating and interpreting statistical indices. *Psychological Methods*, *21*(2), 137-150.

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016b). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, *98*(3), 223-237.

Rogers, C. R. (1949). The attitude and orientation of the counselor in client-centered therapy. *Journal of Consulting Psychology*, *13*(2), 82.

Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology*, *21*(2), 95-103.

Rogers, C. R. (1959). *Client-centered therapy: Its current practice, implications, and theory*. Boston: Houghton Mifflin.

Rogers, C. R. (1973). My philosophy of interpersonal relationships and how it grew. *Journal of Humanistic Psychology*, *13*(2), 3-15.

Rogers, C. R. (1975). Empathic: An unappreciated way of being. *The Counseling Psychologist*, *5*(2), 2-10.

Rogers, C.R., & Dymond, R. (1954). *Psychotherapy and Personality Change*. Chicago, IL: University of Chicago Press.

Roosevelt, T. (1897). A layman's views on specific nomenclature. *Science, 5*(122), 685-688.

Rosenthal, R., Hall, A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). *Sensitivity to nonverbal communication: The PONS test*. Baltimore, MD: Johns Hopkins University Press.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 10). New York: Academic Press.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36.

Rousseau, J. J. (1762/1921). Emile, or on education (trans: Foxley, B.). London: Dent.

Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychological Assessment*, *24*(2), 282-292.

Ryan, A., Macdonald, A., & Walker, E. (2013). The treatment of adolescents with schizotypal personality disorder and related conditions: A practice-oriented review of the literature. *Clinical Psychology: Science and Practice*, *20*(4), 408-424.

Sallquist, J., Eisenberg, N., Spinrad, T. L., Eggum, N. D., & Gaertner, B. M. (2009). Assessment of preschoolers' positive empathy: Concurrent and longitudinal relations with positive emotion, social competence, and sympathy. *The Journal of Positive Psychology, 4*(3), 223-233.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph Supplement, 17*(4, Pt. 2).

Samejima, F. (1997). Graded response model. In *Handbook of modern item response theory* (pp. 85-100). New York: Springer.

Sarbin, T. R. (1944). The logic of prediction in psychology. *Psychological Review, 51*(4), 210-228.

Sayre-McCord, G. (2015). Hume and Smith on sympathy, approbation, and moral judgment. In E. Schliesser (Ed.), *Sympathy: A history* (208-246). Oxford UP.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience 1. *Behavioral and Brain Sciences, 36*(4), 393-414.

Schlegel, K., Boone, R. T., & Hall, J. A. (2017). Individual differences in interpersonal accuracy: A multi-level meta-analysis to assess whether judging other people is one skill or many. *Journal of Nonverbal Behavior, 41,* 103–137.

Schmid, J., Leiman, J.M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22,* 53-61.

Schopenhauer, A. (1840/1903) *The Basis of Morality,* translated by A. Broderick Bullocks. London: Swan Sonneschein.

Schutte, N. S., & Stilinović, E. J. (2017). Facilitating empathy through virtual reality. *Motivation and Emotion, 41*(6), 708-712.

Sellbom, M., & Tellegen, A. (2019). Factor analysis in psychological assessment research: Common pitfalls and recommendations. *Psychological Assessment*.

Shamay-Tsoory, S., & Lamm, C. (2018). The neuroscience of empathy–from past to present and future. *Neuropsychologia*, *116*, 1-4.

Shamay-Tsoory, S. G., Shur, S., Harari, H., & Levkovitz, Y. (2007). Neurocognitive basis of impaired empathy in schizophrenia. *Neuropsychology*, *21*(4), 431.

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.

Shaw, L. L., Batson, C. D., & Todd, R. M. (1994). Empathy avoidance: Forestalling feeling for another in order to escape the motivational consequences. *Journal of Personality and Social Psychology*, *67*(5), 879-887.

Skinner, B. F. (1990). Can psychology be a science of mind?. *American Psychologist*, *45*(11), 1206-1210.

Sleep, C. E., Weiss, B., Lynam, D. R., & Miller, J. D. (2019). An examination of the Triarchic Model of psychopathy's nomological network: A meta-analytic review. *Clinical Psychology Review*.

Singer, T., & Klimecki, O. M. (2014). Empathy and compassion. *Current Biology*, *24*(18), R875-R878.

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.

Singer, T., & Steinbeis, N. (2009). Differential roles of fairness-and compassion-based motivations for cooperation, defection, and punishment. *Annals of the New York Academy of Sciences*, *1167*(1), 41-50.

Siu, A. M., & Shek, D. T. (2005). Validation of the Interpersonal Reactivity Index in a Chinese context. *Research on Social Work Practice*, *15*, 118–126.

Sleep, C. E., Lynam, D. R., Hyatt, C. S., & Miller, J. D. (2017). Perils of partialing redux: The case of the Dark Triad. *Journal of Abnormal Psychology*, *126*(7), 939-950.

Smith, A. (2009). *The Theory of Moral Sentiments*, ed. by Ryan Patrick Hanley. New York: Penguin.

Smith, G. T., McCarthy, D. M., & Zapolski, T. C. B. (2009). On the value of homogeneous constructs for construct validation, theory testing, and the description of psychopathology. *Psychological Assessment, 21*(3), 272-284.

Smith, D. W., & McIntyre, R. (1971). Intentionality via intensions. *The Journal of Philosophy*, *68*(18), 541-561.

Southard, E. E. (1918). The empathic index in the diagnosis of mental diseases. *The Journal of Abnormal Psychology*, *13*(4), 199-214.

Southard, E. E., Jarrett, M. C., & Pound, R. (1922). *The Kingdom of Evils: Psychiatric Social Work Presented in One Hundred Case Histories, Together with a Classification of Social Divisions of Evil*. New York: Macmillan.

Spearman, C.E. (1904). “General intelligence,” objectively determined and measured. *American Journal of Psychology,* 15, 201–293

Spreng, R. N., McKinnon, M. C., Mar, R. A., & Levine, B. (2009). The Toronto Empathy Questionnaire: Scale development and initial validation of a factor-analytic solution to multiple empathy measures. *Journal of Personality Assessment, 91*(1), 62.

Staub, E. (1974). Helping a person in distress: Social, personality, and stimulus determinants. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 7, pp. 293–341). New York: Academic Press.

Stead, R., Fekken, G. C., Kay, A., & McDermott, K. (2012). Conceptualizing the Dark Triad of personality: Links to social symptomatology. *Personality and Individual Differences*, *53*(8), 1023-1028.

Stein, E. (1917/ 1989). *On the problem of empathy*, 3rd ed, trad. Waltraut Stein, Washington: ICS Publications.

Steinberg, J. (2014). An epistemic case for empathy. *Pacific Philosophical Quarterly*, *95*(1), 47-71.

Stern, J. A., & Cassidy, J. (2018). Empathy from infancy to adolescence: An attachment perspective on the development of individual differences. *Developmental Review*, *47*, 1-22.

Stinson, L., & Ickes, W. (1992). Empathic accuracy in the interactions of male friends versus male strangers. *Journal of Personality and Social Psychology*, *62*(5), 787–797.

Stone, A. A., Walentynowicz, M., Schneider, S., Junghaenel, D. U., & Wen, C. K. (2019). MTurk participants have substantially lower evaluative subjective well-being than other survey participants. *Computers in Human Behavior*, *94*, 1-8.

Strayer, J. (1987). Affective and cognitive perspectives on empathy. In N. Eisenberg, & J. Strayer (Eds.), *Empathy and its Development* (pp. 218–245). Cambridge: Cambridge University Press.

Stucky, B. D., & Edelen, M. O. (2014). Using hierarchical IRT models to create unidimensional measures from multidimensional data. In S. P. Reise & D. A. Revicki (Eds.), *Handbook of Item Response Theory Modeling: Applications to typical performance assessment* (pp.183–206). New York: Routledge.

Sunderland, M., Afzali, M. H., Batterham, P. J., Calear, A. L., Carragher, N., Hobbs, M., ... & Slade, T. (2019). Comparing Scores From Full Length, Short Form, and Adaptive Tests of the Social Interaction Anxiety and Social Phobia Scales. *Assessment*, 1073191119832657.

Swami, V., Chamorro-Premuzic, T., Sinniah, D., Maniam, T., Kannan, K., Stanistreet, D., & Furnham, A. (2007). General health mediates the relationship between loneliness, life satisfaction and depression. *Social Psychiatry and Psychiatric Epidemiology*, *42*(2), 161-166.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using Multivariate Statistics (Vol. 5)*. Nedham Heights, MA: Allyn & Bacon.

Telle, N. T., & Pfister, H. R. (2016). Positive empathy and prosocial behavior: A neglected link. *Emotion Review*, *8*(2), 154-163.

Tellegen, A., & Atkinson, G. (1974). Openness to absorbing and self-altering experiences (" absorption"), a trait related to hypnotic susceptibility. *Journal of Abnormal Psychology*, *83*(3), 268.

Tellegen, A., & Waller, N. G. (2008). Exploring personality through test construction: Development of the Multidimensional Personality Questionnaire. *The SAGE Handbook of Personality Theory and Assessment, 2*, 261-292.

Thakkar, K. N., & Park, S. (2010). Empathy, schizotypy, and visuospatial transformations. *Cognitive Neuropsychiatry, 15*(5), 477-500.

Thomas, K. A., & Clifford, S. (2017). Validity and Mechanical Turk: An assessment of exclusion methods and interactive experiments. *Computers in Human Behavior, 77*, 184-197.

Thomas, M. L. (2019). Advances in applications of item response theory to clinical assessment. *Psychological Assessment*.

Thurstone, L. L. (1935) Vectors of the Mind: Multiple-factor analysis for the isolation of primary traits. Chicago: University of Chicago Press.

Thurstone, L. L. (1944). Second-order factors. *Psychometrika, 9*(2), 71-100.

Timmers, I., Park, A. L., Fischer, M. D., Kronman, C. A., Heathcote, L. C., Hernandez, J. M., & Simons, L. E. (2018). Is empathy for pain unique in its neural correlates? A meta-analysis of neuroimaging studies of empathy. *Frontiers in Behavioral Neuroscience, 12*.

Titchener, E. B. (1896). *An Outline of Psychology*. New York: Macmillan.

Titchener, E. B. (1909). *Lectures on the experimental psychology of the thought processes*. New York, NY: Macmillan

Titchener, Edward B. 1915. *A Beginner's Psychology*. New York: Macmillan

Thoma, P., Zalewski, I., von Reventlow, H. G., Norra, C., Juckel, G., & Daum, I. (2011). Cognitive and affective empathy in depression linked to executive control. *Psychiatry Research, 189*(3), 373-378.

Thomas, M. L. (2011). The value of item response theory in clinical assessment: a review. *Assessment, 18*(3), 291-307.

Truax, C. (1967). A scale for the rating of accurate **empathy**. In C.Rogers, E.Gendlin, D.Kiesler, & C.Truax (Eds.), *The Therapeutic Relationship and its Impact* (pp. 555–568). Madison, WI: University of Wisconsin Press.

Vachon, D. D., & Lynam, D. R. (2016). Fixing the problem with empathy: Development and validation of the affective and cognitive measure of empathy. *Assessment*, *23*(2), 135-149.

Vachon, D. D., Lynam, D. R., & Johnson, J. A. (2014). The (non) relation between empathy and aggression: Surprising results from a meta-analysis. *Psychological Bulletin*, *140*(3), 751.

Vazire, S. (2010). Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, *98*(2), 281-300.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*(3), 321-327.

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In *Problems and Solutions in Human Assessment* (pp. 41-71). Boston, MA: Springer

Verhaert, G. A., & Van den Poel, D. (2011). Empathy as added value in predicting donation behavior. *Journal of Business Research*, *64*(12), 1288-1295.

Vernon, P. A., Villani, V. C., Vickers, L. C., & Harris, J. A. (2008). A behavioral genetic investigation of the Dark Triad and the Big 5. *Personality and individual Differences*, *44*(2), 445-452.

Verona, E., & Miller, G. A. (2015). Analysis of covariance. In R. L. Cautin & S. O. Lilienfeld (Eds.), *The Encyclopedia of Clinical Psychology*. Chichester, West Sussex Malden, MA: John Wiley and Sons, Inc.

Vilagut, G., Forero, C. G., Castro-rodriguez, J. I., Olariu, E., Barbaglia, G., Astals, M., ... & Sanchez-gil, C. (2019). Measurement equivalence of Promis Depression in Spain and the United States. *Psychological Assessment*, *31*(2), 248-264.

Vischer, R. (1873/1994). On the optical sense of form (H. F.Mallgrave & E.Ikonomou, Trans.). In H. F.Mallgrave & E.Ikonomou (Eds.), *Empathy, form and space: Problems in German aesthetics* (pp. 89–123). Santa Monica, CA: Getty Center for the History of Art and the Humanities.

Vossen, H. G., Piotrowski, J. T., & Valkenburg, P. M. (2015). Development of the adolescent measure of empathy and sympathy (AMES). *Personality and Individual Differences, 74*, 66-71.

Waller, N. G. (2018). Direct Schmid–Leiman Transformations and Rank-Deficient Loadings Matrices. *Psychometrika*, *83*(4), 858-870.

Waller, N.G. (2019). *fungible: Psychometric Functions from the Waller Lab* (R package version 1.86).

Wang, S., Lilienfeld, S. O., & Rochat, P. (2019). Schadenfreude deconstructed and reconstructed: A tripartite motivational model. *New Ideas in Psychology*, *52*, 1-11.

Wang, Y., Neumann, D. L., Shum, D. H., Liu, W. H., Shi, H. S., Yan, C., ... & Chan, R. C. (2013). Cognitive empathy partially mediates the association between negative schizotypy traits and social functioning. *Psychiatry research*, *210*(1), 62-68.

Ward, M. K., & Meade, A. W. (2018). Applying social psychology to prevent careless responding during online surveys. *Applied Psychology*, *67*(2), 231-263.

Watson, D., Clark, L. A., Chmielewski, M., & Kotov, R. (2013). The value of suppressor effects in explicating the construct validity of symptom measures. *Psychological Assessment*, *25*(3), 929-941.

Watson, P. J., Grisham, S. O., Trotter, M. V., & Biderman, M. D. (1984). Narcissism and empathy: Validity evidence for the Narcissistic Personality Inventory. *Journal of Personality Assessment*, *48*(3), 301-305.

Watts, A. L., Lilienfeld, S. O., Edens, J. F., Douglas, K. S., Skeem, J. L., Verschuere, B., & LoPilato, A. C. (2016). Does response distortion statistically affect the relations between self-report psychopathy measures and external criteria?. *Psychological Assessment*, *28*(3), 294.

Weeks, J. W., Heimberg, R. G., Fresco, D. M., Hart, T. A., Turk, C. L., Schneier, F. R., & Liebowitz, M. R. (2005). Empirical validation and psychometric evaluation of the Brief Fear of Negative Evaluation Scale in patients with social anxiety disorder. *Psychological Assessment*, *17*(2), 179-190.

Weisz, E., & Zaki, J. (2018). Motivated empathy: A social neuroscience perspective. *Current Opinion in Psychology*, *24*, 67-71.

Welp, L. R., & Brown, C. M. (2014). Self-compassion, empathy, and helping intentions. *The Journal of Positive Psychology*, *9*(1), 54-65.

White, B. A. (2014). Who cares when nobody is watching? Psychopathic traits and empathy in prosocial behaviors. *Personality and Individual Differences*, *56*, 116-121.

Wickramasekera, I. E., & Szlyk, J. P. (2003). Could empathy be a predictor of hypnotic ability?. *International Journal of Clinical and Experimental Hypnosis*, *51*(4), 390-399.

Wispé, L. (1986). The distinction between sympathy and empathy: To call forth a concept, a word is needed. *Journal of Personality and Social Psychology*, *50*(2), 314-321.

Wispé, L. (1987) History of the concept of empathy. In: *Empathy and its Development*, ed. N. Eisenberg & J. Strayer. Cambridge University Press. 17-37.

Wittgenstein, Ludwig. (1953, 1968). *Philosophical Investigations*, G.E.M. Anscombe, trans. New York: Macmillan.

Wondra, J. D., & Ellsworth, P. C. (2015). An appraisal theory of empathy and other vicarious emotional experiences. *Psychological Review*, *122*(3), 411.

Wood, D., Harms, P. D., Lowman, G. H., & DeSimone, J. A. (2017). Response speed and response consistency as mutually validating indicators of data quality in online samples. *Social Psychological and Personality Science*, *8*(4), 454-464.

Woods, C. M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment*, *28*, 189-194.

Zahavi, D., & Rochat, P. (2015). Empathy≠ sharing: Perspectives from phenomenology and developmental psychology. *Consciousness and Cognition*, *36*, 543-553.

Zaki, J. (2014). Empathy: a motivated account. *Psychological Bulletin*, *140*(6), 1608.

Zaki, J. (2017). Moving beyond stereotypes of empathy. *Trends in Cognitive Sciences*, *21*(2), 59-60.

Zell, E., & Krizan, Z. (2014). Do people have insight into their abilities? A metasynthesis. *Perspectives on Psychological Science*, *9*(2), 111-125.

Zuckerman, M., & Larrance, D. (1979). Individual differences in perceived encoding and decoding abilities. In R. Rosenthal (Ed.), *Skill in Nonverbal Communication* (pp. 171–203). Cambridge, MA: Oelegeschlager, Gunn, & Hain.

Tables and Figures

(in order of mention within the text)

Table 1. Five empathy dimensions revealed by Reniers et al. (2011)

Construct	*Example item*
Metacognition of Mindreading Ability	*"I can easily work out what another person might want to talk about"*
Perspective-Taking Reasonableness	*"I try to look at everybody's side of a disagreement before I make a decision"*
Empathic Concern	*"It makes me feel good to help someone in need"*
Empathic Contagion	*"If I see someone fidgeting, I'll start feeling anxious too"*
Fantasy	*"Becoming extremely involved in a good book or movie is somewhat rare for me"*

Table 2. Hypothesized new or revised empathic tendencies dimensions	
Construct	*Example item*
Perspective-Taking Reasonableness	*"I make a big effort to sympathetically understand people who disagree with me"*
Concern for Happiness of Others	*"I go out of my way to try to bring joy to others"*
Concern for Suffering of Others	*"When I see someone suffering, I feel a strong urge to help comfort that person"*
Empathic Restraint	*"Even when I am upset, I never lash out to cause pain to someone"*
Confidante Listening	*"People see me as a very empathic, attentive listener"*
Empathic Approach-Avoidance	*"When someone starts to talk to me about his or her emotional struggles, I usually change the subject to something less serious"*
Aversive Empathic Contagion	*"If I am around someone who is feeling nervous, I get really nervous"*
Enjoyable Empathic Contagion	*"When I am with someone who is really happy, it makes me joyful too"*
Empathic Absorption	*"I tend to get emotionally involved with the stories people tell me about their life experiences"*
Empathic Distress	*"When people talk to me about something terrible that has happened to them, I get really nervous"*
Empathy Valuing	*"When choosing friends, I care much more about a person's compassionate nature than about his or her successfulness"*
Note: I expect that some of these different measurement domains will collapse into one another factor-analytically, but I do not have strong specific predictions in this regard.	

Table 3. Non-empathy measures administered and analyzed, studies 1 and 2

Study 1	
HEXACO-60	
Honesty-Humility (H)	Agreeableness (A)
Non-Sentimental Emotionality (NSE)	Conscientiousness (C)
Extraversion (X)	Openness (O)
Study 2	
Buss-Perry Aggression Questionnaire (BPAQ)	
Anger	Physical Aggression (Physical)
Hostility	Verbal Aggression (Verbal)
Triarchic Psychopathy Measure (TriPM)	
Meanness	Disinhibition
Machiavellian Personality Scales (MPS)	
Amorality (AMO)	Desire for Status (DS)
Distrust of Others (DO)	Desire for Control (DC)
Schizotypal Personality Questionnaire Brief Revised (SPQ-BR)	
Social Anxiety (SA)	Odd Speech (OS)
Unusual Perceptions (UP)	Suspiciousness/Ideas of Reference (SUIR)
Eccentric Behavior (EB)	Magical Thinking (MT)
Lack Close Friendships/Constricted Affect (CFCA)	
Penn State Worry Questionnaire (PSWQ)	
Satisfaction with Life Scale (SLS)	
Forms of Self-Criticising and Reassuring (FSCRS; re-labeled as “Self-Compassion”)	

Figure 1. ICLUST diagram of 34 empathy-related negative emotionality items, study 1

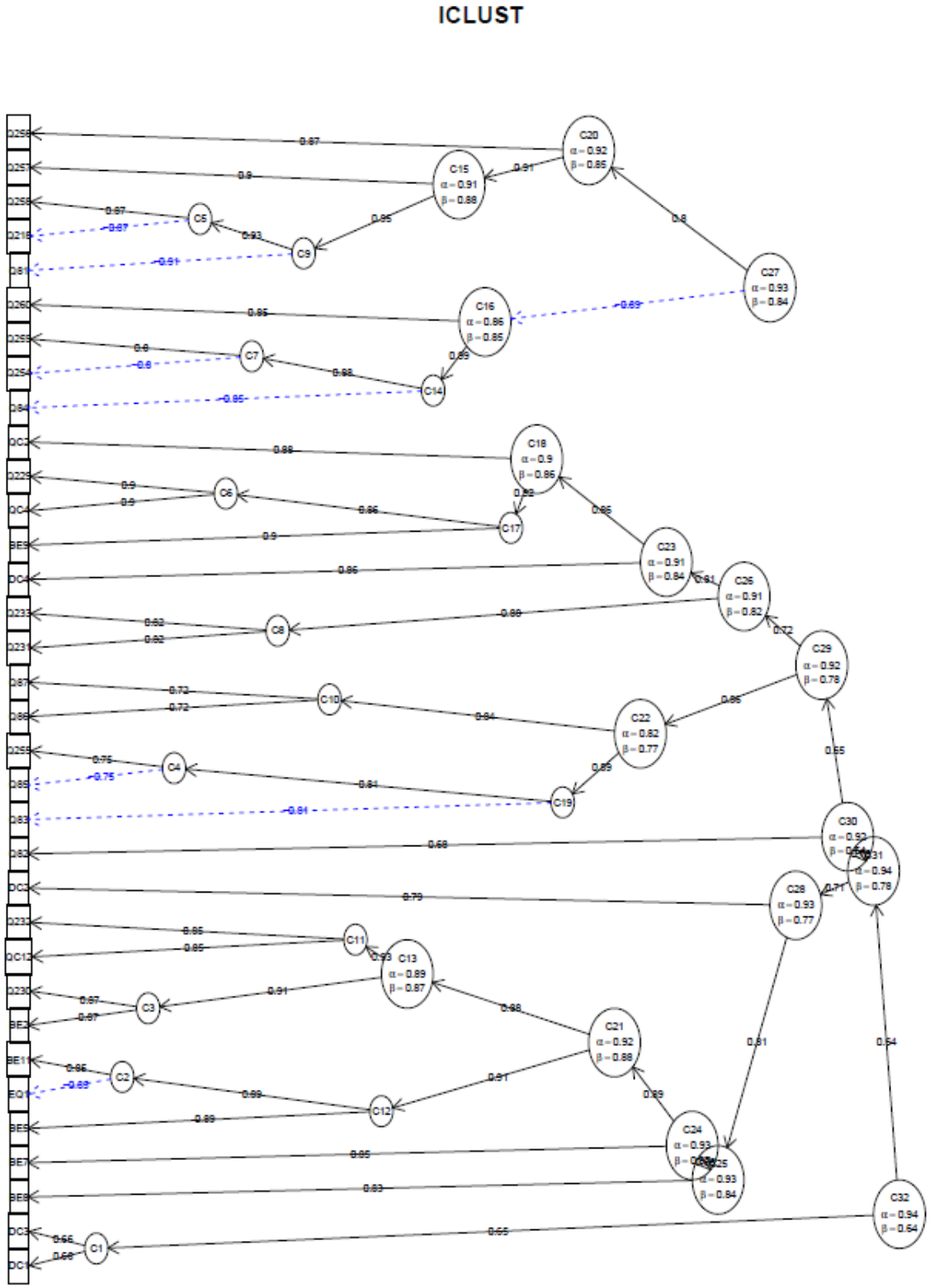


Figure 2. Scree plot and parallel analysis, 137 empathic caring items, study 1

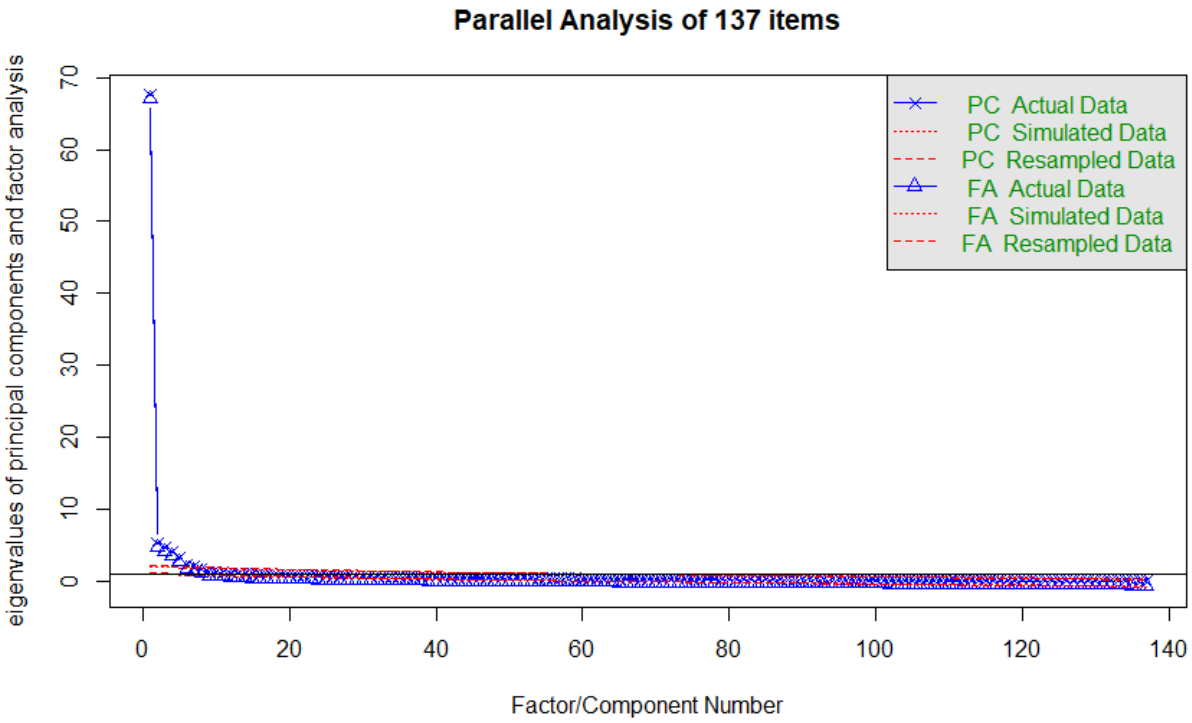


Figure 3. Schmid-Leiman transformation, 35 items of 6 preliminary specific content empathic caring scales, study 1

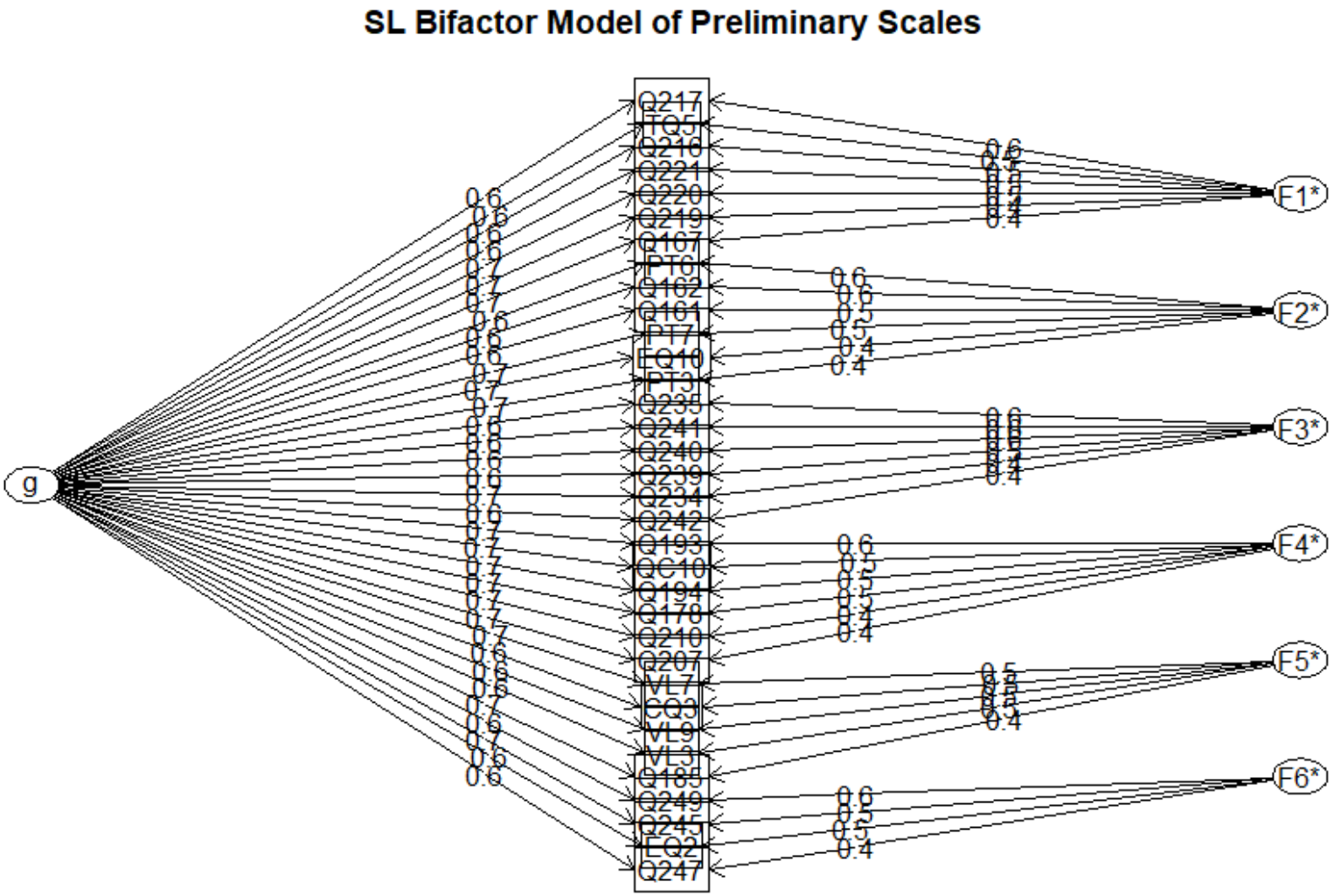


Table 4. Study 1 bifactor indices, preliminary scales, SL EBFA

	Total omega	Omega H or Omega HS	FD	H
General Factor	.98	.84	.95	.96
Empathic Avoidance Factor	.93	.34	.85	.70
Perspective-Taking Factor	.93	.36	.87	.71
Appetitive Contagion Factor	.90	.37	.85	.69
Confidante Listening Factor	.93	.32	.85	.67
Empathic Restraint Factor	.89	.33	.82	.61
Empathic Absorption Factor	.88	.36	.84	.61

Note. SL EBFA = Schmid-Leiman bifactor transformation; FD = factor determinacy (the correlation between factor scores and factors); H = correlation between a factor and an optimally-weighted composite

Table 5. Empathy scales inter-correlations, study 1 (n = 620-712)

(standardized residuals controlling for GFEC total, displayed below diagonal)

	General Factor, 137-item pool	GFEC total	TEQ total	PT	CL	EAV	AC	ER	EAB	EID	ANXCON	SADCON
General Factor, 137-item pool	-	.94*	.94*	.73*	.78*	-.81*	.72*	.80*	.74*	-.58*	.21*	.54*
GFEC total		-	.89*	.67*	.78*	-.74*	.69*	.71*	.71*	-.54*	.18*	.50*
TEQ total			-	.64*	.71*	-.81*	.73*	.76*	.70*	-.57*	.22*	.58*
PT				-	.60*	-.57*	.55*	.56*	.52*	-.45*	.07	.34*
CL				.16*	-	-.65*	.51*	.57*	.53*	-.58*	.01	.29*
EAV				-.14*	-.18*	-	-.51*	-.62*	-.56*	.76*	-.05	-.36*
AC				.17*	-.06	-.00	-	.55*	.62*	-.34*	.25*	.56*
ER				.16*	.04	-.19*	.12*	-	.50*	-.37*	.19*	.46*
EAB				.09	-.05	-.08	.27*	-.00	-	-.33*	.40*	.67*
EID				-.15*	-.30*	.65*	.04	.01	.09	-	.25*	-.04
ANXCON				-.07	-.21*	.12*	.18*	.09	.39*	.41*	-	.64*
SADCON				.01	-.18*	.02	.34*	.18*	.52*	.30*	.65*	-

Note. Below the diagonal, correlations are between variables which are both standardized residuals of regressing each on the GFEC.

Table 6. Relationships with HEXACO personality dimensions, study 1 (n = 620-712)

	H	NSE	X	A	C	O
General Factor, (137 item)	.42*	.24*	.34*	.49*	.37*	.36*
GFEC	.37*	.22*	.40*	.45*	.33*	.32*
TEQ	.41*	.26*	.31*	.43*	.35*	.37*
PT	.31*	.04	.35*	.57*	.29*	.32*
PT residual	.08	-.14*	.11*	.36*	.10*	.14*
CL	.29*	.07	.40*	.45*	.34*	.37*
CL residual	-.00	-.16*	.14*	.16*	.14*	.18*
EAV	-.41*	-.16*	-.28*	-.43*	-.32*	-.34*
EAV residual	-.21*	.00	.02	-.13*	-.11*	-.16*
AC	.25*	.17*	.45*	.40*	.26*	.24*
AC residual	-.00	.03	.24*	.11*	.05	.03
ER	.41*	.25*	.16*	.49*	.30*	.27*
ER residual	.22*	.13*	-.18*	.24*	.10	.06
EAB	.17*	.36*	.20*	.25*	.13*	.25*
EAB residual	-.13*	.29*	-.11*	-.10*	-.15*	.03
EID	-.30*	.12*	-.39*	-.38*	-.29*	-.28*
EID residual	-.12*	.28*	-.21*	-.16*	-.13*	-.13*
ANXCON	-.10*	.67*	-.23*	-.13*	-.17*	-.02
ANXCON residual	-.17*	.65*	-.31*	-.21*	-.23*	-.08
SADCON	.10*	.54*	.00	.14*	.03	.13*
SADCON residual	-.09	.49*	-.23*	-.10*	-.15*	-.04

Note. * = $p < .01$; H = Honesty-Humility; NSE = Emotionality minus the Sentimentality items; X = Extraversion; A = Agreeableness; C = Conscientiousness; O = Openness

Table 7. Study 2, bifactor indices, empathic caring scales only, orthogonal bifactor CFA

	Total Omega	Omega H or Omega HS	FD	H
General Factor	.92	.62	.94	.95
Empathic Avoidance Factor	.94	.41	.88	.77
Perspective-Taking Factor	.91	.37	.85	.70
Appetitive Contagion Factor	.89	.42	.86	.71
Confidante Listening Factor	.92	.21	.76	.56
Empathic Restraint Factor	.85	.35	.81	.60

Note. FD = factor determinacy (the correlation between factor scores and factors); H = correlation between a factor and an optimally-weighted composite

Table 8. Study 2, bifactor indices, GFEC plus specific empathic caring scales, oblique bifactor CFA

	Total omega	Omega H or Omega HS	FD	H
General Factor	.96	.87	.98	.98
Empathic Avoidance Factor	.94	.42	.92	.78
Perspective-Taking Factor	.91	.32	.87	.66
Appetitive Contagion Factor	.89	.39	.89	.69
Confidante Listening Factor	.92	.23	.86	.59
Empathic Restraint Factor	.85	.31	.83	.57

Note. FD = factor determinacy (the correlation between factor scores and factors); H = correlation between a factor and an optimally-weighted composite

Table 9. Empathy scales inter-correlations, study 2 (n = 335)

(standardized residuals controlling for GFEC total, displayed below diagonal)

	TEQ	PT	CL	EAV	AC	ER	EID	ANXCON	SADCON
GFEC	.88*	.74*	.80*	-.68*	.66*	.68*	-.58*	.14	.39*
TEQ	-	.66*	.71*	-.76*	.73*	.73*	-.63*	.14	.41*
PT		-	.63*	-.48*	.55*	.52*	-.41*	.05	.25*
CL		.09	-	-.61*	.51*	.56*	-.58*	.01	.23*
EAV		.06	-.14	-	-.52*	-.51*	.86*	.09	-.13
AC		.11	-.05	-.13	-	.47*	-.48*	.13	.33*
ER		.03	.04	-.09	.04	-	-.38*	.18*	.35*
EID		.03	-.24*	.78*	-.16*	.03	-	.26*	.06
ANXCON		-.08	-.17*	.25*	.06	.12	.42*	-	.74*
SADCON		-.07	-.16*	.20*	.10	.12	.38*	.75*	-

Note. * = $p < .01$; Below the diagonal, correlations are between variables which are both standardized residuals of regressing each on the GFEC. GFEC = Generaral Factor of Empathic Caring scale; TEQ = Toronto Empathy Questionnaire; PT = Perspective-Taking; CL = Confidante Listening; EAV = Empathic Avoidance; AC = Appetitive Contagion; ER = Empathic Restraint; EID = Empathic Intimacy Discomfort; ANXCON = Anxiety Contagion; SADCON = Sadness Contagion

Table 10. Bifactor EFA loadings for the 48 empathic caring items, Study 2 (n = 335)

	General	EAV	AC	PT	ER	#5
GF1: Honestly, I tend to supportively nurture people around me who are struggling.	0.85	-	-	-	-	-
GF2: I go out of my way to bring joy to others.	0.78	-	-	-	-	-
GF3: It is very important to me that I help others feel truly understood.	0.84	-	-	-	-	-
GF4: When I see that someone is lonely, I often try to sincerely emotionally connect with the person.	0.81	-	-	-	-	-
GF5: I work to be a healing presence for people who feel down about themselves.	0.85	-	-	-	-	-
GF6: I get a strong urge to help when I see someone who is upset.	0.81	-	-	-	-	0.32
GF7: I really enjoy caring for other people.	0.77	-	-	-	-	-
GF8: I make lots of small efforts to bring happiness to others.	0.74	-	-	-	-	-
GF9: I care a lot about being a very encouraging person to others.	0.86	-	-	-	-	-
GF10: More than anything else, I want to be remembered as a compassionate person.	0.68	-	-	-	-	-
GF11: I try to make sure that the people around me feel good about themselves.	0.84	-	-	-	-	-
GF12: I would describe myself as a pretty soft-hearted person.	0.7	-	-	-	-	-
GF13: I love boosting other people's self esteem.	0.8	-	-	-	-	-
GF14: I often have tender, concerned feelings for people less fortunate than me.	0.76	-	-	-	-	-
GF15: I often try to help people feel better when they are upset.	0.81	-	-	-	-	-
AC1 (R): Even when I am around people who are joyful, it is hard for me to really feel that joy.	0.56	-	0.55	-	-	-
AC2: When I see strangers having a lot of fun, I can't help but feel pleasure too.	0.51	-	0.57	-	-	-

Note. All loadings below .3 not presented. GF = General Factor Empathic Caring scale, AC = Appetitive Contagion scale, CL = Confidante Listening scale, EA = Empathic Avoidance scale (all items reverse-scored), ER = Empathic Restraint scale, PT = Perspective-Taking Motivation scale

Table 10. *Bifactor EFA loadings for the 48 empathic caring items, Study 2 (n = 335), cont.*

	General	EAV	AC	PT	ER	#5
AC3: When someone around me is really excited about his or her accomplishment, I get genuinely excited as well.	0.74	-	0.35	-	-	-
AC4: When strangers around me are joyful, my heart starts to fill with joy too.	0.64	-	0.59	-	-	-
AC6 (R): When other people are cheerful and merry, I do not usually catch their emotions myself.	0.63	-	0.51	-	-	-
AC7: If strangers around me are really happy, I tend to also start feeling really happy	0.59	-	0.56	-	-	-
CL1: People feel quite comfortable opening up to me about their hopes and dreams.	0.78	-	-	-	-	-
CL2: People tell me that I am a comforting person to talk to.	0.85	-	-	-	-	-
CL3: Friends talk to me about their problems as they say that I am very understanding.	0.86	-	-	-	-	-
CL4: I am a calming presence for people when they are suffering.	0.85	-	-	-	-	-
CL5: People feel safe opening up to me because they know I won't be judgmental.	0.78	-	-	-	-	-
CL6: Compared to others, I am a particularly supportive listener.	0.74	-	-	-	-	-
CL7: I am often like a kind therapist with my friends.	0.73	-	-	-	-	-
EA1: If people start to talk about really sad things in their lives, I try to get out of there pretty quickly.	0.64	0.6	-	-	-	-
EA2: When a friend starts to talk about his/her problems, I try to steer the conversation towards something else.	0.68	0.48	-	-	-	-
EA3: When someone starts to talk to me about his or her emotional struggles, I usually change the subject to something less serious.	0.66	0.59	-	-	-	-

Note. All loadings below .3 not presented. GF = General Factor Empathic Caring scale, AC = Appetitive Contagion scale, CL = Confidante Listening scale, EA = Empathic Avoidance scale (all items reverse-scored), ER = Empathic Restraint scale, PT = Perspective-Taking Motivation scale

Table 10. *Bifactor EFA loadings for the 48 empathic caring items, Study 2 (n = 335), cont.*

	General	EAV	AC	PT	ER	#5
EA4: In truth, I am not all that interested in how most other people experience their emotions.	0.73	0.35	-	-	-	-
EA5: I tend to purposefully avoid people who are feeling sad.	0.63	0.51	-	-	-	-
EA6: When people start talking about their own painful feelings, I sometimes only half-listen.	0.66	0.46	-	-	-	-
EA7: To be perfectly honest, I do not actually like it when people start talking a lot about their feelings.	0.69	0.5	-	-	-	-
EA8: I try to avoid having to listen to people talk about their deep emotions.	0.73	0.55	-	-	-	-
ER1: I feel awful when I hurt someone's feelings.	0.67	-	-	-	0.58	-
ER2: I avoid hurting other people's feelings.	0.54	-	-	-	0.56	-
ER3: If I see that I am doing something that hurts someone, I will quickly stop.	0.58	-	-	-	0.55	-
ER4 (R): I don't worry much about hurting people's feelings.	0.62	-	-	-	0.43	-
ER5: When I can tell that what I am saying is beginning to hurt someone, it makes me want to be very gentle towards that person.	0.72	-	-	-	0.36	-
PT1: When I'm upset at someone, I usually try to "put myself in his shoes" for a while.	0.66	-	-	0.63	-	-
PT2: If I start to feel frustrated with someone, I focus on understanding that person's point of view.	0.61	-	-	0.51	-	-
PT3: I find it easy to put myself in somebody else's shoes.	0.72	-	-	0.39	-	-
PT4: Before criticizing somebody, I try to imagine how I would feel if I were in their place.	0.65	-	-	0.5	-	-
PT5: I sometimes try to understand my friends better by imagining how things look from their perspective.	0.75	-	-	0.4	-	-
PT6: Even when I have a conflict with someone, I think deeply about how things must feel from his/her point of view.	0.61	-	-	0.57	-	-
PT7: I make sure to sympathetically imagine the feelings of other people even when they are aggressive towards me.	0.56	-	-	0.43	-	-

Note. All loadings below .3 not presented. GF = General Factor Empathic Caring scale, AC = Appetitive Contagion scale, CL = Confidante Listening scale, EA = Empathic Avoidance scale (all items reverse-scored), ER = Empathic Restraint scale, PT = Perspective-Taking Motivation scale

Table 11. Relationships with TriPM and BPAQ scales, study 2 (n = 335)

	TriPM Meanness	TriPM Disinhibition	BPAQ Anger	BPAQ Hostility	BPAQ Physical	BPAQ Verbal
GFEC	-.71*	-.29*	-.24*	-.24*	-.33*	-.28*
TEQ	-.81*	-.39*	-.29*	-.31*	-.38*	-.32*
PT	-.54*	-.26*	-.28*	-.25*	-.28*	-.19*
PT residual	-.01	-.06	-.15*	-.12	-.06	.02
CL	-.57*	-.30*	-.28*	-.26*	-.25*	-.20*
CL residual	.00	-.12	-.14*	-.12	.02	.03
EAV	.67*	.43*	.33*	.41*	.34*	.27*
EAV residual	.26*	.32*	.22*	.33*	.16*	.11
AC	-.53*	-.23*	-.23*	-.37*	-.33*	-.23*
AC residual	-.09	-.05	-.10	-.28*	-.15*	-.06
ER	-.77*	-.40*	-.33*	-.21*	-.40*	-.43*
ER residual	-.39*	-.27*	-.23*	-.06	-.24*	-.34*
EID	.53*	.44*	.36*	.50*	.27*	.26*
EID residual	.15*	.33*	.26*	.45*	.10	.12
ANXCON	-.04	.31*	.31*	.48*	.03	.05
ANXCON residual	.06	.35*	.34*	.52*	.07	.09
SADCON	-.29*	.15*	.18*	.31*	-.09	-.05
SADCON residual	-.01	.29*	.30*	.44*	.05	.07

Note. * = $p < .01$; TriPM = Triarchic Psychopathy Measure; BPAQ = Buss-Perry Aggression Questionnaire

Table 12. Relationships with Emotional Distress and MPS Scales,sStudy 2 (n = 335)

	Self-Compassion	PSWQ	SLS	MPS DO	MPS DS	MPS DC	MPS AMO
GFEC	.27*	-.03	.32*	-.36*	-.19*	-.03	-.42*
TEQ	.26*	.02	.25*	-.46*	-.24*	-.10	-.55*
PT	.30*	-.13	.31*	-.30*	-.18*	-.05	-.27*
PT residual	.16*	-.17*	.11	-.05	-.06	-.04	.06
CL	.33*	-.12	.29*	-.27*	-.05	.04	-.30*
CL residual	.20*	-.17*	.07	.02	.17*	.11	.06
EAV	-.31*	.10	-.24*	.50*	.23*	.11	.52*
EAV residual	-.18*	.12	-.03	.35*	.14	.12	.31*
AC	.41*	-.14*	.31*	-.41*	-.08	-.01	-.36*
AC residual	.31*	-.16*	.14*	-.23*	.06	.01	-.11
ER	.12	.08	.15*	-.38*	-.21*	-.19*	-.50*
ER residual	-.08	.13	-.09	-.19*	-.10	-.23*	-.29*
EID	-.45*	.26*	-.35*	.51*	.13	.05	.43*
EID residual	-.36*	.30*	-.20*	.37*	.02	.04	.23*
ANXCON	-.47*	.63*	-.25*	.24*	.03	.03	.08
ANXCON residual	-.51*	.64*	-.30*	.29*	.06	.04	.14
SADCON	-.28*	.45*	-.15*	.10	-.09	-.04	-.10
SADCON residual	-.42*	.50*	-.30*	.26*	-.01	-.03	.08

Note. * = $p < .01$; PSWQ = Penn State Worry Questionnaire; SLS = Satisfaction with Life Scale; MPS = Machiavellian Personality Scale; DO = Distrust of Others; DS = Desire for Status; DC = Desire for Control; AMO = Amorality

Table 13. Relationships with schizotypal personality traits (SPQ-BR), study 2 (n = 335)

	SA	CFCA	EB	OS	SUIR	UP	MT
GFEC	-.18*	-.47*	-.17*	-.07	-.18*	-.01	.06
TEQ	-.15*	-.51*	-.18*	-.12	-.28*	-.11	-.03
PT	-.18*	-.39*	-.13	-.15*	-.14*	-.01	-.02
PT residual	-.08	-.07	.00	-.14	-.01	.00	-.10
CL	-.24*	-.51*	-.20*	-.17*	-.20*	.00	.06
CL residual	-.17*	-.23*	-.10	-.19*	-.10	.02	.01
EAV	.28*	.58*	.28*	.26*	.39*	.20*	.10
EAV residual	.21*	.36*	.22*	.29*	.37*	.27*	.19*
AC	-.27*	-.52*	-.21*	-.11	-.28*	-.07	.08
AC residual	-.20*	-.28*	-.13	-.09	-.21*	-.09	.06
ER	.02	-.25*	-.12	-.09	-.19*	-.14	-.13
ER residual	.18*	.09	-.00	-.05	-.09	-.18*	-.23*
EID	.44*	.69*	.35*	.32*	.47*	.23*	.01
EID residual	.41*	.52*	.31*	.35*	.44*	.27*	.06
ANXCON	.55*	.25*	.23*	.43*	.44*	.29*	.14
ANXCON residual	.58*	.32*	.25*	.44*	.46*	.30*	.13
SADCON	.37*	.06	.09	.33*	.27*	.18*	.10
SADCON residual	.48*	.27*	.17*	.39*	.37*	.20*	.08

Note. * = $p < .01$; SPQ-BR = Schizotypal Personality Questionnaire – Brief Revised; SA = Social Anxiety; CFCA = Constricted Affect/Lack Friendships; EB = Eccentric Behavior; OS = Odd Speech; SUIR = Suspiciousness/Ideas of Reference; UP = Unusual Perceptions; MT = Magical Thinking

Table 14. General and specific content scales, strongest external correlates, studies 1 and 2

Zero-Order Associations at or above r = |.5|
GFEC: TriPM Meanness (-.71)
CL: TriPM Meanness (-.57)
PT: HEX Agreeableness (.57), TriPM Meanness (-.54)
EAV: TriPM Meanness (.67), SPQ CFCA (.58), MPS Amorality (.52), MPS Distrust of Others (.50)
AC: TriPM Meanness (-.53), SPQ CFCA (-.52)
ER: TriPM Meanness (-.77), MPS Amorality (-.50)
EID: SPQ CFCA (.69), TriPM Meanness (.53), MPS Distrust of Others (.50)
ANXCON: HEX NSE (.67), PSWQ Worry (.63), SPQ SA (.55)
SADCON: HEX NSE (.54)

Unique Variance Associations at or above r = |.3|
CL: none

PT: HEX Agreeableness (.36)

EAV: (SPQ CFCA (.36), SPQ SUIR (.37), BPAQ Hostility (.33), MPS Distrust of Others (.35), TriPM Disinhibition (.32), MPS Amorality (.31)

AC: Self-Compassion (.31)

ER: TriPM Meanness (-.39), BPAQ Verbal (.34)

EID: SPQ CFCA (.52), BPAQ Hostility (.45), SPQ SUIR (.44), SPQ SA (.41), MPS Distrust of Others (.37), Self-Compassion (-.36), SPQ OS (.35), TriPM Disinhibition (.33), SPQ EB (.31), PSWQ Worry (.30)

ANXCON: HEX NSE (.65), PSWQ Worry (.64), SPQ SA (.58), BPAQ Hostility (.52), Self-Compassion (-.51), SPQ SUIR (.46), SPQ OS (.44), TriPM Disinhbition (.35), BPAQ Anger (.34), SPQ CFCA (.32), HEX Extraversion (-.31), SLS Life Satisfaction (-.30)

SADCON: PSWQ Worry (.50), HEX NSE (.49), SPQ SA (.48), BPAQ Hostility (.44), Self-Compassion (-.42), SPQ OS (.39), SPQ SUIR (.37), BPAQ Anger (.30), SLS Life Satisfaction (-.30)

Note. TriPM = Triarchic Psychopathy Measure; SPQ = Schizotypal Personality Questionaire – Brief Revised; HEX = Hexaco-60; MPS = Machiavellian Personality Scales, PSWQ = Penn State Worry Questionnaire; BPAQ = Buss-Perry Aggression Questionnaire; CFCA = Constricted Affect/Lack Friendships; SA = Social Anxiety; OS = Odd Speech; SUIR = Suspiciousness/Ideas of Reference; EB = Eccentric Behavior; NSE = Emotionality minus Sentimentality items

Appendix A. Empathy items administered in Study 1

IRI items (not revised)

I sometimes find it difficult to see things from the "other guy's" point of view.

I try to look at everybody's side of a disagreement before I make a decision.

I sometimes try to understand my friends better by imagining how things look from their perspective.

If I'm sure I'm right about something, I don't waste much time listening to other people's arguments.

I believe that there are two sides to every question and try to look at them both.

When I'm upset at someone, I usually try to "put myself in his shoes" for a while.

Before criticizing somebody, I try to imagine how I would feel if I were in their place.

I often have tender, concerned feelings for people less fortunate than me.

Sometimes I don't feel very sorry for other people when they are having problems.

When I see someone being taken advantage of, I feel kind of protective towards them.

Other people's misfortunes do not usually disturb me a great deal.

When I see someone being treated unfairly, I sometimes don't feel very much pity for them.

I am often quite touched by things that I see happen.

I would describe myself as a pretty soft-hearted person.

IRI items (revised)

I really get involved with the feelings of people when they are telling me stories about themselves.

I am usually objective when I hear someone's story, and I don't often get completely caught up in it.

Becoming emotionally involved in another person's story is very rare for me.

After hearing a person tell me her story, I have felt as though I were that person

When I hear a good story from someone, I can very easily put myself in the place of the person.

When I am hearing a person's story, I imagine how I would feel if the events in the story were happening to me.

When people share their deep emotions, I feel apprehensive and ill-at-ease.

I sometimes feel helpless when I listen to someone talk about their suffering.

When I see someone terribly upset, I tend to remain calm.

I feel scared when I am in a conversation with someone who is opening up about his/her deepest worries.

I am usually pretty calm when dealing with other people's emotions of distress.

I tend to lose control during really emotional conversations.

When I spend time with someone who is suffering from emotional trauma, I often fall to pieces.

EI items (as identified in Murphy, Costello, & Lilienfeld, 2018)

If I see someone who is excited, I will feel excited myself.

I sometimes find myself feeling the emotions of the people around me, even if I don't try to feel what they're feeling.

If I'm watching a movie and a character injures their leg, I will feel pain in my leg.

If I hear a story in which someone is scared, I will imagine how scared I would be in that situation and begin to feel scared myself.

I can't watch shows in which an animal is being hunted by another because I feel nervous as if I am being hunted.

If I see someone fidgeting, I'll start feeling anxious too.

I catch myself crossing my arms or legs just like the person I'm talking to.

If I see a video of a baby smiling, I find myself smiling.

If I'm watching someone walking on a balance beam, I will lean when they lean.

If I'm having a conversation with someone and they scratch their nose, I will also scratch my nose.

EQ (ER scale only) items

Seeing people cry doesn't really upset me.

I tend to get emotionally involved with a friend's problems.

I get upset if I see people suffering on news programs.

It is hard for me to see why some things upset people so much.

Other people often say that I am insensitive, though I don't always see why.

I really enjoy caring for other people.

If I say something that someone else is offended by, I think that that's their problem, not mine.

I usually stay emotionally detached when watching a film.

Friends usually talk to me about their problems as they say that I am very understanding.

I find it easy to put myself in somebody else's shoes.

I can't always see why someone should have felt offended by a remark.

Non-duplicative items from ECQ

I care what happens to other people.

I'm sensitive to the feelings of others.

I avoid hurting other people's feelings.

TEQ (non-duplicative items)

When someone else is feeling excited, I tend to get excited too.

It upsets me to see someone being treated disrespectfully.

I remain unaffected when someone close to me is happy.

I enjoy making other people feel better.

When a friend starts to talk about his/her problems, I try to steer the conversation towards something else.

I can tell when others are sad even when they do not say anything.

I find that I am "in tune" with other people's moods.

I do not feel sympathy for people who cause their own serious illnesses.

I become irritated when someone cries.

I am not really interested in how other people feel.

I get a strong urge to help when I see someone who is upset.

I find it silly for people to cry out of happiness.

ACME (AR scale only) items

It makes me feel good to help someone in need.

I get excited to give someone a gift that I think they will enjoy.

I don't worry much about hurting people's feelings.

I don't really care if other people feel happy.

I don't really care if people are feeling depressed.

Other people's feelings don't bother me at all.

I feel awful when I hurt someone's feelings.

Other people's misfortunes don't bother me much.

If I see that I am doing something that hurts someone, I will quickly stop.

I often try to help people feel better when they are upset.

I enjoy making others happy.

People have told me that I'm insensitive.

BES items (Affective scale only, revised by removing references to friends)

Other people's emotions don't affect me much.

After being with someone who is sad about something, I usually feel sad.

I get frightened when I watch characters in a good scary movie.

I get caught up in other people's feelings easily.

I don't become sad when I see other people crying.

Other people's feeling don't bother me at all.

I often become sad when watching sad things on TV or in films.

Seeing a person who has been angered has no effect on my feelings.

I tend to feel scared when I am with people who are afraid.

I often get swept up in other people's feelings.

Other people's unhappiness doesn't make me feel anything.

QCAE contagion, PT, CON, and EC (non-duplicative)

I am happy when I am with a cheerful group and sad when the others are glum.

It worries me when others are worrying and panicky.

People I am with have a strong influence on my mood.

I am inclined to get nervous when others around me seem to be nervous.

I always try to consider the other fellow's feelings before I do something.

I can usually appreciate the other person's viewpoint, even if I do not agree with it.

Before I do something I try to consider how my friends will react to it.

I often get deeply involved with the feelings of a character in a film, play, or novel.

I often get emotionally involved with my friends' problems.

Friends talk to me about their problems as they say that I am very understanding.

It affects me very much when one of my friends seems upset.

I get very upset when I see someone cry.

New Items

Perspective-Taking Reasonableness/Motivation

Truthfully, I make a big effort to sympathetically understand people when I have conflicts with them.

When I start to get upset with people, it is really hard for me to see things from their perspectives.

If I start to feel frustrated with someone, I focus on understanding that person's point of view.

When someone disagrees with me, I think deeply about how things must feel from his/her point of view.

When I have a conflict with someone, I usually make a real effort to see things from his/her emotional perspective.

When I am at odds with someone, I sympathetically imagine how things might feel like from the other person's perspective.

To be perfectly honest, I don't care all that much about how other people experience their lives.

I sincerely try hard to understand how other people experience the world.

In truth, I am not all that interested in how most other people experience their emotions.

Positive Emotionality Caring

I go out of my way to bring joy to others.

I try to help other people feel really great about their accomplishments.

I make lots of small efforts to bring happiness to others.

I try to make sure that the people around me feel good about themselves.

Even if I don't succeed at something, I am still sincerely happy for those who do succeed.

I do lots of little things to nurture happiness in the people around me.

I love helping people feel proud of themselves.

I love boosting other people's self esteem.

I care a lot about being a very encouraging person to others.

I give out sincere compliments pretty frequently.

Suffering Caring

I am a calming presence for people when they are suffering.

I work to be a healing presence for people who feel down about themselves.

I tend to nurture and support people around me who are struggling.

Empathic Restraint

Even when I am upset, I almost never lash out in way that causes pain to someone.

I sometimes say hurtful words to others when I get upset.

I am known to have a pretty sharp tongue when I am frustrated.

To be honest, I can have a pretty aggressive temper sometimes.

When I can tell that what I am saying is beginning to hurt someone, it makes me want to be very gentle towards that person.

I have insulted some people this past month.

I sometimes use harsh words to criticize people to their faces.

When I am arguing with someone and they begin to feel sad, it makes me feel genuine compassion for them.

I try to make sure people are never afraid of my temper.

If I can see that someone is intimidated by me, I try to make them feel safe and relaxed.

I never try to shame people when I have a conflict with them.

Confidante Listening

People see me as a truly non-judgmental listener.

People feel quite comfortable opening up to me about their hopes and dreams.

People tell me that I am a comforting person to talk to.

People tell me that I am good at understanding them.

I put a lot of effort into being a caring listener.

Honestly, I am not all that motivated to be a really supportive listener.

Whenever someone opens up to me, I ask gentle questions to help them express their feelings.

I am often like a kind therapist with my friends.

When I am listening to someone, I put in real effort to make him/her feel sincerely understood.

When I am listening to someone, I ask questions to help me understand what it feels like to be that person.

When someone opens up to me, I focus on helping the person truly feel heard.

When someone opens up to me, I focus on understanding how it must feel to be that person.

When people talk to me about their personal struggles, I frequently check with them to make sure I am understanding their feelings accurately.

People see me as an emotionally safe person to open up to about their troubles.

When someone opens up to me, I ask gentle questions to help the person unpack their emotions.

Compared to others, I am a particularly supportive listener.

When my friends talk to me about their problems, I listen deeply to understand their feelings.

When people tell me their personal stories, they can tell that I genuinely care about their feelings.

People feel good opening up to me because they know I won't be judgmental.

When I am listening to someone, I try to genuinely emotionally connect with the person.

When I am talking to someone, I usually focus more on my own feelings than on the other person's feelings.

When I am listening to someone, I work to put my own feelings to the side in order to better listen.

When someone tells me about their struggles, it usually becomes a conversation about my own feelings.

When someone opens up to me, I focus on their feelings much more than my own.

Empathic Approach-Avoidance

When someone starts to talk to me about his or her emotional struggles, I usually change the subject to something less serious.

If people start to talk about really sad things in their lives, I try to get out of there pretty quickly.

I just don't feel very comfortable listening to people talk about their emotions.

To be perfectly honest, I do not actually like it when people start talking a lot about their feelings.

When people start talking about their own painful feelings, I sometimes only half-listen.

I tend to purposefully avoid people who are feeling sad.

If I can tell that an acquaintance is really upset, I usually make an effort to reach out to them.

When I see that someone is lonely, I often try to sincerely emotionally connect with the person.

I am quite comfortable when people open up about their deepest emotions.

If I can tell that someone needs a supportive listener, I go out of my way to give them my time and attention.

I tend to avoid people who have low self-esteem.

I usually avoid people who are depressed.

People who feel sorry for themselves tend to irritate me.

Aversive Contagion

If I am around someone who is feeling nervous, I get really nervous.

I tend to get pretty sad when the people around me feel sad.

I find myself feeling angry when I am around someone who is angry.

When I am near people who are upset, I can't help but feel upset too.

When I spend time with someone who is depressed, I start feeling pretty depressed too.

Appetitive Contagion

When someone around me is really excited about his or her accomplishment, I get genuinely excited as well.

Even when I am around people who are joyful, it is hard for me to really feel that joy.

If I see a stranger who is really happy, I tend to then also feel genuinely happy.

If I am with someone who is very happy, that person's emotions do not infect me all that much.

When I see strangers smiling at each other, I feel like smiling too.

When other people are reallycheerful, I do not usually catch their emotions myself.

When I see strangers having a lot of fun, I can't help but feel pleasure too.

When strangers around me are joyful, my heart starts to fill with joy too.

When people around me are laughing, I can't help but feel like laughing too.

When people around me are really happy, my heart usually swells with happiness too.

When I see people hugging each other, I feel warm inside.

Empathic Absorption

I tend to get emotionally involved with the stories people tell me about their life experiences.

When someone reveals his or her stories to me, I instinctively imagine what it might feel like to be that person.

When people open up to me, I can't help but feel imaginatively involved in the stories they tell.

Most people's experiences are just not very interesting to me.

When the people close to me talk about their experiences, I get really swept up in them.

I get really absorbed in the stories that people tell me about themselves.

I like to imagine myself living the lives of my friends and family members.

Most peoples' lives just seem really boring to me.

To be honest, most people are not very interesting to think about.

Empathic Distress

I get really nervous when people talk to me about something terrible that has happened to them.

When someone tells me about something really upsetting he or she has experienced, I can easily stay calm for the person.

Emotional intimacy usually distresses me.

Conversations about deep emotions just stress me out.

I get uncomfortable when people start talking about their emotional struggles.

I tend to get anxious when people start talking about their emotional trauma.

I tend to get quite uneasy when people start talking about their painful past experiences.

Empathy Valuing

When choosing friends, I care more about a person's compassionate nature than about his or her other qualities.

More than anything else, I want to be remembered as a compassionate person.

A big goal in my life is to sympathetically understand how other people feel.

I spend a lot of time thinking about how I can be more caring to my friends and family.

I would rather be seen as “rich but somewhat self-centered” than “poor but very compassionate”.

I strive to be extremely merciful towards the people in my life.

Compassion is one of the most important virtues I wish to model in my own life.

Loving kindness is worth more to me than money or success.

It is very important to me that I help others feel truly understood.

I aim to make sure that my friends feel truly seen and understood.

Appendix B. Finalized GFEC Scale and Specific Content Scales

General Factor of Empathic Caring Scale (GFEC)

- ***Formed through a combination of a basic correlated factors EFA model, with further item culling based upon agreements between exploratory bifactor models with correlated or uncorrelated specific factors. Final items also selected using IRT.***

- ***All items identified as quasi-pure indicators of the general empathic caring construct. The scale appears to be strongly unidimensional and internally consistent.***

- ***For use as either a stand-alone general empathic caring scale or as a control variable***

- Honestly, I tend to supportively nurture people around me who are struggling.
- I go out of my way to bring joy to others.
- It is very important to me that I help others feel truly understood.
- When I see that someone is lonely, I often try to sincerely emotionally connect with the person.
- I work to be a healing presence for people who feel down about themselves.
- I get a strong urge to help when I see someone who is upset.
- I really enjoy caring for other people.
- I make lots of small efforts to bring happiness to others.
- I care a lot about being a very encouraging person to others.
- More than anything else, I want to be remembered as a compassionate person.
- I try to make sure that the people around me feel good about themselves.
- I would describe myself as a pretty soft-hearted person.
- I love boosting other people's self-esteem.
- I often have tender, concerned feelings for people less fortunate than me.
- I often try to help people feel better when they are upset.

Specific Content Scales of Empathic Caring

NOTE: all of these scales can be reduced in number of items, if a brief version is required

Appetitive Contagion (AC)

- Even when I am around people who are joyful, it is hard for me to really feel that joy.
- When I see strangers having a lot of fun, I can't help but feel pleasure too.
- When someone around me is really excited about his or her accomplishment, I get genuinely excited as well.
- When strangers around me are joyful, my heart starts to fill with joy too.
- When other people are cheerful and merry, I do not usually catch their emotions myself.
- If strangers around me are really happy, I tend to also start feeling really happy

Perspective-Taking (PT)

- When I'm upset at someone, I usually try to "put myself in his shoes" for a while.
- If I start to feel frustrated with someone, I focus on understanding that person's point of view.
- I find it easy to put myself in somebody else's shoes.
- Before criticizing somebody, I try to imagine how I would feel if I were in their place.
- I sometimes try to understand my friends better by imagining how things look from their perspective.
- Even when I have a conflict with someone, I think deeply about how things must feel from his/her point of view.
- I make sure to sympathetically imagine the feelings of other people even when they are aggressive towards me.

Empathic Avoidance (EAV)

- If people start to talk about really sad things in their lives, I try to get out of there pretty quickly.
- When a friend starts to talk about his/her problems, I try to steer the conversation towards something else.
- When someone starts to talk to me about his or her emotional struggles, I usually change the subject to something less serious.
- In truth, I am not all that interested in how most other people experience their emotions.
- I tend to purposefully avoid people who are feeling sad.
- When people start talking about their own painful feelings, I sometimes only half-listen.
- To be perfectly honest, I do not actually like it when people start talking a lot about their feelings.
- I try to avoid having to listen to people talk about their deep emotions.

Empathic Restraint (ER)

- I feel awful when I hurt someone's feelings.
- I avoid hurting other people's feelings.
- If I see that I am doing something that hurts someone, I will quickly stop.
- I don't worry much about hurting people's feelings.
- When I can tell that what I am saying is beginning to hurt someone, it makes me want to be very gentle towards that person.

Confidante Listening (CL)

-Note: this scale is not reliably distinguishable from the GFEC scale (or the general factor of empathic caring). I recommend only using this scale if you are specifically interested in the construct of confidante listening.

- People feel quite comfortable opening up to me about their hopes and dreams.
- People tell me that I am a comforting person to talk to.
- Friends talk to me about their problems as they say that I am very understanding.
- I am a calming presence for people when they are suffering.
- People feel safe opening up to me because they know I won't be judgmental.
- Compared to others, I am a particularly supportive listener.
- I am often like a kind therapist with my friends.

Specific Content Scales of Empathy-Related Negative Emotionality

Empathic Intimacy Discomfort

I get uncomfortable when people start talking about their emotional struggles.
Emotional intimacy usually distresses me.
When people share their deep emotions, I feel apprehensive and ill-at-ease.
I tend to get quite uneasy when people start talking about their painful past experiences.
Conversations about deep emotions just stress me out.
I tend to get anxious when people start talking about their emotional trauma.
I just don't feel very comfortable listening to people talk about their emotions.

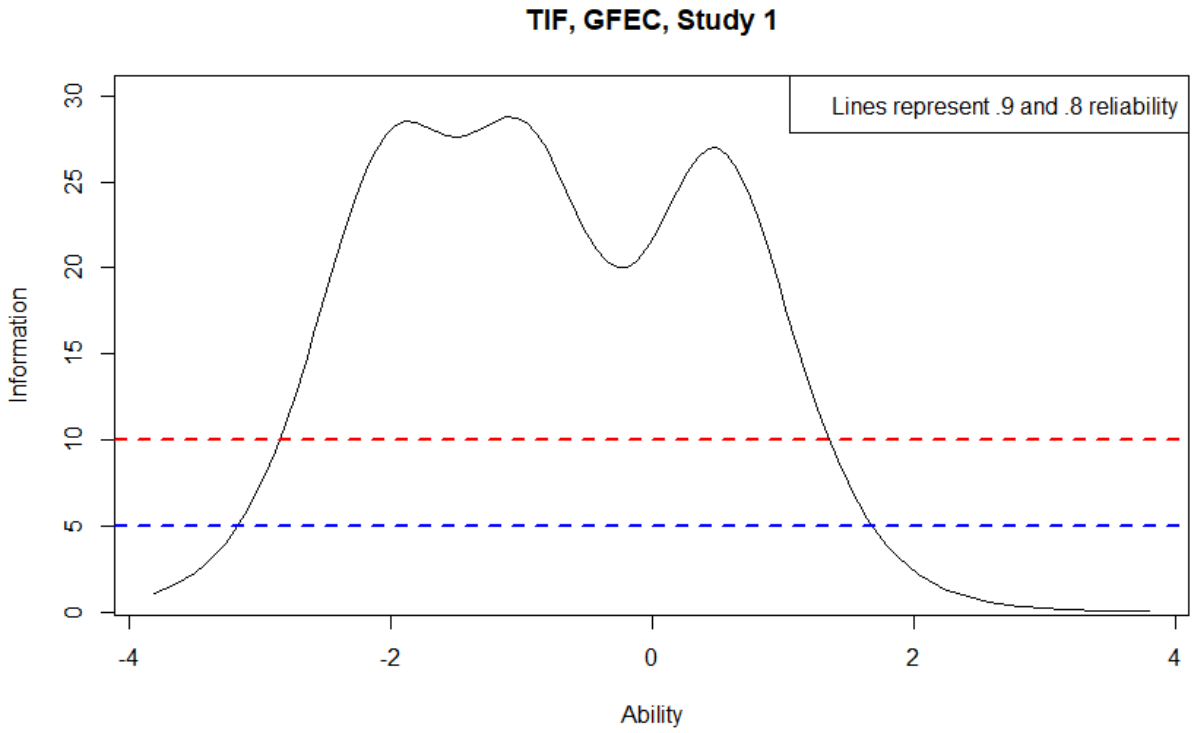
Contagion for Sadness

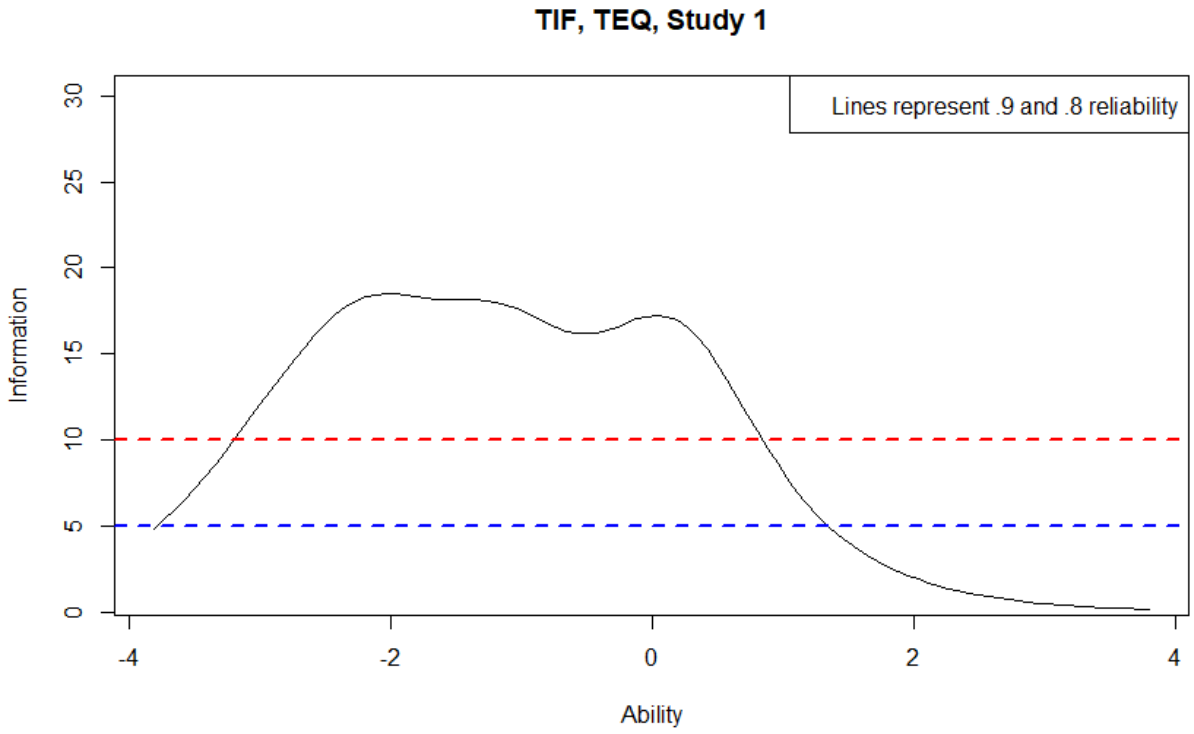
I tend to get pretty sad when I am around someone who is sad.
After being with someone who is sad about something, I usually feel sad.
When I am near people who are upset, I can't help but feel upset too.
When I spend time with someone who is depressed, I start feeling pretty depressed too.
If I am around someone who is feeling sorrowful, I can't help but feel really sorrowful as well.
When I see someone crying, it makes me feel like crying too.

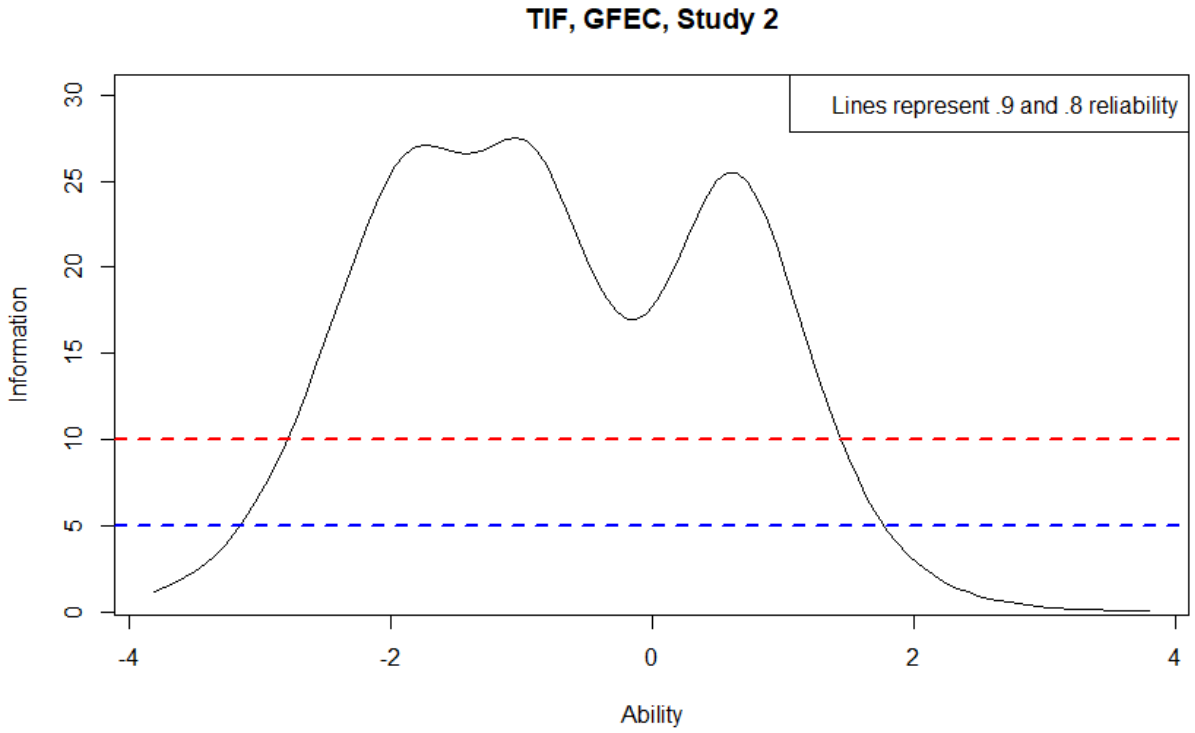
Contagion for Anxiety/Fear

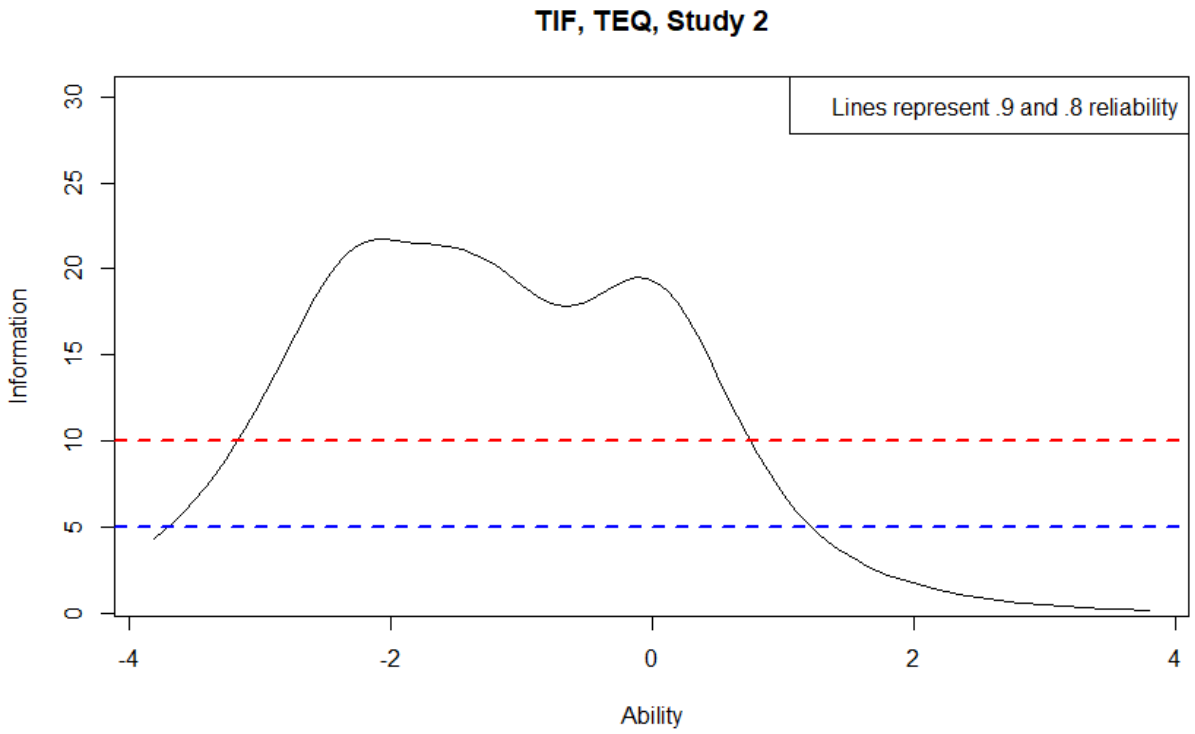
If I am around someone who is feeling nervous, I get really nervous.
If I see someone fidgeting, I'll start feeling anxious too.
I tend to feel scared when I am with people who are afraid.
I am inclined to get nervous when others around me seem to be nervous.
It worries me when others are worrying and panicky.
I start feeling really anxious when the people around me are anxious.
If I see someone feeling really stressed out, I start to get quite stressed out too.

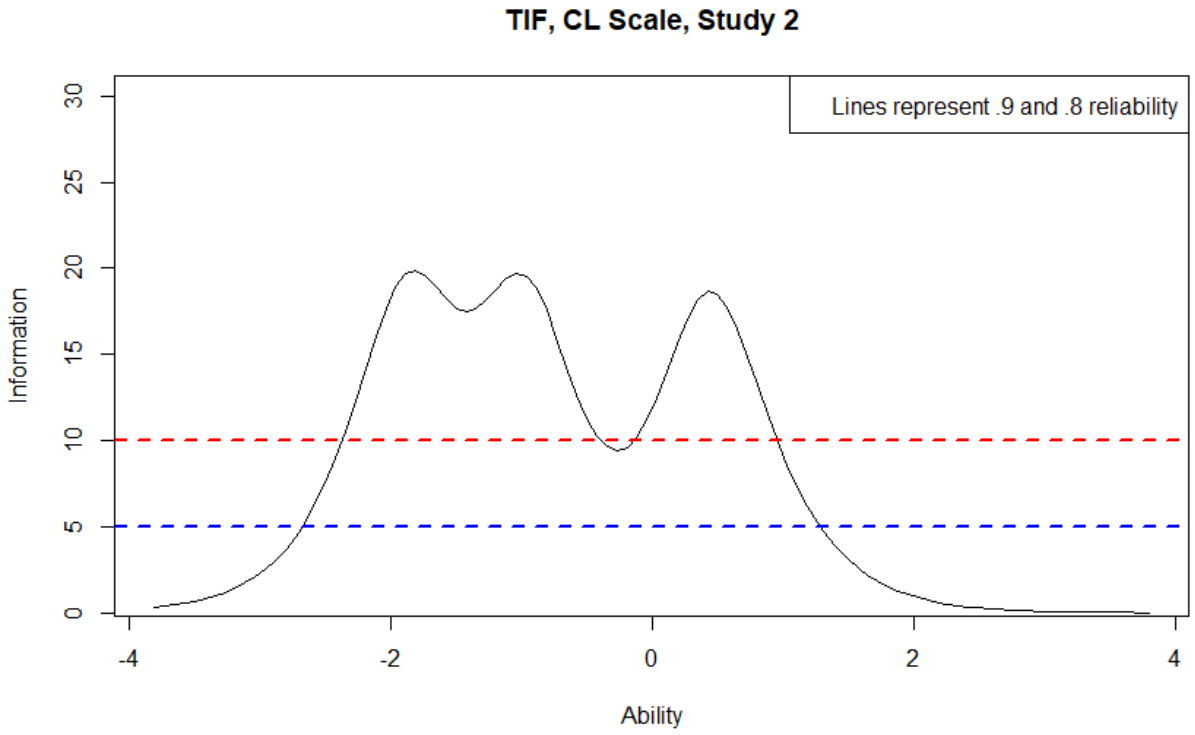
Appendix C. Test Information Function (TIF) Curves, Studies 1 and 2

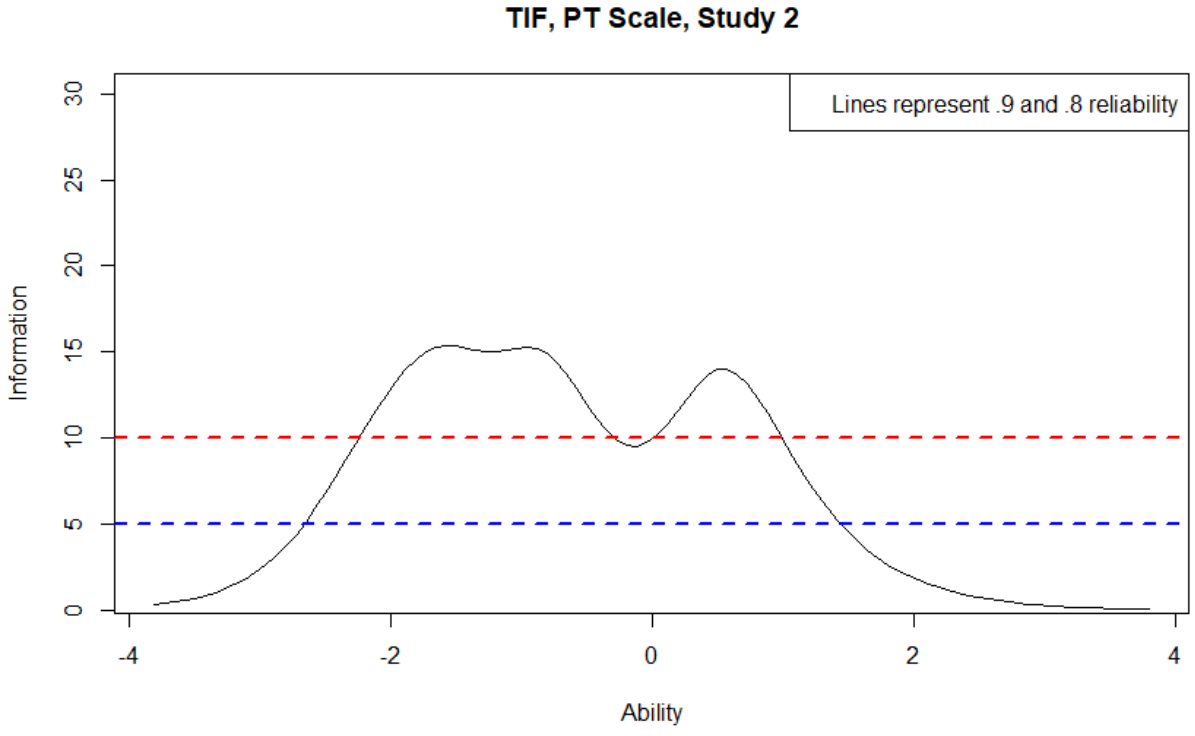


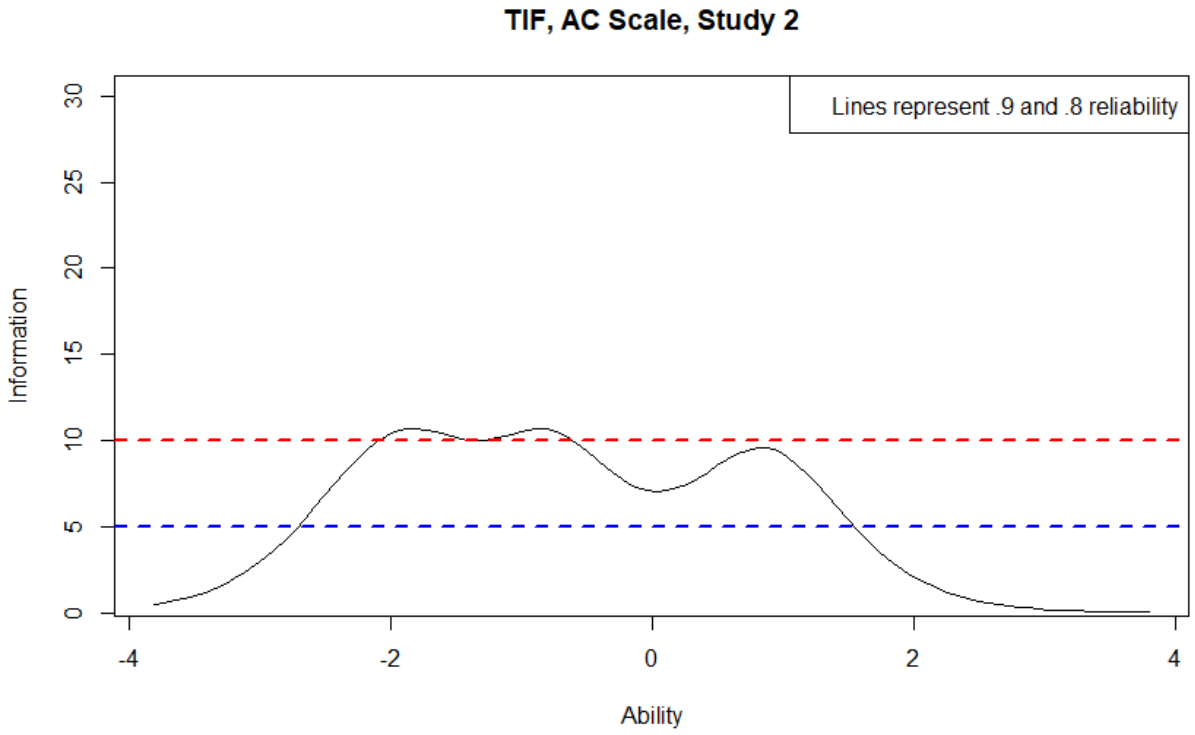


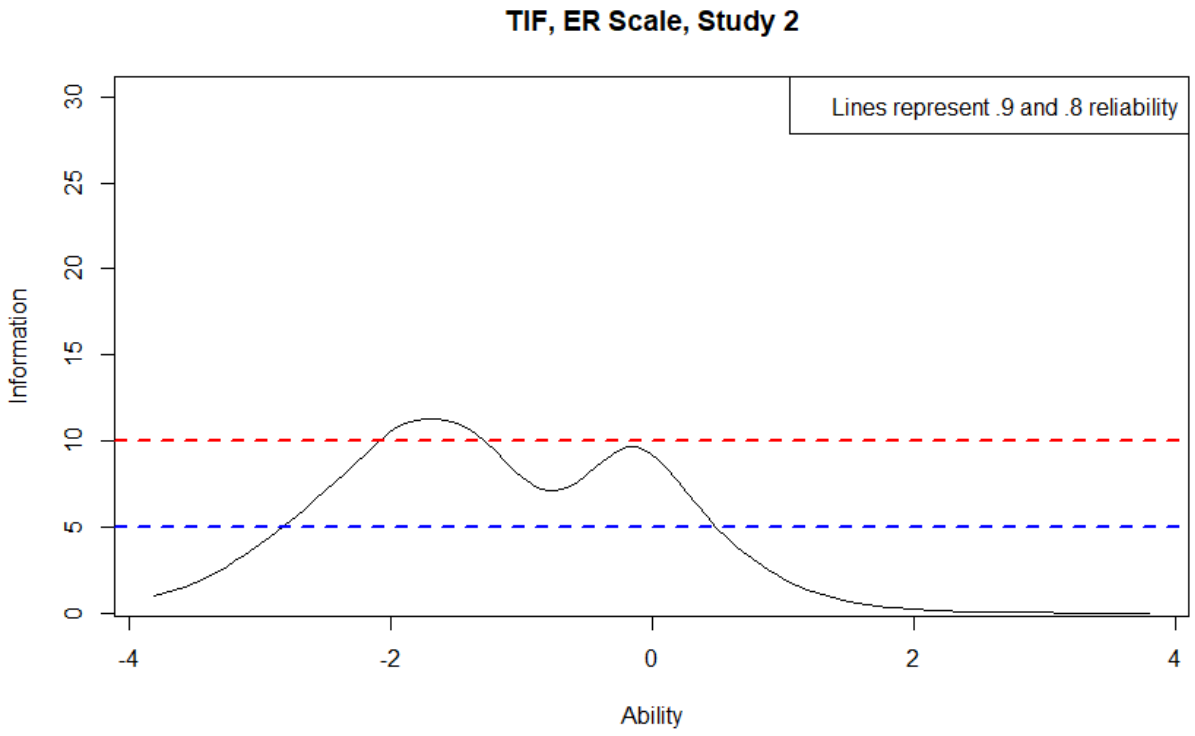

TIF, GFEC, Study 2
Lines represent .9 and .8 reliability
Information
0
5
10
15
20
25
30
-4
-2
0
2
4
Ability

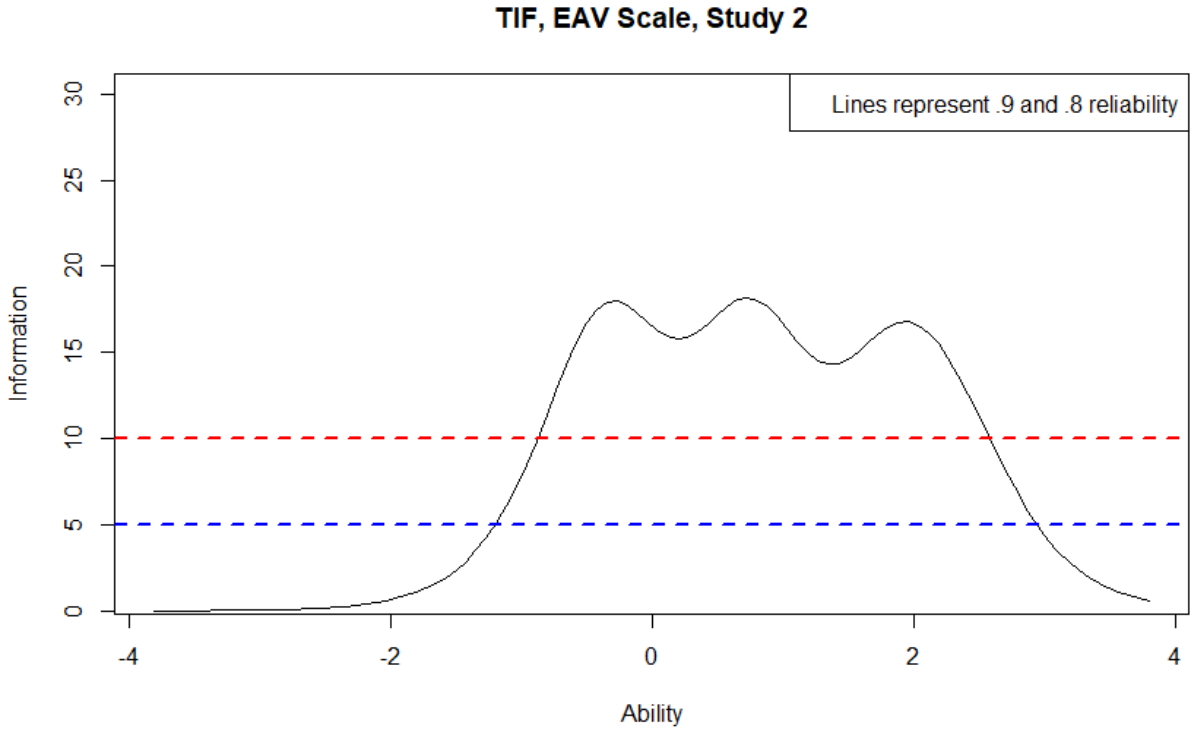

TIF, TEQ, Study 2
Lines represent .9 and .8 reliability
Information
0
5
10
15
20
25
30
-4
-2
0
2
4
Ability

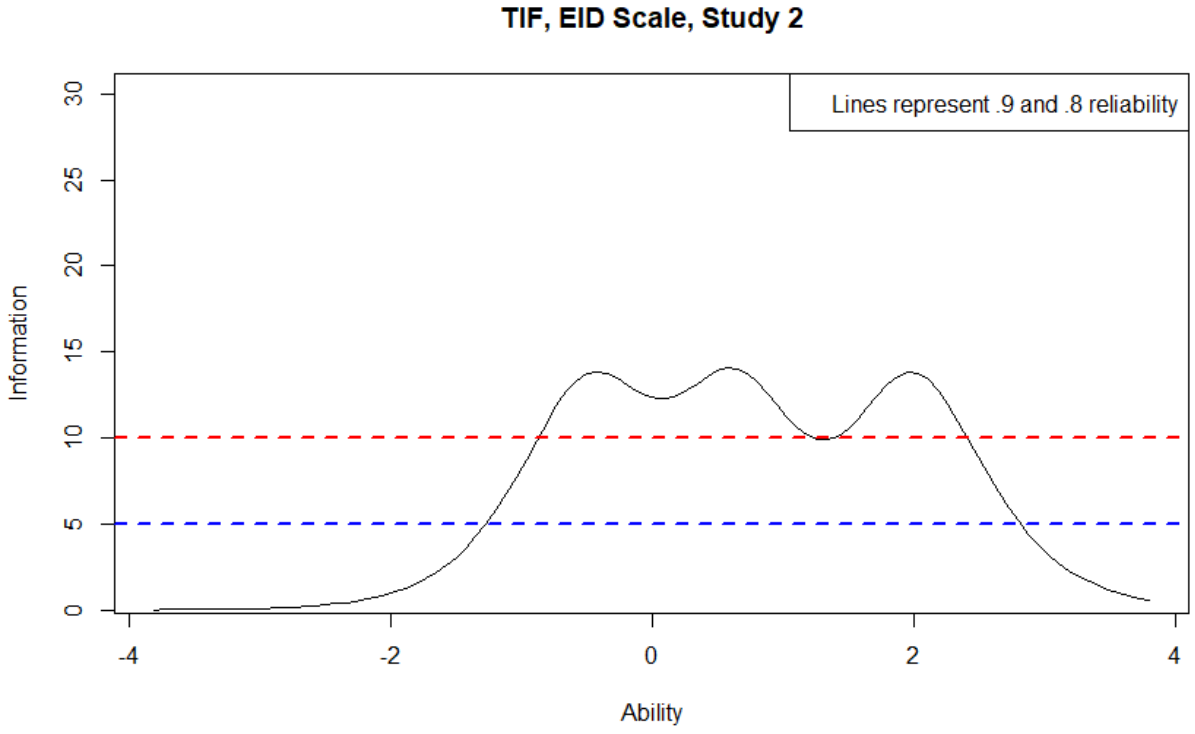

TIF, CL Scale, Study 2
Lines represent .9 and .8 reliability
Information
0
5
10
15
20
25
30
-4
-2
0
2
4
Ability

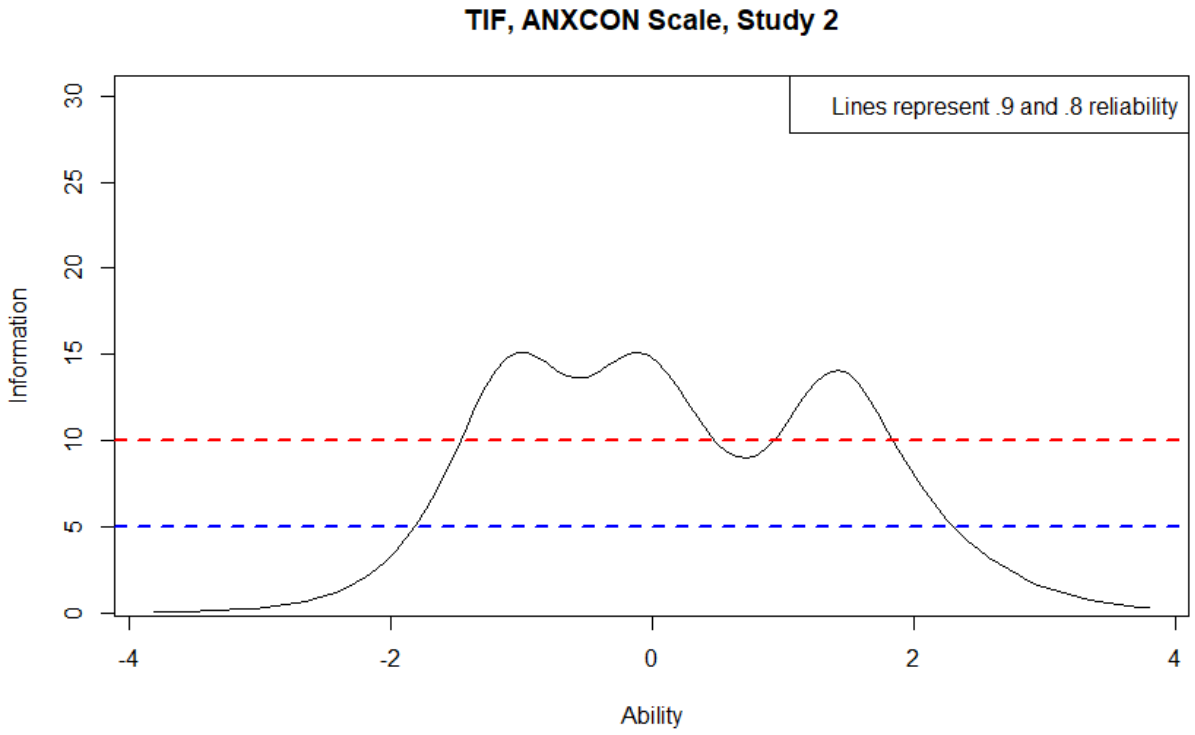

TIF, PT Scale, Study 2
Lines represent .9 and .8 reliability
Information
0
5
10
15
20
25
30
-4
-2
0
2
4
Ability

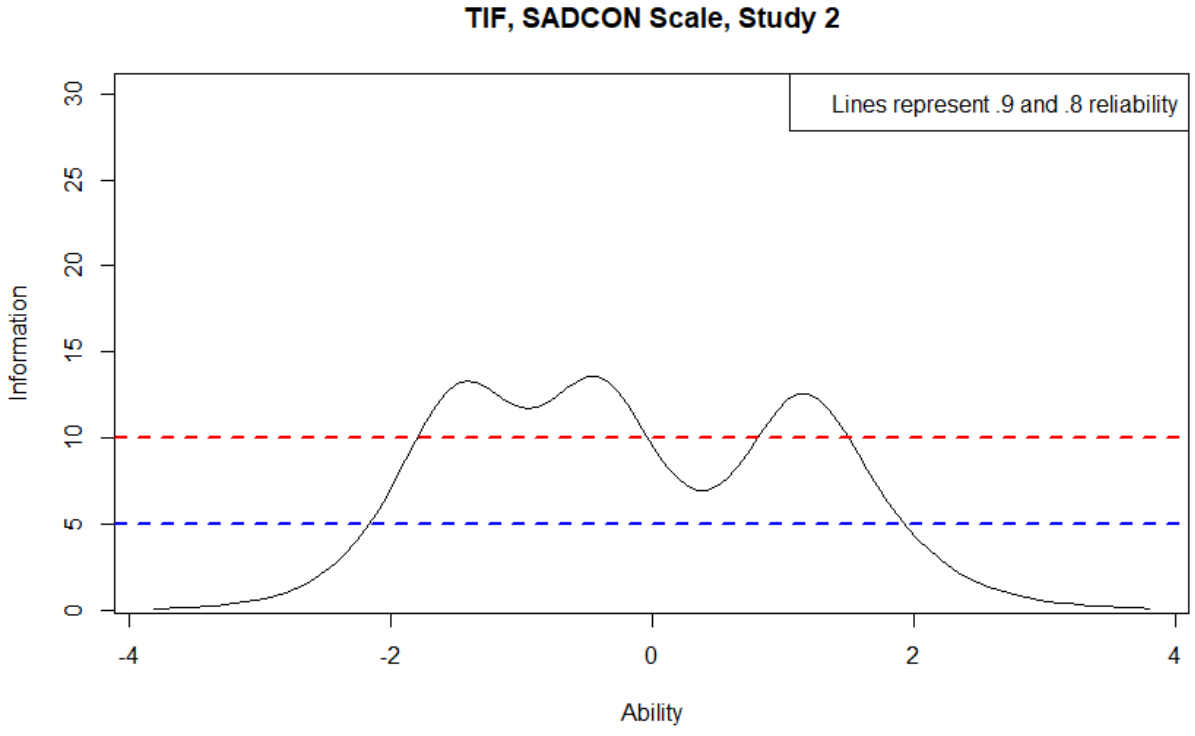

TIF, AC Scale, Study 2
Lines represent .9 and .8 reliability
Information
0
5
10
15
20
25
30
-4
-2
0
2
4
Ability


TIF, ER Scale, Study 2
Lines represent .9 and .8 reliability
Information
0
5
10
15
20
25
30
-4
-2
0
2
4
Ability


TIF, EAV Scale, Study 2
Lines represent .9 and .8 reliability
Information
0
5
10
15
20
25
30
-4
-2
0
2
4
Ability


TIF, EID Scale, Study 2
Lines represent .9 and .8 reliability
Information
0
5
10
15
20
25
30
-4
-2
0
2
4
Ability


TIF, ANXCON Scale, Study 2
Lines represent .9 and .8 reliability
Information
0
5
10
15
20
25
30
-4
-2
0
2
4
Ability


TIF, SADCON Scale, Study 2
Lines represent .9 and .8 reliability
Information
0
5
10
15
20
25
30
-4
-2
0
2
4
Ability

Appendix D. Item Keys, Study 1 full empathy item pool

IRIPT1_R	I sometimes find it difficult to see things from the "other guy's" point of view.
IRIPT2	I try to look at everybody's side of a disagreement before I make a decision.
IRIPT3	I sometimes try to understand my friends better by imagining how things look from their perspective.
IRIPT4_R	If I'm sure I'm right about something, I don't waste much time listening to other people's arguments.
IRIPT5	I believe that there are two sides to every question and try to look at them both.
IRIPT6	When I'm upset at someone, I usually try to "put myself in his shoes" for a while.
IRIPT7	Before criticizing somebody, I try to imagine how I would feel if I were in their place.
IRIEC1	I often have tender, concerned feelings for people less fortunate than me.
IRIEC2_R	Sometimes I don't feel very sorry for other people when they are having problems.
IRIEC3	When I see someone being taken advantage of, I feel kind of protective towards them.
IRIEC4_R	Other people's misfortunes do not usually disturb me a great deal.
IRIEC5_R	When I see someone being treated unfairly, I sometimes don't feel very much pity for them.
IRIEC6	I am often quite touched by things that I see happen.
IRIEC7	I would describe myself as a pretty soft-hearted person.
Q75	I really get involved with the feelings of people when they are telling me stories about themselves.
Q76	I am usually objective when I hear someone's story, and I don't often get completely caught up in it.
Q77	Becoming emotionally involved in another person's story is very rare for me.
Q78	After hearing a person tell me her story, I have felt as though I were that person to some extent.
Q79	When I hear a good story from someone, I can very easily put myself in the place of the person.
Q80	When I am hearing a person's story, I imagine how I would feel if the events in the story were happening to me.
Q81	When people share their deep emotions, I feel apprehensive and ill-at-ease.
Q82	I sometimes feel helpless when I listen to someone talk about their suffering.
Q83	When I see someone terribly upset, I tend to remain calm.
Q84	I feel scared when I am in a conversation with someone who is opening up about his or her deepest worries.
Q85	I am usually pretty calm when dealing with other people's emotions of distress.
Q86	I tend to lose control during really emotional conversations.
Q87	When I spend time with someone who is suffering from emotional trauma, I often fall to pieces.
EIPNC1	If I see someone who is excited, I will feel excited myself.
EIPNC2	I sometimes find myself feeling the emotions of the people around me, even if I don't try to feel what they're feeling.
EIDC1	If I'm watching a movie and a character injures their leg, I will feel pain in my leg.

EIDC2	If I hear a story in which someone is scared, I will imagine how scared I would be in that situation and begin to feel scared myself.
EIDC3	I can't watch shows in which an animal is being hunted by another because I feel nervous as if I am being hunted.
EIDC4	If I see someone fidgeting, I'll start feeling anxious too.
EIPM1	I catch myself crossing my arms or legs just like the person I'm talking to.
EIPNC3	If I see a video of a baby smiling, I find myself smiling.
EIPM2	If I'm watching someone walking on a balance beam, I will lean when they lean.
EIPM3	If I'm having a conversation with someone and they scratch their nose, I will also scratch my nose.
EQ1	Seeing people cry doesn't really upset me.
EQ2	I tend to get emotionally involved with a friend's problems.
EQ3	I get upset if I see people suffering on news programs.
EQ4	It is hard for me to see why some things upset people so much.
EQ5	Other people often say that I am insensitive, though I don't always see why.
EQ6	I really enjoy caring for other people.
EQ7	If I say something that someone else is offended by, I think that that's their problem, not mine.
EQ8	I usually stay emotionally detached when watching a film.
EQ9	Friends usually talk to me about their problems as they say that I am very understanding.
EQ10	I find it easy to put myself in somebody else's shoes.
EQ11	I can't always see why someone should have felt offended by a remark.
ECQ1	I care what happens to other people.
ECQ2	I'm sensitive to the feelings of others.
ECQ3	I avoid hurting other people's feelings.
TEQ1	When someone else is feeling excited, I tend to get excited too.
TEQ2	It upsets me to see someone being treated disrespectfully.
TEQ3	I remain unaffected when someone close to me is happy.
TEQ4	I enjoy making other people feel better.
TEQ5	When a friend starts to talk about his/her problems, I try to steer the conversation towards something else.
TEQ6	I can tell when others are sad even when they do not say anything.
TEQ7	I find that I am "in tune" with other people's moods.
TEQ8	I do not feel sympathy for people who cause their own serious illnesses.
TEQ9	I become irritated when someone cries.
TEQ10	I am not really interested in how other people feel.
TEQ11	I get a strong urge to help when I see someone who is upset.
TEQ12	I find it silly for people to cry out of happiness.
ACME1	It makes me feel good to help someone in need.
ACME2	I get excited to give someone a gift that I think they will enjoy.
ACME3	I don't worry much about hurting people's feelings.
ACME4	I don't really care if other people feel happy.
ACME5	I don't really care if people are feeling depressed.

ACME6	Other people's feelings don't bother me at all.
ACME7	I feel awful when I hurt someone's feelings.
ACME8	Other people's misfortunes don't bother me much.
ACME9	If I see that I am doing something that hurts someone, I will quickly stop.
ACME10	I often try to help people feel better when they are upset.
ACME11	I enjoy making others happy.
ACME12	People have told me that I'm insensitive.
BES1	Other people's emotions don't affect me much.
BES2	After being with someone who is sad about something, I usually feel sad.
BES3	I get frightened when I watch characters in a good scary movie.
BES4	I get caught up in other people's feelings easily.
BES5	I don't become sad when I see other people crying.
BES6	Other people's feeling don't bother me at all.
BES7	I often become sad when watching sad things on TV or in films.
BES8	Seeing a person who has been angered has no effect on my feelings.
BES9	I tend to feel scared when I am with people who are afraid.
BES10	I often get swept up in other people's feelings.
BES11	Other people's unhappiness doesn't make me feel anything.
QCAE1	I am happy when I am with a cheerful group and sad when the others are glum.
QCAE2	It worries me when others are worrying and panicky.
QCAE3	People I am with have a strong influence on my mood.
QCAE4	I am inclined to get nervous when others around me seem to be nervous.
QCAE5	I always try to consider the other fellow's feelings before I do something.
QCAE6	I can usually appreciate the other person's viewpoint, even if I do not agree with it.
QCAE7	Before I do something I try to consider how my friends will react to it.
QCAE8	I often get deeply involved with the feelings of a character in a film, play, or novel.
QCAE9	I often get emotionally involved with my friends' problems.
QCAE10	Friends talk to me about their problems as they say that I am very understanding.
QCAE11	It affects me very much when one of my friends seems upset.
QCAE12	I get very upset when I see someone cry.
Q159	Truthfully, I make a big effort to sympathetically understand people when I have conflicts with them.
Q160	When I start to get upset with people, it is really hard for me to see things from their perspectives.
Q161	If I start to feel frustrated with someone, I focus on understanding that person's point of view.
Q162	Even when I have a conflict with someone, I think deeply about how things must feel from his/her point of view.
Q163	When I have a conflict with someone, I usually make a real effort to see things from his/her emotional perspective.
Q164	When I am at odds with someone, I sympathetically imagine how things might feel like from the other person's perspective.
Q165	To be perfectly honest, I don't care all that much about how other people experience their lives.

Q166	I sincerely try hard to understand how other people experience the world.
Q167	In truth, I am not all that interested in how most other people experience their emotions.
Q168	I go out of my way to bring joy to others.
Q169	I don't try very hard to help other people feel really great about their accomplishments.
Q170	I make lots of small efforts to bring happiness to others.
Q171	I try to make sure that the people around me feel good about themselves.
Q172	Even if I don't succeed at something, I am still sincerely happy for those who do succeed.
Q173	I do lots of little things to nurture happiness in the people around me.
Q174	I love helping people feel proud of themselves.
Q175	I love boosting other people's self esteem.
Q176	I care a lot about being a very encouraging person to others.
Q177	I give out sincere compliments pretty frequently.
Q178	I am a calming presence for people when they are suffering.
Q179	I work to be a healing presence for people who feel down about themselves.
Q180	Honestly, I tend to supportively nurture people around me who are struggling.
Q181	Even when I am upset, I almost never lash out in way that causes pain to someone.
Q182	I sometimes say hurtful words to others when I get upset.
Q183	I am known to have a pretty sharp tongue when I am frustrated.
Q184	To be honest, I can have a pretty aggressive temper sometimes.
Q185	When I can tell that what I am saying is beginning to hurt someone, it makes me want to be very gentle towards that person.
Q186	I have insulted some people this past month.
Q187	I sometimes use harsh words to criticize people to their faces.
Q188	When I am arguing with someone and they begin to feel sad, it makes me feel genuine compassion for them.
Q189	I try to make sure people are never afraid of my temper.
Q190	If I can see that someone is intimidated by me, I try to make them feel safe and relaxed.
Q191	I never try to shame people when I have a conflict with them.
Q192	Compared to others, people see me as a truly non-judgmental listener.
Q193	People feel quite comfortable opening up to me about their hopes and dreams.
Q194	People tell me that I am a comforting person to talk to.
Q195	People tell me that I am good at helping them feel understood.
Q196	I put a lot of effort into being a caring listener.
Q197	Honestly, I am not all that motivated to be a really supportive listener.
Q198	Whenever someone opens up to me, I ask gentle questions to help them express their feelings.
Q199	I am often like a kind therapist with my friends.
Q200	When I am listening to someone, I put in real effort to make him/her feel sincerely understood.
Q201	When I am listening to someone, I ask questions to help me understand what it feels like to be that person.
Q202	When someone opens up to me, I focus on helping the person feel deeply understood.

Q203	When someone opens up to me, I focus on understanding how it must feel to be that person.
Q204	When people talk to me about their personal struggles, I frequently check with them to make sure I am understanding their feelings accurately.
Q205	Compared to others, people see me as an emotionally safe person to open up to about their troubles.
Q206	When someone opens up to me, I ask gentle questions to help the person unpack their emotions.
Q207	Compared to others, I am a particularly supportive listener.
Q208	When my friends talk to me about their problems, I listen deeply to understand their feelings.
Q209	When people tell me their personal stories, they can tell that I genuinely care about their feelings.
Q210	People feel safe opening up to me because they know I won't be judgmental.
Q211	When I am listening to someone, I genuinely emotionally connect with the person.
Q212	When I am talking to someone, I usually focus more on my own feelings than on the other person's feelings.
Q213	When I am listening to someone, put my own feelings to the side in order to better listen.
Q214	When someone tells me about their struggles, it often becomes a conversation more about my own feelings.
Q215	When someone opens up to me, I focus on their feelings much more than my own.
Q216	When someone starts to talk to me about his or her emotional struggles, I usually change the subject to something less serious.
Q217	If people start to talk about really sad things in their lives, I try to get out of there pretty quickly.
Q218	I just don't feel very comfortable listening to people talk about their emotions.
Q219	To be perfectly honest, I do not actually like it when people start talking a lot about their feelings.
Q220	When people start talking about their own painful feelings, I sometimes only half-listen.
Q221	I tend to purposefully avoid people who are feeling sad.
Q222	If I can tell that an acquaintance is really upset, I usually make an effort to reach out to them.
Q223	When I see that someone is lonely, I often try to sincerely emotionally connect with the person.
Q224	I am quite comfortable when people open up about their deepest emotions.
Q225	If I can tell that someone needs a supportive listener, I go out of my way to give them my time and attention.
Q226	I tend to avoid people who have low self-esteem.
Q227	I usually avoid people who are depressed.
Q228	People who feel sorry for themselves tend to irritate me.
Q229	If I am around someone who is feeling nervous, I get really nervous.
Q230	I tend to get pretty sad when I am around someone who is sad.
Q231	I find myself feeling angry when I am around someone who is angry.
Q232	When I am near people who are upset, I can't help but feel upset too.

Q233	When I spend time with someone who is depressed, I start feeling pretty depressed too.
Q234	When someone around me is really excited about his or her accomplishment, I get genuinely excited as well.
Q235	Even when I am around people who are joyful, it is hard for me to really feel that joy.
Q236	If I see a stranger who is really happy, I tend to then also feel genuinely happy.
Q237	If I am with someone who is very happy, that person's emotions do not infect me all that much.
Q238	When I see strangers smiling at each other, I feel like smiling too.
Q239	When other people are cheerful and merry, I do not usually catch their emotions myself.
Q240	When I see strangers having a lot of fun, I can't help but feel pleasure too.
Q241	When strangers around me are joyful, my heart starts to fill with joy too.
Q242	When people around me are laughing, I can't help but feel like laughing too.
Q243	When people around me are really happy, my heart usually swells with happiness too.
Q244	When I see people hugging each other, I feel warm inside.
Q245	I tend to get emotionally involved with the stories people tell me about their life experiences.
Q246	When someone reveals his or her stories to me, I instinctively imagine what it might feel like to be that person.
Q247	When people open up to me, I can't help but feel imaginatively involved in the stories they tell.
Q248	Most people's experiences are just not very interesting to me.
Q249	When the people close to me talk about their experiences, I get really swept up in them.
Q250	I get really absorbed in the stories that people tell me about themselves.
Q251	I like to imagine myself living the lives of my friends and family members.
Q252	Most peoples' lives just seem boring to me.
Q253	To be honest, most people are not very interesting to think about.
Q254	I get really nervous when people talk to me about something terrible that has happened to them.
Q255	When someone tells me about something really upsetting he or she has experienced, I can easily stay calmly focused on listening.
Q256	Emotional intimacy usually distresses me.
Q257	Conversations about deep emotions just stress me out.
Q258	I get uncomfortable when people start talking about their emotional struggles.
Q259	I tend to get anxious when people start talking about their emotional trauma.
Q260	I tend to get quite uneasy when people start talking about their painful past experiences.
Q261	When choosing friends, I care more about a person's compassionate nature than about his or her other qualities.
Q262	More than anything else, I want to be remembered as a compassionate person.
Q263	A big goal in my life is to sympathetically understand how other people feel.
Q264	I spend a lot of time thinking about how I can be more caring to my friends and family.
Q265	I would rather be seen as "rich but somewhat self-centered" than "poor but very compassionate".
Q266	I strive to be extremely merciful towards the people in my life.
Q267	Loving kindness is one of the most important virtues I wish to model in my own life.

Q268	I would rather be extremely successful than be extremely kind.
Q269	It is very important to me that I help others feel truly understood.
Q270	Truthfully, I aim to make sure that my friends feel truly seen and understood.