INDIVIDUAL DIFFERENCES IN SOCIAL COGNITION AND BEHAVIOR: A PERSONALITY PSYCHOLOGY FRAMEWORK

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY

Scott Douglas Blain

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Colin G. DeYoung, Ph.D., advisor
Robert F. Krueger, Ph.D., co-advisor

August 2021

Acknowledgments

First, I would like to thank my advisor, Colin DeYoung, for his mentorship, guidance, and support. Without your enthusiasm and insight, there is no way I would be where I am today as a scientist or person. I would also like to thank my other committee members—Robert Krueger, Scott Sponheim, Angus MacDonald III, and James Lee. Moreover, I would like to acknowledge the foundational contributions of my undergraduate advisors and mentors, including Blythe Corbett, Sohee Park, Reyna Gordon, Rachel Aaron, Craig Smith, and Meg Saylor.

I appreciate the continued inspiration and collaboration from my current and former lab mates and colleagues Amanda Rueter, Aisha Udochi, Tyler Sassenberg, Matt Rogers, Edward Chou, Julia Longenecker, Yizhou Ma, Xiaoqin Wang, and Caroline Schlüter, as well as the contributions of the research assistants I have worked with over the past three years: Noemi Hedrick, Daiqing Zhao, Li Shen Chong, Muchen Xi, Seth Miller, Amanda Sati, Anna Peddle, Wade Simpson, Leroy Medrano, and others.

I would like to recognize the contributions of my co-authors and collaborators for the roles they played in the work presented throughout this dissertation. The work presented in Chapter 1 includes data analytic and writing contributions from Aisha Udochi, Timothy Allen, Muchen Xi, and Colin DeYoung.

Aisha Udochi also provided substantial contributions to Chapter 2 and serves as co-first author for a version of that work currently under review; other contributors to the work presented in Chapter 2 include Tyler Sassenberg, Phillip Burton, Leroy Medrano, and Colin DeYoung. The data presented in Chapter 2 of this dissertation were provided by the HCP, WU-Minnesota Consortium (Principal Investigators: Essen and Ugurbil;

1U54MH091657) funded by the NIH Institutes and Centers that support the Blueprint for Neuroscience Research and by the McDonnell Center at Washington University.

The work presented in Chapter 3 is adapted from material published by the American Psychological Association in the *Journal of Personality and Social Psychology* (https://doi.org/10.1037/pspp0000371), of which I am first author. I acknowledge the vital roles of my coauthors, Colin DeYoung, Tyler Sassenberg, Daiqing (Iris) Zhao, and Muchen Xi, for their contributions to data analysis, literature review, and writing.

The work presented in Chapter 4 is an evolution of research begun by Yanna Weisberg and Colin DeYoung several years ago, inspired by work on the interpersonal circumplex and the neurobiology of Affiliation. I would like to thank Yanna for her vital contributions to the creation and validation of the Trait Affiliation Scale presented in this chapter. I would also like to acknowledge the role of David Condon and collaborators from Synthetic Aperture Personality Assessment project for providing data for and making analytical contributions to the work presented in Chapter 4. Finally, thank you to Edward Chou and Muchen Xi for their feedback and contributions to the project.

Thank you to U of M's Graduate School, College of Liberal Arts, Department of Psychology, Institute of Child Development, and Counsel of Graduate Students, Augsburg University, Ruhr-Universität Bochum, the National Science Foundation, Association for Research in Personality, Society for Interpersonal Theory and Research, Society for Multivariate Experimental Psychology, and Society for Personality and Social Psychology for financial support throughout my U of M tenure.

Thank you to my friends—especially Ethan, Angelica, Colt, Matthew, Tim, and Arthur—and family, for your never-ending support (and proofreading). And finally, thank you, Weijie, for being with me throughout the dissertation writing process and through these crazy two years that have been 2020 and 2021.

Abstract

Though humans are universally social, we vary considerably in our ability and motivation to form and maintain relationships. One approach to explaining this variation looks to identify the mechanisms that facilitate social behavior, including social cognition and reward sensitivity. Much of this work, however, is methodologically lacking and fails to provide comprehensive explanatory frameworks. This dissertation applies insights from personality psychology to improve our understanding of individual differences in social cognition and interpersonal functioning, focusing on the broad traits most descriptive of social behavior: Agreeableness, Extraversion, and Trait Affiliation. Across four studies attempting to elucidate the neurocognitive mechanisms of these traits, various methods—including questionnaires, behavioral tasks, fMRI, and psychometric techniques—were used to elucidate how and why individuals vary in their social abilities, behaviors, and associated outcomes.

Study 1 was a multi-task investigation of how three Agreeableness-Antagonism subfactors differentially predict social cognitive ability. Study 2 used fMRI, along with personality questionnaires and behavioral tasks, to examine associations among Agreeableness, social cognitive ability, and function of the brain's default network, applying structural equation modeling and a Bayesian individualized cortical parcellation approach. Study 3 failed to replicate classic associations demonstrated between measures of depressivity and reward sensitivity, suggesting that instead, reward sensitivity is related primarily to Extraversion. Finally, Study 4 explored Trait Affiliation, an important dimension at the intersection of Agreeableness and Extraversion, and presents a new

Trait Affiliation Scale, along with evidence for its reliability, validity, and practical utility.

Collectively, this work represents a high standard of statistical power and methodological rigor, utilizing a total of eight independent samples ranging from $N = 195$ to $N = 25{,}732$. Across these studies, social cognitive ability and reward sensitivity are further established as important psychological mechanisms underlying individual differences in social functioning. The work presented here also offers methodological contributions and broader theoretical insights into the understanding of personality and its relation to psychopathology. In sum, this dissertation paves the way to a better understanding of how and why individuals vary in our social abilities, interpersonal interactions, and relationship success, in addition to serving as an argument for the broad utility of personality psychology's methods and theories.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BDI-II	Beck Depression Inventory
BFAS	Big Five Aspect Scales
BFI	Big Five Inventory
CAT-PD	Computer Adaptive Test for Personality Disorders (Statistic Form)
CFI	Comparative Fit Index
DN	Default Network
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders (Fourth Edition)
DSM-5	Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition)
ESCS	Eugene-Springfield Community Sample
ESI-BF	Externalizing Spectrum Inventory Brief Form
ESEM	Exploratory Structural Equation Modeling
fMRI	Functional Magnetic Resonance Imaging
FPCN	Frontoparietal Control Network
FSL	Functional Magnetic Resonance Imaging of the Brain Software Library
GPIP	Group Prior Individual Parcellation
HCP	Human Connectome Project
ICA	Independent Components Analysis
IPC	Interpersonal Circumplex
IPIP	International Personality Item Pool
IRI	Interpersonal Reactivity Index
LAVAAN	Latent Variable Analysis (R Package)

NEO-FFI	NEO Five Factor Inventory
NEO PI-R	NEO Personality Inventory, Revised
NIH	National Institutes of Health
PCA	Principal Components Analysis
PID-5	Personality Inventory for DSM-5
RMSEA	Root Mean Square Error of Approximation
SAPA	Synthetic Aperture Personality Assessment
SEM	Structural Equation Modeling
SRMR	Standardized Root Mean Squared Residual
TLI	Tucker Lewis Index
ToM	Theory of Mind

INTRODUCTION

Though humans are universally social, we vary remarkably in our ability and motivation to form and maintain healthy relationships. Why is it that for some individuals, networking, dating, and interviewing come naturally, but for others, getting through a trip to the coffee shop or a phone call with mom can be a struggle? What makes one person charismatic and another socially awkward? This question has been explored by many subfields of psychology—from social and personality to clinical and biological.

One promising approach to this question is seeking to identify the neurocognitive mechanisms underlying social interaction. For example, researchers have identified social cognitive processes—including emotion perception, theory of mind, and empathy—by which we recognize and interpret the mental states of others (Barrett et al., 2011; Gallagher & Frith, 2003; Premack & Woodruff, 1978; Singer & Klimecki, 2014). Underlying neural correlates of these social cognitive processes have also been identified and include brain regions such as the dorsal medial prefrontal cortex and temporoparietal junction (Allen et al., 2017; Andrews-Hanna et al., 2014; Gallese et al., 2004; Saxe & Kanwisher, 2003; Schilbach et al., 2008; Schurz et al., 2014; Spunt & Lieberman, 2012; Vogeley et al., 2001). By understanding potential underlying sources of variation, we might gain a better understanding of why, when it comes to social interaction and relationships, some people fail, and others flourish (DeYoung & Weisberg, 2018).

Despite the prevalence of research on individual differences in interpersonal behavior, however, much of this work occurs in isolated subfields with limited disciplinary crosstalk. Moreover, a large portion of research on social cognition and social neuroscience is lacking in comprehensive theories, statistical power, and attention to psychometrics (Button et al., 2013; Mar et al., 2013; Vul et al., 2009). To advance our understanding of interpersonal functioning, a shift toward methodological rigor and integrative frameworks is essential. One solution is adopting insights from personality psychology—that is, connecting research on social cognition and behavior with models that identify and explain major dimensions of psychological variation.

Personality as a Unifying Framework for Individual Differences Research

A majority of research in personality has focused on the measurement and description of *traits*, which refer to relatively stable patterns of motivation, emotion, cognition, and behavior (DeYoung & Blain, 2020; Fleeson, 2001; McAdams & Pals, 2007; Zillig et al., 2002). The most thoroughly validated and widely used model of personality traits is the *Five Factor Model* or *Big Five*, which describes the major dimensions of covariation among human personality traits—Conscientiousness, Agreeableness, Neuroticism, Openness/Intellect, and Extraversion (Costa & McRae, 1992; DeYoung, 2015; Hofstee et al., 1992; John et al., 2008). Regardless of which specific trait model is used, personality psychology attempts to answer some of the most fundamental questions about people: Why are individuals the way they are? How and why do we differ from one another? And what biological substrates and behavioral outcomes are associated with individual differences in personality? Because the mission of personality psychology is to explain the entire person, there is perhaps no better place to look for a framework that can integrate findings regarding any specific class of individual differences, including variation in social cognition and behavior.

When it comes to understanding individual differences in social functioning, two of the Big Five domains are particularly relevant: Agreeableness and Extraversion. *Agreeableness* describes the tendency to be cooperative and altruistic as opposed to selfish and exploitative. *Extraversion* describes the tendency toward sociability, reward sensitivity, and positive emotionality. Agreeableness and Extraversion predict a variety of interpersonal outcomes and have been linked to individual differences in social cognitive processing and reward sensitivity, respectively (Allen et al., 2017; Lucas et al., 2000; Nettle & Liddle, 2008; Smillie et al., 2012). Recent research has further explored the component parts of these traits, revealing that each of the Big Five can be reliably decomposed into two distinct aspects (DeYoung et al., 2007; Soto & John, 2016).

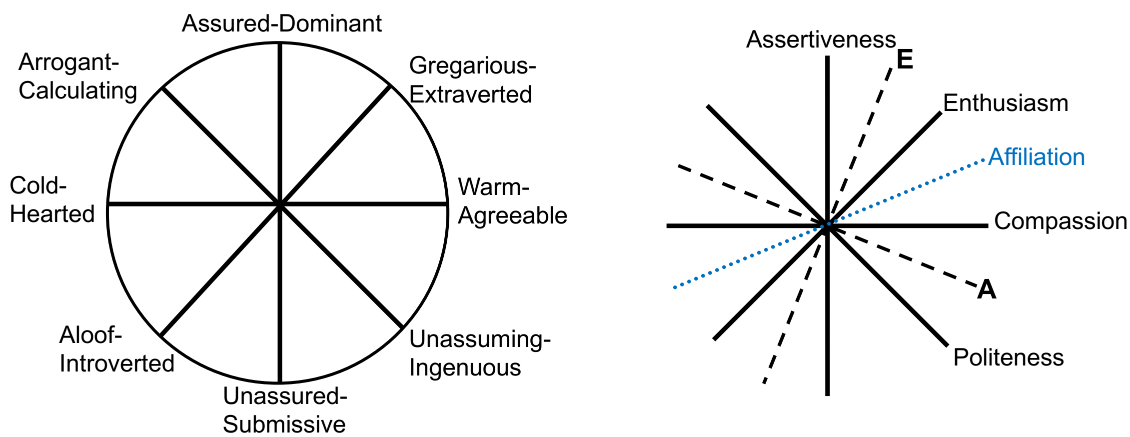
The two aspects within Extraversion are labeled Assertiveness and Enthusiasm, and the two aspects within Agreeableness are labeled Compassion and Politeness (DeYoung et al., 2007). Assertiveness includes tendencies related to leadership, dominance, and drive, whereas Enthusiasm includes both outgoing friendliness or sociability and the tendency to experience and express positive emotion. Compassion reflects empathy, sympathy, and caring for others, whereas Politeness reflects respect for others' needs and desires, as well as a tendency to refrain from aggression. Together, the traits of Extraversion and Agreeableness, along with their component aspects, capture a broad array of individual differences when it comes to social behavior and interpersonal functioning. Moreover, they can be easily united with theoretical frameworks already at the forefront of integrating trait and process approaches to interpersonal theory, such as the *Interpersonal Circumplex* (IPC).

The IPC organizes interpersonal traits and behaviors in relation to two orthogonal dimensions—Status and Love (Gurtman, 2009; Leary, 1957; Wiggins, 1979). The IPC is frequently visualized using a circle; locations along the circumplex can be specified using angular projections between 0º and 360º, each representing a specific interpersonal style that can be conceptualized as a blend of low to high Status and low to high Love. The IPC factor space is also sometimes described in terms of eight subdivisions known as octants (e.g., the Gregarious-Extraverted, Aloof-Introverted, and Unassured-Submissive octants). Since the IPC and Big Five are two of the most used and influential models of individual differences in personality and social behavior, substantial efforts have been taken to unify these systems (Barford et al., 2015; DeYoung et al., 2013; McCrae & Costa, 1989; Pincus, 2002; Wiggins & Pincus, 1994).

Research suggests that Extraversion and Agreeableness describe the same two-dimensional space as the IPC, constituting a slight rotation of the Love and Status axes, with Extraversion and Agreeableness falling near 60° and 330º, respectively (Barford et al., 2015; DeYoung et al., 2013; McCrae & Costa, 1989; Pincus, 2002; Wiggins & Pincus, 1994). Aspects of Extraversion and Agreeableness also map onto the IPC, with Compassion falling at 0º (equivalent to the Love Axis), Enthusiasm at 45º (or the Warm-Gregarious octant), Assertiveness at 90º (equivalent to the Status Axis), and Politeness at 315º (or the Unassuming-Ingenious octant). Another trait particularly relevant to individual differences in social behavior can be represented by the blend of Compassion and Enthusiasm, which falls at 22.5º in the IPC. This particular location on the IPC is associated with Warmth, Affection, and Social Closeness and has been conceptualized as *Trait Affiliation* (DeYoung et al., 2013).

Trait Affiliation is an interstitial facet between Agreeableness and Extraversion, and thus, may stem from variation in both the Agreeableness-related processes of empathy, cooperation, and social cognition, as well as Extraversion-related processes involving reward sensitivity (Depue & Morrone-Strupinsky, 2005; DeYoung, 2015; DeYoung et al., 2013; DeYoung & Weisberg, 2018). A full integration of the IPC and relevant Big Five traits is pictured in Figure 0.1. Together, the Big Five and IPC provide broad coverage of traits relevant to individual differences in social functioning (in terms of trait tendencies and behaviors in a given context), and they can be usefully bootstrapped to forward interpersonal research broadly.

Figure 0.1.



Integration of the Big Five, Interpersonal Circumplex, and Trait Affiliation.

Models such as the IPC and Big Five may also help us understand some of the social deficits seen across a variety of mental disorders, such as antisocial personality, autism, and schizophrenia (Jones et al., 2010; Lozier et al., 2014; Pardini et al., 2003; Pinkham et al., 2008; Sebastian et al., 2012). The major dimensions of psychopathology align with those of normal personality variation and are often associated with similar outcomes (DeYoung & Krueger, 2018; DeYoung et al., 2016; Kotov et al., 2010; Krueger & Markon, 2014; Markon et al., 2005; Widiger, 2011). For example, Antagonism-related symptoms such as aggression, callousness, and deceit are equivalent to pathological low Agreeableness. Likewise, Detachment-related symptoms such as intimacy avoidance, social withdrawal, and anhedonia are equivalent to pathological low Extraversion. Thus, it is highly probable that Detachment and Antagonism share many of the same underlying neurocognitive mechanisms of (low) Extraversion and (low) Agreeableness, respectively. By leveraging frameworks such as the Five Factor Model and IPC and applying these existing models to the study of underlying psychological and neurobiological mechanisms of social functioning, across both normal-range and pathological variation, we can gain a fuller picture of how and why individuals vary in their interpersonal behaviors and outcomes.

The Current Dissertation

In this dissertation, I examine individual differences in social cognition and behavior through the lens of the personality traits Agreeableness and Extraversion, as well as their constituent sub-factors, pathological counterparts, and intersection as the interstitial trait of Affiliation. In doing so, I elucidate the psychological and neurobiological underpinnings of these personality traits and associated individual differences in social functioning. My overarching approach uses latent variable modeling to assess relations among personality traits (in both normal and pathological ranges), psychological processes including social cognition and reward sensitivity, and the function of brain systems such as the default network.

In the chapters that follow, I review a selection of literature from social cognition, social neuroscience, and psychopathology research perspectives, while arguing for the utility of personality psychology's methods and theories. I also present original research examining individual differences in interpersonal functioning from a personality perspective. These studies are organized across four interrelated lines of work—1) determining associations of Agreeableness-Antagonism subfactors with social cognitive ability, 2) examining neural correlates of Agreeableness and social cognitive ability, 3) testing reward sensitivity as an underlying correlate of depressivity and Extraversion, and 4) creating and validating a new Trait Affiliation Scale.

Beginning with Chapter 1, I examine how social cognitive ability is associated with facets of the Agreeableness-Antagonism dimension. Previous research suggests that Agreeableness is associated with better mentalizing—or the ability to understand and interpret the mental states of others—but that this positive association might hold only for specific sub-factors of Agreeableness (Allen et al., 2017; Nettle & Liddle, 2008). In my research, I replicate and extend this work, examining how questionnaire measures of Agreeableness and Antagonism map onto three lower-order sub-factors, as well as how these subfactors differentially predict individual differences in social cognitive ability.

In Chapter 1, using a multi-task design and exploratory structural equation modeling, I find that higher Compassion and Pacificism predict better social cognitive ability, but that higher Honesty predicts worse social cognition. This study extends work on the underlying mechanisms of Agreeableness and its pathological low variants, while also meshing well with research on the potential advantages of Machiavellian intelligence and research highlighting the importance of examining traits at multiple levels of the personality trait hierarchy rather than just the Big Five.

In Chapter 2, I integrate longstanding topics from social neuroscience with perspectives from personality and network neuroscience, using a large, publicly available dataset with both extensive behavioral phenotyping and neuroimaging. Social cognitive abilities have been linked to function of a broad set of brain regions, often collectively referred to as the social brain (Amodio & Frith, 2006; Frith & Frith, 2006; Saxe & Kanwisher, 2003; Saxe & Powell, 2006; Saxe & Wexler, 2005; Schurz et al., 2014; Spunt & Lieberman, 2012; Vogeley et al., 2001; Young et al., 2010). This includes structures such as the dorsal medial prefrontal cortex and temporoparietal junction—regions that are also a part of the brain's so-called default network (Andrews-Hanna et al., 2014; Allen et al., 2017; Mars et al., 2012; Meyer, 2019; Schilbach et al., 2008; 2012; Vogeley et al., 2001).

In Chapter 2, using a Bayesian individualized cortical parcellation approach, combined with structural equation modeling and multiple behavioral tasks, I find that levels of activation in portions of the brain's default network during a social cognition task are positively associated with levels of Agreeableness and social cognitive abilities. I discuss these findings in terms of their implications for our knowledge of social cognition and the default network, while also highlighting broader theoretical and methodological impact for social and personality neuroscience.

Moving to Chapter 3, I pivot from investigating the correlates of Agreeableness to examining the trait of Extraversion. Though Agreeableness and associated interpersonal information processing abilities are certainly essential for relationships, there is also an important motivational component: if a person does not find interactions worthwhile, they are unlikely to have fulfilling relationships, regardless of interpersonal acumen. Extraversion is the personality trait most related to this approach motivation and the desire to affiliate with others, and these tendencies appear collectively underpinned by a broader sensitivity to reward. As blunted reward sensitivity is also a core feature observed in several mental disorders (e.g., depression, schizophrenia, and various personality disorders) measures of reward sensitivity have featured prominently in recent psychiatry and clinical psychology research (Andreasen et al., 2012; Di Nicola, 2013; Pizzagalli et al., 2008a, 2008b; Kwapil & Barrantes-Vidal, 2015; Snaith, 1993).

In Chapter 3, I examine how performance on one commonly used probabilistic reward task is associated with Extraversion and associated variance in depression symptoms. In a large community sample, I find that—in contrast to previous studies with smaller samples—reward sensitivity is not associated with depressivity but is positively associated with Extraversion. These results suggest reward sensitivity—as measured by this task—may be related primarily to Extraversion and its pathological manifestations, rather than to depression *per se*. These findings are consistent with existing models that conceptualize depressive symptoms as combining features of high Neuroticism and low Extraversion. Findings are consistent with theories proposing reward sensitivity as a key mechanism of Extraversion and are discussed in broader contexts of dimensional psychopathology frameworks, replicable science, and behavioral task reliability.

Finally, turning to Chapter 4, I move from focusing on the individual Big Five traits of Agreeableness and Extraversion to examining the importance of their intersection: Trait Affiliation. Trait Affiliation represents the tendency to seek out, develop, and maintain relationships and represents a blend of the Compassion aspect from Agreeableness and the Enthusiasm aspect from Extraversion. Trait Affiliation is easily integrated into other models, such as the IPC, where it is an interstitial trait between the Gregarious-Extraverted and Warm-Agreeable octants (Barford et al., 2015; DeYoung et al., 2013). Despite a variety of Affiliation-related measures existing in the Big Five and IPC factor spaces (e.g., measures of warmth or social closeness), there is currently a lack of questionnaires specifically designed to measure this consequential and interstitial trait (DeYoung et al., 2013). My final dissertation chapter documents the creation and validation of a new ten-item Trait Affiliation Scale.

In Chapter 4, I draw upon several samples and statistical approaches to provide evidence for reliability and validity of my Trait Affiliation Scale. I organize this work into six related sub-studies. Study 4a focuses on scale creation, including item selection and initial evidence of construct validity. Study 4b focuses on the application of item response theory to evaluate item information and create a ten-item scale from an initial set of 24 candidate items. Study 4c provides evidence of structural, convergent, and discriminant validity by examining factor structure, internal consistency, and associations of the 10-item Trait Affiliation Scale with various other personality questionnaires. Study 4d provides evidence of test-retest reliability using a four-wave longitudinal dataset. Study 4e examines evidence of criterion and incremental validity, testing associations of Trait Affiliation with relevant outcome variables (e.g., social goals and social network size) above and beyond Agreeableness, Extraversion, and their aspects. Finally, Study 4e focuses on emotion induction and how the Trait Affiliation Scale predicts both baseline affiliative states and response to affiliative video clips designed to induce warmth, affection, and a desire to bond with others. After presenting my results, I discuss the importance of Affiliation as a trait and provide recommendations for use of this scale in future research.

I end my dissertation by synthesizing the results of these four studies, offering recommendations for future work on interpersonal functioning, and arguing that personality can be used as an integrative framework for research on virtually any topic of psychological inquiry involving individual differences.

CHAPTER 1:

Theory of Mind and the Agreeableness-Antagonism Dimension: Differential Associations with Callousness, Aggression, and Manipulativeness

Social cognitive processes encompass the various skills essential for successfully navigating social interactions (Barrett et al, 2011; Singer & Klimecki, 2014). People vary in their proficiency in these skills, and research in psychopathology has consistently reported social cognitive deficits across a broad range of symptoms and disorders. For example, the social cognitive process known as *theory of mind (ToM)* or *mentalizing*, which refers to a person's ability to recognize, understand, and utilize the thoughts, feelings, and beliefs of other people (Premack & Woodruff, 1978), has been negatively associated with autism spectrum disorders (Baron-Cohen et al., 1986). Given that better mentalizing is associated with increased social competence (Liddle & Nettle, 2006; Jensen-Campbell et al., 2002) and reduced aggression and antisocial behavior (Mohr et al., 2007; Meier et al., 2006), exploring the personality correlates of this social cognitive process could contribute to both basic and clinical psychology research.

Perhaps the most logical place to look for psychopathology research relevant to social cognition and social neuroscience is in relation to constructs and disorders where antisocial behavior and relationship problems are core defining features. This includes research on personality disorders related to high levels of aggression, callousness, and relationship difficulties—namely, antisocial, narcissistic, histrionic, and borderline personalities (American Psychiatric Association, 2013; Krueger & Markon, 2014).

Indeed, those with antisocial or borderline personality disorder perform worse on certain ToM tasks (Dolan & Fullam, 2004; Preißler et al., 2010). In addition to work examining ToM in relation to categorical conceptualizations from the Diagnostic and Statistical Manual of Mental Disorders (DSM; American Psychiatric Association, 2013), there has been related work examining ToM deficits in the context of psychopathy and children with callous unemotional traits. Though broad social cognitive deficits are seen in psychopathy and children with callous unemotional traits, problems with *affective empathy*—or the tendency to vicariously experience the emotional states of others rather than simply understand their mental states—seem to be more prevalent than problems with ToM, and the exact associations between ToM and psychopathy remain unclear (Jones et al., 2010; Shamay-Tsoory et al., 2010; Pardini et al., 2003).

Another useful framework for understanding how antagonistic traits might relate to individual differences in ToM is the dark triad—consisting of the traits Narcissism, Machiavellianism, and Psychopathy—which has been examined in both clinical and non-psychiatric populations (Furnham et al., 2013; Paulhus & Williams, 2002). Research on these traits and their relation to social cognitive functioning is mixed, but, taken as a whole, it suggests that while psychopathy is negatively associated with ToM abilities, Narcissism (characterized by entitlement, grandiosity, and attention seeking) and Machiavellianism (characterized by manipulativeness and deceit) may be unrelated to or even positively associated with individual differences in ToM (Jonason & Krause, 2013; Kajonius & Björkman, 2020; Paal & Bereczkei, 2007; Schimmenti et al., 2019; Stellwagen & Kerig, 2013; Vonk et al., 2015; Wai & Tiliopoulos, 2012). One useful approach to integrating and further explaining ToM findings from the DSM and dark triad perspectives is mapping these constructs onto existing models of individual differences, such as the five factor model of personality and corresponding dimensional psychopathology questionnaires.

Personality and Social Cognition

Certain personality traits have already been linked to adeptness in mentalizing ability. The *Big Five* personality traits capture five broad dimensions of personality that comprehensively organize most personality traits and descriptors. Moreover, the Big Five dimensions are very similar to the dimensions that emerge from patterns of covariation in symptoms of psychopathology—including not only personality disorders but other disorders too (DeYoung & Krueger, 2018; Kotov et al., 2010; Kotov et al., 2017). Many psychiatric symptoms can be described as risky or maladaptive variants of behaviors described by normal personality variation (DeYoung & Krueger, 2018). For instance, maladaptively low Agreeableness has been labeled “Antagonism.” Thus, personality frameworks such as the Big Five may provide useful frameworks for more thoroughly describing and explaining individual differences in mentalizing ability, including the mentalizing deficits seen across a variety of psychopathology symptom dimensions and mental disorders.

Each of the Big Five domains can be broken down into two unique but correlated aspects of personality, which can then be broken down into the many facets of personality that make up that aspect (DeYoung et al., 2007). One Big Five domain in particular—Agreeableness (referred to as Antagonism at its low pole)—has been associated with many of the same correlates as ToM, including social competence and social network size (Liddle & Nettle, 2006; Jensen-Campbell et al., 2002; Meier et al., 2006). Research directly examining the relation between Agreeableness and ToM has also begun to emerge, and there is some evidence for a positive correlation between the two constructs (Nettle & Liddle, 2008). This research, however, did not include a Big Five measure with subscales to assess the finer-grained aspects and facets of Agreeableness and their relation to ToM; moreover, the 10-item Agreeableness scale used by Nettle & Liddle (2008) contains mostly items that are related to the Compassion aspect of Agreeableness and not to the broader construct or its other facets.

As previously stated, Antagonism is synonymous with low levels of Agreeableness, forming an Agreeableness-Antagonism continuum from high to low levels of Agreeableness (Gore & Widiger, 2013; Suzuki et al., 2015). Little research has been dedicated to understanding how Agreeableness and Antagonism—along with their lower-order factors—predict general or social cognitive abilities, with current findings yielding inconsistent results (Paal & Bereczkei, 2007; Lyons et al., 2010). Allen et al. (2017) extended Nettle and Liddle's (2008) investigation of ToM and Agreeableness by examining differential associations between mentalizing and lower-order aspects of Agreeableness, known as *Compassion* and *Politeness*. Allen et al. (2017) found that mentalizing was positively correlated with Compassion and negatively correlated with Politeness. Subsequent analyses of multiple Agreeableness and Antagonism facet-level scales suggested three factors that differentially predicted mentalizing. The Empathy or Compassion subfactor and a Non-aggression or Pacifism subfactor positively predicted mentalizing ability but an honesty subfactor negatively predicted ToM. Nonetheless, these findings regarding subfactors were discovered in *post hoc* analyses and warrant replication using a confirmatory framework and additional measures of ToM ability. In the current research, I attempted to further examine whether ToM ability was related to these three Agreeableness-Antagonism subfactors of Compassion-Callousness, Honesty-Manipulativeness, and Pacifism-Aggression.

Reliability, Multi-task Designs, and Latent Variable Modeling

To justify claims regarding the underlying associations among constructs—for example, Agreeableness and social cognition—we must first be able to assess each of those constructs, individually, in a way that is reliable and valid. Concerns of reliability and validity are especially important when using behavioral tasks, as even tasks that can detect robust effects at the group level (e.g., tests of implicit bias or self-regulation) often fail to produce reliable measurement of individual differences (Hedge et al., 2018; Enkavi et al., 2019a; Schnabel et al., 2008). Fortunately, questionnaire measures of personality and tests of general or social cognitive ability tend to have better reliabilities than many of the measures commonly used in other areas of psychology (Hedge et al., 2018; Morrison et al., 2019; Pinkham et al., 2018; Vellante et al., 2013).

Single-task performance-based indicators are often limited in their scope and measure constructs more narrowly than those they purport to represent (Apperly, 2012; Blain, Longenecker et al., 2020). Performance on any given task is influenced by a number of task-specific factors but using multi-indicator designs allows us to move toward measuring constructs more reliably as what is shared across multiple tasks, thereby avoiding underestimation of true effect sizes (Blain, Longenecker et al., 2020; Campbell & Fiske, 1959; Eisenberg et al., 2019; Enkavi et al., 2019a; 2019b; Nosek & Smythe, 2007).

We can further increase our ability to reliably measure these constructs and estimate their associations with other variables by using latent variable methods, such as structural equation modeling (SEM), which models the prediction of latent variables by other latent variables. Latent variables represent the shared variance of multiple measured (or *manifest*) variables (Schumacker & Lomax, 2004). For example, a latent social cognitive ability variable might be modeled as the shared variance of accuracy scores across different social cognition tasks. Assessing variables of interest at the latent level allows for more robust conclusions, as latent variables capture only the shared variance of their indicators, thereby eliminating error variance and more accurately capturing variability in the underlying constructs of interest (Keith, 2006). Modeling social cognitive ability as the shared variance in performance across tasks should give a better representation of true variance in social cognitive ability by factoring out unique task variance (which includes a combination of task-specific variance and error).

The Current Study

In the current study, I hoped to replicate and extend the findings from Nettle & Liddle (2008) and Allen et al. (2017) by analyzing the relation of ToM ability and individuals' trait levels along the Agreeableness-Antagonism continuum. My approach further breaks down the personality hierarchy to explore how Agreeableness-Antagonism subfactors (i.e., Compassion-Callousness, Pacifism-Aggression, and Honesty-Manipulativeness) relate to ToM.

First, I hypothesized that accuracy scores from multiple tests of ToM ability would be positively correlated with one another and map onto a single latent factor. Mirroring the pattern of findings from Allen et al. (2017), I also hypothesized that when computing an oblique, three-factor solution while factor analyzing multiple self-report measures of Agreeableness and Antagonism, dimensions would emerge that correspond to Compassion-Callousness, Honesty-Manipulativeness, and Pacifism-Aggression. Finally, I hypothesized that ToM accuracy (modeled as a latent variable and as scores from individual tasks) would positively correlate with subfactors for Compassion and Pacifism and negatively correlate with Honesty.

Although the work of Allen et al. (2017) showed support for a three-factor structure of Agreeableness-Antagonism subfactors that differentially predict ToM abilities, their analyses were *post hoc* and warrant replication. Further, this study utilizes a broad battery of Agreeableness-Antagonism facet scales, a multitask design, and an exploratory structural equation modeling (ESEM) analytical approach—all of which are advantages over previous work done on the topic.

Method

Participants

Participants were recruited via a combination of Qualtrics panels and from the campus of the University of Minnesota – Twin Cities as part of a study on social cognition, personality, and psychopathology. No explicit exclusion criteria for psychopathology were implemented in an attempt to sample a broad, representative range of pathological and normal personality variation from the general population. The original sample consisted of 389 individuals, but 54 individuals were excluded for having high amounts of random or invalid responding, leaving a total valid sample of 335. Participants ranged from 18 to 75 in age (*M* = 26.4, *SD* = 13.6). There were 267 females (79.7%), 67 males (20%), and 1 intersex individual (0.3%). In terms of race/ethnicity, 235 participants identified as White or Caucasian (70.1%), 59 as Asian or Pacific Islander (17.6%), 7 as Black or African American (2.1%), 4 as Latino or Hispanic (1.2%), and 30 as multiracial or other (9.0%). 283 participants were native English speakers (84.5%).

Participants reviewed an online informed consent document before beginning the study, then completed an online battery of questionnaires and behavioral tasks, including self-report measures of demographics, personality, psychopathology, and social functioning and several tests of social cognition. All protocols were approved by the University of Minnesota Institutional Review Board (ID# STUDY00003741).

Self-Report Measures

Big Five Aspect Scales

The Big Five Aspect Scales (BFAS; DeYoung et al., 2007) is a 100-item questionnaire that assesses the Big Five personality domains, including two component aspects for each of the Big Five. Each aspect is measured by a total of ten items, including a combination of standard and reverse-coded items. Participants answered each question using a five-point Likert scale ranging from 1 (“Strongly disagree”) to 5

("Strongly agree"). The current study used the two Agreeableness aspects scales, measuring Compassion and Politeness.

Computer Adaptive Test of Personality Disorders: Static Form

The Computer Adaptive Test of Personality Disorders: Static Form (CAT-PD SF; Simms et al., 2011; Wright & Simms, 2014), a selection of 212 items from the 1366-item CAT-PD, is a measure that assesses 33 maladaptive personality traits (e.g., Callousness and Manipulativeness) that can be grouped into five broad categories similar to the Big Five (i.e., Negative Emotionality, Detachment, Antagonism, Disconstraint, and Psychoticism). Participants rated items on a 5-point scale ranging from 1 ("very untrue of me") to 5 ("very true of me"). The current study used the Antagonism-related scales of Callousness, Manipulativeness, Hostile Aggression, and Domineering.

Externalizing Spectrum Inventory Brief Form

The Externalizing Spectrum Inventory Brief Form (ESI-BF; Patrick et al., 2013) is a shortened version of the 415-item ESI. This 160-item questionnaire assesses general Disinhibition, Substance Abuse, Callous Aggression, and 23 lower-order facets of the externalizing spectrum. Participants rated each item on a 4-point scale, with higher scores corresponding to greater agreement with the item. The current study used the lower-order facet scales most strongly correlated with Antagonism and Agreeableness, including Physical Aggression, Relational Aggression, Destructive Aggression, Fraud, Theft, Empathy, and Honesty.

Interpersonal Reactivity Index

The Interpersonal Reactivity Index (IRI; Davis et al., 1980) is a 28-item questionnaire that assesses individual differences in empathy across four dimensions—Empathic Concern, Fantasy, Personal Distress, and Perspective Taking. Each dimension is measured by a 7-item subscale, and participants rated items on a 5-point Likert scale ranging from 0 (“does not describe me well”) to 4 (“describes me very well”), with a combination of standard and reverse-scored items. In the current study, I only used the Empathic Concern scale (i.e., other-oriented emotions centered on helping people in need) because it is the IRI dimension most strongly associated with Agreeableness (Melchers et al., 2016).

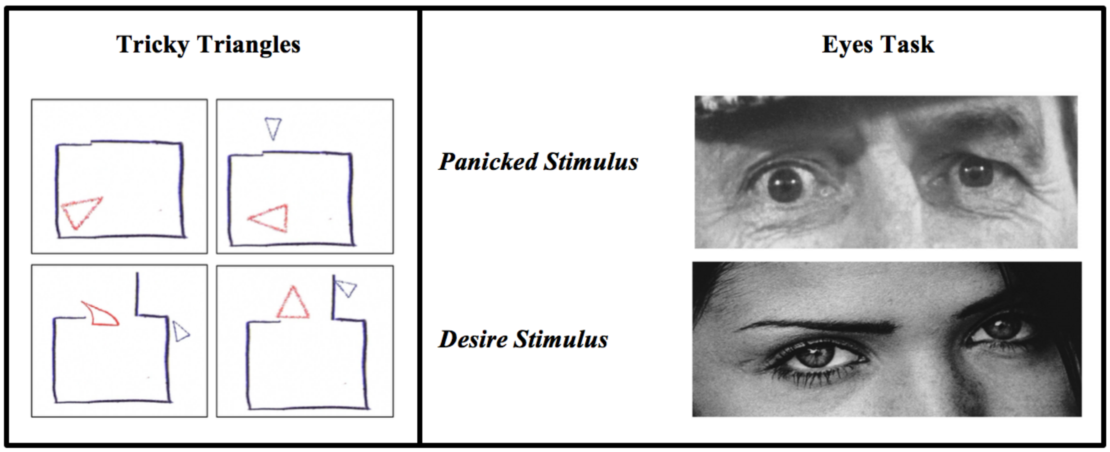
Personality Inventory for DSM-5

The Personality Inventory for DSM-5 (PID-5; Krueger et al., 2012) is a 220-item questionnaire that assesses 25 maladaptive personality trait facets that, like the CAT-PD, can be grouped into five categories (i.e., Antagonism, Detachment, Disinhibition, Negative Affect, and Psychoticism). Facet scales range from 4 to 14 items and are rated on a 4-point scale ranging from 1 (“very false or often false”) to 4 (“very true or often true”). The present study used three of the facets belonging to the Antagonism domain, including Callousness, Deceitfulness, and Manipulativeness.

Theory of Mind Tasks

Participants also completed three tests of social cognition, including a triangle animation task in which participants labeled animations as random, physical, or social (Abell et al., 2000), a mentalizing stories task in which participants had to answer questions about characters’ factual and social knowledge (Stiller & Dunbar, 2007), and a perceptual ToM task using eye stimuli (Baron-Cohen et al., 1997). Example stimuli from the triangles and eyes tasks are pictured in Figure 1.1.

Figure 1.1.



Social cognition tasks from Chapter 1.

Theory of Mind Vignettes

The ToM vignette task (Stiller & Dunbar, 2007) comprises five short stories depicting social situations. Each story describes a social interaction involving multiple characters. Participants read each story, after which they answered five ToM questions and five memory questions pertaining to the story. All questions are in true-false format. Memory questions are designed to measure the participants' ability to retain the factual contents of the story, and the number of facts that the participant must retain varies from two to six in each question. Performance on memory questions within the task can be used as a covariate to ensure that any associations with variables of interest are due to participants' ToM ability rather than their memory for the details of the story. ToM questions required that the participant reason, or infer, a character's perspective in the story. Questions vary across five levels of difficulty, with each successive level requiring the participant to track an additional character or level of perspective. For example, in second-level questions, participants tracked their own mental state and the mental state of one character (e.g., "John wanted to go home after work"). In fourth-level questions, participants tracked the mental state of three characters (e.g., "John thought that Penny knew what Sheila wanted to do"). In order to assess performance on the task, I adopted the procedure used by Nettle and Liddle (2008) and Allen et al. (2017) and computed simple sums of correct responses to memory questions and ToM questions for each participant.

Tricky Triangles Task

In the triangles task (Abell et al., 2000), participants are presented with a series of computerized animations of shapes interacting in a way that was random, physical, or social. In the random condition, the shapes did not interact with each other, but rather moved around purposelessly (e.g., bouncing or drifting). In the physical condition, the shapes moved in a goal-directed manner without invoking ToM or mentalizing (e.g., fishing or swimming). In the social condition, shapes enacted a social sequence, such as coaxing, seducing, or mocking. Participants were tasked with indicating whether each animation was random, physical, or social in nature, then scored for their accuracy in correctly categorizing each animation in a series of 22 clips.

Reading the Mind in the Eyes Task

The eyes task (Baron-Cohen et al., 2001) consists of 36 grey-scale photos of people taken from magazines. These photos were cropped and rescaled so that only the area around the eyes could be seen. Each photo was accompanied by four mental state terms, from which the participant was instructed to choose the word that best described what the person in the photo was thinking or feeling. Only one of the four items was correct (as judged by consensus from an independent panel of judges in the initial psychometric study). Participants were scored for their accuracy across all 36 stimuli.

Analyses

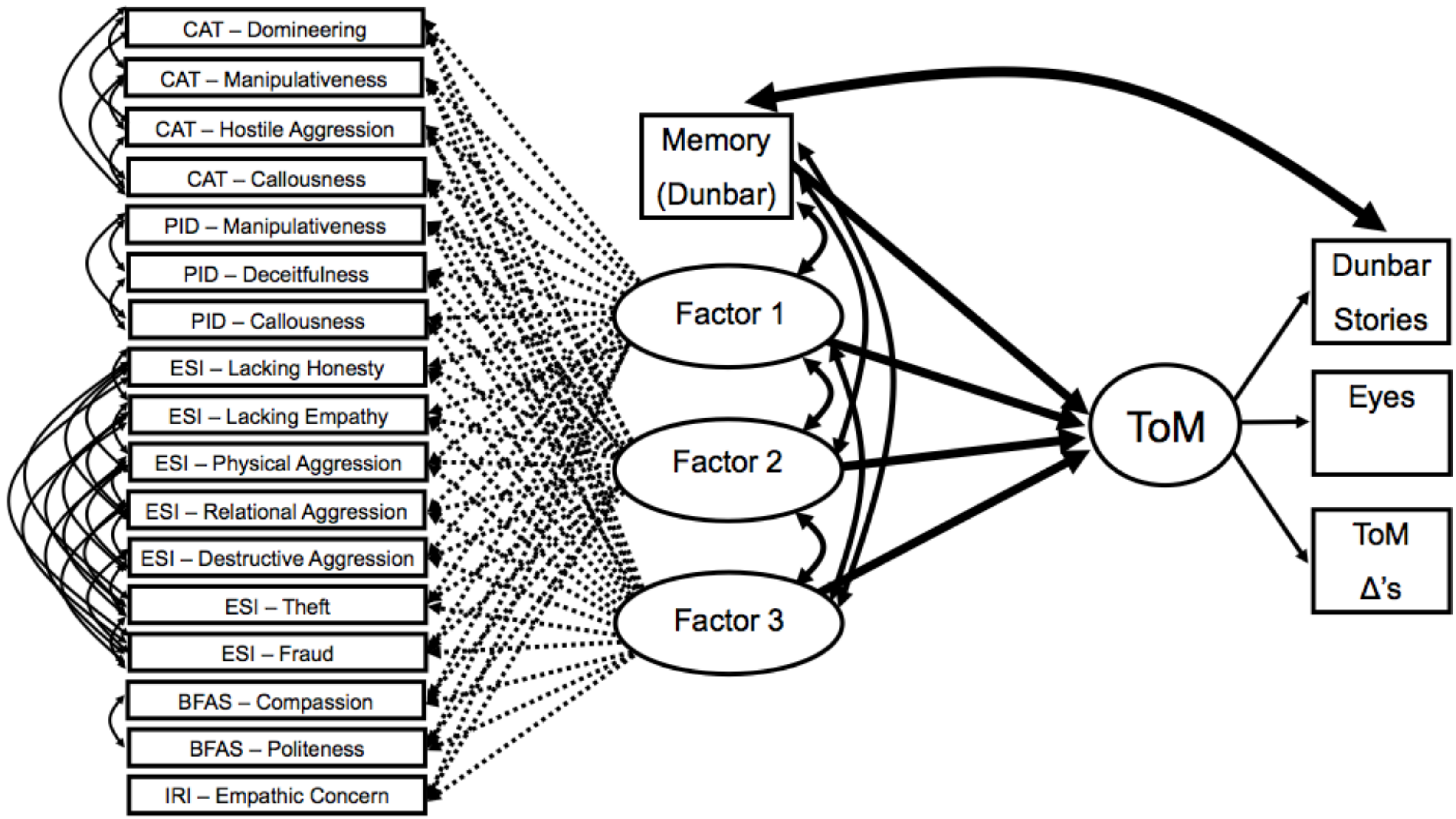
Descriptive statistics were calculated for all task performance and personality variables. Cronbach's α (Cronbach, 1951) and ω_t (McDonald, 1999; Revelle & Condon, 2019) were computed to assess internal consistency reliability. MPLUS was then used for latent variable modeling (Muthén & Muthén, 2017); all models were estimated using full-information, robust maximum likelihood estimation (MLR). First, I computed a single-factor confirmatory factor analytic (CFA) model using accuracy scores from the three ToM tasks to examine how well these tasks represent a single coherent latent variable. To further assess whether a single factor model was appropriate for explaining shared variance across tasks, ω_{t} was also computed.

Next, Exploratory Structural Equation Modeling (ESEM) was implemented. Exploratory Agreeableness-Antagonism subfactors were derived from relevant BFAS, IRI, ESI-BF, PID-5, and CAT-PD SF subscales. First, I conducted a Velicer's minimum average partial (MAP) test (O'Connor, 2000) to see how many factors were empirically suggested in this data. Then, a three-factor solution was computed, given my hypotheses and aims to conceptually replicate and extend the work of Allen et al. (2017). I used factoring with a constrained oblimin rotation (γ = -.80). Relative to traditional structural equation modeling (SEM) with confirmatory factors, computing exploratory factors better accounts for the nontrivial cross-loadings of indicators (Asparouhov & Muthén, 2009). ESEM also allows for more accurate model estimation vs. simple use of observed factor score estimates in SEM. Nonetheless, results for models in the current study were substantively equal if factor score estimates derived using exploratory factor analysis followed by the regression method were used (rather than full ESEM).

Residual covariances of subscales from the same questionnaire (e.g., BFAS Politeness and Compassion) were freely estimated, which resulted in better fit vs. constraining these covariances to zero (ΔS-B χ^2 = 218.3, $p < .001$). Additionally, the residual covariance between memory accuracy and ToM accuracy from the vignette task was freely estimated, resulting in significantly improved fit (ΔS-B χ^2 = 23.4, $p < .001$).

Subsequently, I predicted latent ToM accuracy (shared variance of accuracy across the three tasks) from the Agreeableness-Antagonism factors, allowing predictors to correlate and including performance on the memory questions from the vignette task as an additional, correlated predictor variable. The full ESEM model is presented in Figure 1.2.

Figure 1.2.



Exploratory structural equation model of Agreeableness-Antagonism factors and ToM.

Satorra-Bentler adjusted fit indices were computed and 95% confidence intervals (with standard errors derived using the Huber-White sandwich estimator) were estimated for the path coefficients from predictor variables to latent ToM accuracy (Huber, 1967; Muthén & Muthén, 2017; Satorra, & Bentler, 2001; White, 1980). Finally, for visualization purposes, factor score estimates were computed for Agreeableness-Antagonism and ToM latent variables, using the regression method. ToM factor scores were residualized for scores on the memory condition of the vignette task and standardized, then plotted against standardized factor scores for each of the Agreeableness-Antagonism factors (residualizing for variance in the other two Agreeableness-Antagonism factors).

Results

Descriptive statistics are presented in Table 1.1. Performance was generally high for the behavioral tasks, but variables showed no prominent ceiling effects. Several of the personality and task performance variables showed moderate skewness; thus, analytical methods robust to non-normality were used (i.e., MLR estimation implemented with MPLUS). Values of ω_t and α indicated acceptable internal consistency for the majority of questionnaire and task variables.

Bivariate correlations are presented in Table 1.2. Across measures, Agreeableness variables positively predicted task performance while Antagonism variables negatively predicted task performance. The CFA model showed that accuracy scores on the three ToM tasks loaded onto a single latent variable (Figure 1.3).

Table 1.1. *Descriptive statistics for Chapter 1*

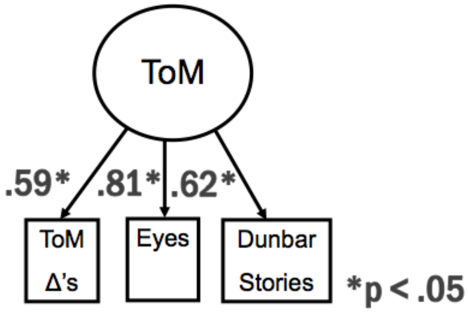
	Mean (SD)	Skewness	[Minimum, Maximum]	α	ω_t
Dunbar ToM	19.4 (3.2)	-1.1	[8, 25]	.66	.77
Dunbar Memory	22.5 (3.7)	-1.6	[10, 26]	.81	.85
Eyes Task	24.3 (5.7)	-0.8	[5, 34]	.77	.80
Triangles	12.6 (3.3)	-0.5	[3, 19]	.62	.68
BFAS Compassion	4.0 (0.6)	-0.4	[2.4, 5.0]	.86	.86
BFAS Politeness	3.8 (0.6)	-0.6	[2.1, 5.0]	.77	.77
IRI Empathic Concern	3.8 (0.7)	-0.2	[1.9, 5.0]	.77	.78
PID Callousness	1.4 (0.5)	2.0	[1.0, 3.8]	.91	.94
PID Deceitfulness	1.6 (0.6)	1.0	[1.0, 3.7]	.88	.89
PID Manipulativeness	1.8 (0.8)	0.6	[1.0, 4.0]	.84	.85
CAT Callousness	1.8 (0.8)	1.1	[1.0, 4.4]	.91	.91
CAT Domineering	2.1 (0.8)	0.7	[1.0, 5.0]	.85	.85
CAT Hostile Aggression	1.6 (0.8)	1.7	[1.0, 5.0]	.93	.93
CAT Manipulativeness	1.7 (0.7)	1.2	[1.0, 4.3]	.87	.87
ESI Theft	1.3 (0.7)	2.1	[1.0, 4.0]	.90	.91
ESI Fraud	1.3 (0.7)	2.2	[1.0, 4.0]	.89	.90
ESI Honesty	3.1 (0.6)	-0.4	[1.0, 4.0]	.76	.77
ESI Physical Aggression	1.4 (0.6)	2.1	[1.0, 4.0]	.91	.92
ESI Destructive Aggression	1.3 (0.5)	2.6	[1.0, 4.0]	.93	.93
ESI Relational Aggression	1.4 (0.6)	1.6	[1.0, 4.0]	.89	.89
ESI Empathy	3.4 (0.5)	-0.9	[1.8, 4.0]	.88	.89

Table 1.2. *Bivariate correlations of Chapter 1 self-report measures and task performance*

Measure	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.
1. Eyes Task	1.00																				
2. Triangles Task	.47	1.00																			
3. Dunbar ToM	.50	.37	1.00																		
4. Dunbar Mem	.53	.40	.72	1.00																	
5. BFAS Compassion	.31	.21	.28	.23	1.00																
6. BFAS Politeness	.23	.19	.21	.22	.58	1.00															
7. IRI Empathic Concern	.35	.30	.28	.27	.63	.49	1.00														
8. PID-5 Callousness	-.51	-.41	-.43	-.49	-.60	-.57	-.58	1.00													
9. PID-5 Manipulativeness	-.32	-.24	-.21	-.35	-.40	-.53	-.40	.67	1.00												
10. PID-5 Deceitfulness	-.36	-.31	-.28	-.38	-.54	-.60	-.50	.78	.80	1.00											
11. CAT-PD Callousness	-.49	-.37	-.39	-.44	-.70	-.58	-.66	.81	.60	.71	1.00										
12. CAT-PD Domineering	-.30	-.25	-.24	-.33	-.52	-.65	-.47	.66	.69	.70	.73	1.00									
13. CAT-PD Hostile Aggression	-.55	-.44	-.47	-.56	-.54	-.59	-.52	.84	.61	.70	.83	.72	1.00								
14. CAT-PD Manipulativeness	-.44	-.40	-.38	-.47	-.55	-.64	-.57	.78	.70	.81	.82	.78	.85	1.00							
15. ESI Theft	-.49	-.41	-.43	-.53	-.39	-.42	-.40	.71	.57	.65	.65	.59	.77	.81	1.00						
16. ESI Fraud	-.51	-.40	-.42	-.54	-.46	-.47	-.44	.76	.62	.72	.69	.63	.77	.76	.87	1.00					
17. ESI Honesty	.17	.26	.18	.21	.27	.33	.47	-.32	-.33	-.46	-.36	-.31	-.34	-.46	-.26	-.29	1.00				
18. ESI Physical Aggression	-.49	-.40	-.41	-.52	-.45	-.45	-.43	.75	.58	.64	.67	.61	.81	.72	.85	.84	-.20	1.00			
19. ESI Destructive Aggression	-.56	-.45	-.46	-.56	-.46	-.46	-.41	.79	.56	.65	.69	.60	.82	.74	.88	.88	-.21	.89	1.00		
20. ESI Relational Aggression	-.43	-.36	-.32	-.42	-.51	-.61	-.48	.77	.67	.76	.75	.71	.79	.81	.80	.85	-.31	.83	.83	1.00	
21. ESI Empathy	.48	.42	.41	.43	.73	.55	.74	-.74	-.53	-.63	-.80	-.63	-.71	-.71	-.61	-,64	.44	-.66	-.65	-.69	1.00

Notes. r ≥ .11 is significant (at α .05). BFAS = Big Five Aspect Scales, IRI = Interpersonal Reactivity Inventory, CAT = Computer Adaptive Test of Personality Disorders, PID-5 = Personality Inventory for DSM-5, ESI = Externalizing Spectrum Inventory.

Figure 1.3.



Measurement model of ToM tasks from Chapter 1.

Factor loadings were moderately high, with performance on the eyes task being the strongest. Fit statistics for all models are presented in Table 1.3. The CFA model was just identified, so meaningful model fit evaluation based on standard fit indices was not possible. Nonetheless, there was evidence that a substantial portion of variance could be explained by a single underlying social cognitive ability factor (ω_t = .71).

Table 1.3. *Fit statistics for Chapter 1 structural equation models*

Models	RMSEA	95% C.I.	SRMR	S-B $\chi 2$	*p*	CFI	TLI
1. ToM	.000	[.000, .000]	.000	0.00	< .001	1.0	1.0
2. ToM, Mem, and Antagonism Factors	.060	[.050, .069]	.023	262.4	< .001	.98	.96

Next, I used ESEM to identify latent factors from Agreeableness and Antagonism scales. Consistent with the notion that Agreeableness can be separated into two correlated aspects and with previous work using a similar set of Agreeableness-Antagonism scales (Allen et al., 2017), conducting a Velicer's MAP test indicated the presence of two factors across the 17 scales. Correlations between these two extracted factors and relevant

BFAS variables showed that the first factor approximated low Politeness ($r = -.66$, $p < .001$) and the second factor strongly resembled Compassion ($r = .79$, $p < .001$).

Since I was interested in parsing subfactors within Politeness and their associations with ToM, I then chose to extract three factors. Factor loadings for the three-factor solution are presented in Table 1.4.

Table 1.4. *Factor loadings of Chapter 1 Agreeableness-Antagonism scales on three exploratory factors*

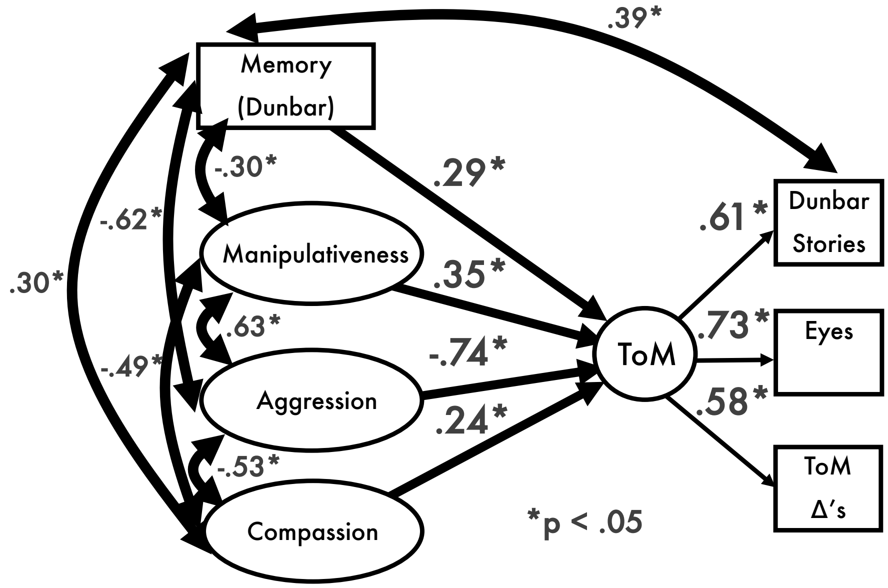
Scale	Aggression	Manipulativeness	Compassion
CAT – Domineering	.21	**.57***	-.18*
CAT – Manipulativeness	**.47***	.46*	-.14*
PID – Manipulativeness	.24	**.60***	-.02
PID – Deceitfulness	.28	**.61***	-.12*
ESI – Honesty	-.02	**-.39***	.16
BFAS – Politeness	-.07	**-.51***	.26*
ESI – Relational Aggression	**.53***	.39*	-.10*
ESI – Physical Aggression	**.82***	.03	-.03
ESI – Destructive Aggression	**.90***	-.03	-.03
ESI – Theft	**.78***	.13	.05
ESI – Fraud	**.70***	.20	-.03
CAT – Hostile Aggression	**.82***	.05	.14*
PID – Callousness	**.66***	.06	.31*
CAT – Callousness	.40*	.16*	**-.52***
ESI – Empathy	-.29*	-.04	**.70***
BFAS – Compassion	-.02	-.08	**.78***
IRI – Empathic Concern	-.05	-.10	**.71***

Note. *$p < .05$ (based on the z-distribution and standard errors computed using the Huber-White sandwich estimator); bolded values indicate the factor for which the given scale had the largest loading.

Based on their content and on previous research, I labeled these three factors Compassion-Callousness, Pacifism-Aggression, and Honesty-Manipulativeness. The first factor corresponded to Honesty-Manipulativeness, showing the strongest positive loadings for PID-5 Deceitfulness and Manipulativeness, CAT-PD Domineering and Manipulativeness, and Relational Aggression, as well as negative loadings for BFAS Politeness and ESI Honesty. The second factor corresponded to Pacifism-Aggression, having strong positive loadings for the ESI aggression subscales and various relevant scales from the CAT-PD SF and PID-5. A final factor corresponded to Compassion-Callousness, with strong negative loadings for the BFAS Compassion and IRI empathic concern scales, as well as positive loadings for the ESI Empathy and CAT-PD SF and PID-5 Callousness scales. Significant cross-loadings were relatively common, across all factors.

Subsequently, I examined the effects of the Aggression, Callousness, and Manipulativeness factors on ToM accuracy across tasks. Results of the full structural model and ToM measurement model are displayed in Figure 1.4. (The full measurement model for Agreeableness-Antagonism factors is not displayed here, due to visual complexity). Residual correlations accounting for shared instrument variance are presented in Table 1.5.

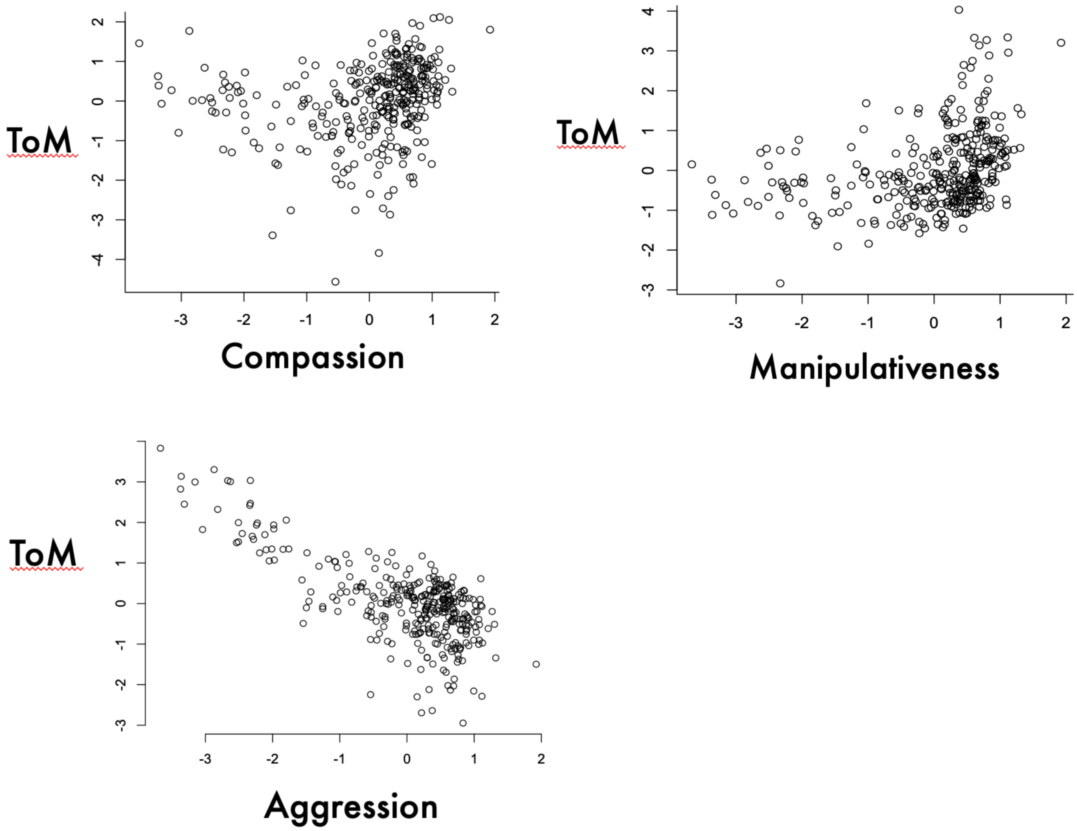
Figure 1.4.



Model of Agreeableness-Antagonism factors and ToM.

I found that ToM was negatively predicted by Aggression (95% CI β: [-.96, -.52]) and positively by Compassion (95% CI β: [.08, .41]) as well as Manipulativeness (95% CI β: [.17., .52]). Memory was a positive predictor of ToM (95% CI β: [.12, .46]). Model fit was reasonable, as indicated by an SRMR ≤ .08, RMSEA ≤ .06, and TLI/CFI ≥ .95 (Hu & Bentler, 1999). Similar results were obtained whether or not memory was included as a covariate. Moreover, specifications that did not freely estimate residual covariances (of subscales from the same questionnaire and of memory and ToM scores from the stories task) yielded results that were substantively equivalent to those reported here. Associations among residualized factor score estimates are visualized in Figure 1.5.

Figure 1.5.



Scatterplots of ToM and Agreeableness-Antagonism factor score estimates.

Table 1.5. *Residual correlations accounting for shared instrument variance in Chapter 1 models of Agreeableness-Antagonism factors and ToM*

Measure	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.
1. Dunbar ToM		—																	
2. Dunbar Mem	.39*																		
3. BFAS Compassion				.24*															
4. BFAS Politeness			.24*																
5. IRI Empathic Concern																			
6. PID-5 Callousness							.29*	.38*											
7. PID-5 Manipulativeness						.29*		.39*											
8. PID-5 Deceitfulness						.37*	.39*												
9. CAT-PD Callousness										.22*	.34*	.29*							
10. CAT-PD Domineering									.22*		.33*	.16							
11. CAT-PD Hostile Aggression									.34*	.33*		.30*							
12. CAT-PD Manipulativeness									.29*	.16	.30*								
13. ESI Theft														.55*	.11	.47*	.56*	.37*	-.21
14. ESI Fraud													.55*		-.13	.44*	.57*	.44*	-.11
15. ESI Honesty													.11	.13		.25*	.27*	.23*	.21*
16. ESI Physical Aggression													.47*	.44*	.25*		.54*	.49*	-.22*
17. ESI Destructive Aggression													.56*	.57*	.27*	.54*		.47*	-.10
18. ESI Relational Aggression													.37*	.44*	.23*	.49*	.47*		-.16
19. ESI Empathy													-.21	-.11	.21*	-.22*	-.10	-.16	

Notes. *$p < .05$. Residual correlations from Model 2 (latent ToM) are displayed below the diagonal and those from Model 3 (individual tasks) are displayed above. BFAS = Big Five Aspect Scales, IRI = Interpersonal Reactivity Inventory, CAT = Computer Adaptive Test of Personality Disorders Static Form, PID-5 = Personality Inventory for DSM-5, ESI = Externalizing Spectrum Inventory

Discussion

The current study investigated how subfactors of the Agreeableness-Antagonism dimension are associated with individual differences in ToM ability. Specifically, I sought to replicate and extend the work of Allen et al. (2017) using multiple behavioral tasks, a more extensive assessment of normal-range and pathological personality traits, and an ESEM approach. Using ESEM, I derived a three-factor structure for a selection of Agreeableness-Antagonism scales; in line with theory and previous work, I labeled these factors Compassion-Callousness, Pacifism-Aggression, and Honesty-Manipulativeness. Results showed that while Compassion-Callousness and Pacifism-Aggression predicted ToM ability in similar directions, the pattern of association diverged for Honesty-Manipulativeness. This replication is particularly worthwhile given that results for these three Agreeableness-Antagonism factors and their association with ToM in Allen et al. (2017) were found in *post hoc* analyses and using a single behavioral task.

As I included two of the tasks used by Nettle and Liddle (2009) in their original study of ToM and Agreeableness, the current work can also be viewed as an extension of that research. Current findings diverge from results reported by Nettle and Liddle because their study found an association only between Agreeableness and ToM as measured by the mentalizing vignettes task, not the eyes task; the current results suggest associations between Agreeableness-Antagonism and ToM may extend beyond just "social-cognitive" ToM (as reported by Nettle and Liddle), to also encompass individual differences in "social-perceptual" ToM. Moreover, similar to Allen et al. (2017), I found that associations between ToM and Agreeableness were specific to certain subfactors;

this suggests Nettle and Liddle's finding of an association between ToM and Agreeableness may have arisen from using a measure highly weighted toward the Compassion aspect. In addition to successfully replicating and extending previous work, the current study extends our understanding of Agreeableness-Antagonism and has broader implications for personality and individual differences research.

Understanding Agreeableness and Antagonism

Human beings are inherently social animals, and therefore, we must work with others to achieve many of our own goals, from finding a partner, to getting a work-related promotion, to fostering rewarding friendships, or raising a family. Though related to numerous psychological processes and personality traits, individual differences in interpersonal functioning seem particularly linked to social cognitive abilities and dispositional tendencies toward Agreeableness vs. Antagonism. It has now been shown in two samples that individual differences in social cognitive abilities are positively associated with Compassion and Pacifism but *negatively* related to Honesty, suggesting that better mentalizing abilities may enable individuals to be more compassionate and less aggressive, but also more deceitful and manipulative.

As there is a robust history of research relating Agreeableness—and particularly its Compassion aspect—to questionnaire measures of trait empathy (del Barrio et al., 2004; Graziano et al., 2007; Mooradian et al., 2008; 2011; Penner et al., 1995), it is unsurprising that the compassion subfactor was positively related to ToM ability. The Compassion-Callousness subfactor in the current study showed strong negative loadings for IRI Empathic Concern and BFAS Compassion, as well as strong positive loadings for

ESI (Lacking) Empathy and CAT-PD Callousness. My finding of a negative association between the callousness factor and ToM ability, paired with the previous work of Allen et al. (2017) and Nettle and Liddle (2009), complements research showing that ToM, trait empathy, and Compassion are associated with positive real-world social outcomes (Cassidy et al., 2003; Devine et al., 2016; Stiller & Dunbar, 2007; Sun et al., 2017). It is possible that individual differences in the ability to accurately perceive, decipher, and react to the mental states of others is one core psychological mechanism underlying Compassion and prosocial behavior. Future work should further explore the relations between social cognition and Compassion by incorporating a broader array of behavioral tasks that tap into additional components of social cognitive ability: for instance, tasks capturing additional components of empathy such as emotion contagion or affective resonance (Zaki et al., 2008). It is possible that performance on these tasks would be even more strongly related to Compassion than tasks like those in the current study, which primarily assess cognitive and perceptual aspects of social processing.

In addition to showing a positive association with Compassion, ToM ability was also positively associated with individual differences in the pacifism (or non-aggression) subfactor. The pacifism-aggression subfactor in the current study showed strong positive loadings for several ESI subscales related to aggression, as well as the CAT-PD Hostile Aggression scale. Counterintuitively, the factor loading of PID-5 Callousness was actually higher for the Pacifism-Aggression factor than the Compassion-Callousness factor; this is likely due to the large number of aggression items in that subscale and is consistent with the fact that the current PID-5 Callousness scale was originally designed to measure two separate facets (callousness and aggression) but was eventually collapsed into a single scale based on analyses using item response theory (DeYoung et al., 2016; Krueger et al., 2012). Somewhat unexpectedly, the Politeness scale from the BFAS did not significantly load onto the aggression subfactor (but instead only loaded on the manipulativeness subfactor). The finding of a negative association between aggression and ToM result replicates a *post hoc* finding from Allen et al. (2017) and is consistent with other work documenting negative associations between aggression and ToM ability (Meier et al., 2006; Mohr et al., 2007). These results suggest that facets of Agreeableness other than just Compassion (and trait empathy) are positively associated with individual differences in ToM ability. Future work could better determine to what extent the specific psychological mechanisms linking lower aggression to better ToM ability are similar to or different from those underlying the association between ToM and Compassion.

The only Agreeableness-Antagonism subfactor showing a diverging pattern of association with ToM ability was Honesty-Manipulativeness. The manipulativeness subfactor was marked by positive loadings for PID-5 Manipulativeness and Deceitfulness, CAT-PD Manipulativeness and Domineering, and ESI (lacking) Honesty, as well as a negative loading for BFAS Politeness. Paired with the original finding in Allen et al. (2017), the current replication of a positive association between manipulativeness and ToM provides further empirical support for the longstanding notion that being able to successfully persuade, manipulate, and deceive others is partially reliant on the ability to understand the thoughts and emotions of others (Byrne, 1996; Byrne & Whiten, 1994; Ding et al., 2015; Lonigro et al., 2014; Slaughter et al., 2013;

Talwar et al., 2007). This suggests that specific traits often conceptualized as pathological (i.e., dishonesty and manipulativeness) may actually be associated with enhanced abilities and outcomes in some circumstances, consistent with how many researchers have discussed potential advantages of Machiavellianism (Byrne & Whiten, 1994; Furnham & Treglown, 2021; Grover & Furnham, 2021).

It is worth noting that the current results contrast with a few studies that found a negative association or no association between ToM ability and Machiavellianism (e.g., Ali & Chamorro-Premuzic, 2010; Lyons et al., 2010). Nonetheless, scales used to assess Machiavellianism in these studies tend to conflate dishonesty and manipulation with other facets of Antagonism, such as immorality or mistrust, indicating that the Honesty-Manipulativeness dimension as construed in the current work and by Allen at al. (2017) may, indeed, be associated with patterns of ToM that diverge from associations with the broader Agreeableness-Antagonism domain. Moreover, studies that do not control for subfactors predicting ToM abilities in opposite directions may have true effects in either direction masked due to statistical suppression (Martinez Gutierrez & Cribbie, 2021; Tzelgov & Henik, 1991).

In addition to unpacking the associations between Agreeableness-Antagonism and ToM ability, it is worth briefly discussing additional factors that might influence the relations among personality, ToM, and social behavior. For instance, one key determinant in lying behavior is the potential cost of lying, with costs taking the form of either tangible punishments or a blow to one's reputation and self-image. Lying is less likely when experimenters can verify the veracity of the participant's behavior, or as monetary punishments increase (Gneezy et al., 2018; Laske et al., 2018). Individuals also limit the extent of their lying to preserve their self-image (Mazar et al., 2008). Importantly, the aversive costs of lying diminish with repeated exposure, which may explain why lying escalates in certain contexts, leading to a "slippery slope" of dishonesty (Garrett et al., 2016). Another potentially influential factor in the associations between Agreeableness-Antagonism subfactors and mentalizing is motivation to engage in mentalizing and associated prosocial behavior. For instance, some research has suggested those with elevated dark triad traits possess the capacity and ability to engage in ToM-related processes, but lack the disposition—and, in most situations, the motivation—to do so (Kajonius & Björkman, 2020).

One final possibility is that superior mentalizing fosters dishonesty when individuals have something to gain from a less powerful counterpart. Indeed, powerful individuals are more likely to lie or exaggerate at the expense of others (Swanner & Beike, 2015; Lammers & Burgmer, 2018). High Antagonism also predicts transgressions against forgiving partners (McNulty & Russell, 2016), and dishonesty is associated with exploitation when one party has more power than the other (Barends et al., 2019). Thus, individual differences in perceived power may influence whether mentalizing capacities are used to deceive and manipulate or to empathize and cooperate. Future research should more thoroughly examine situational, motivational, and relationship-specific factors to determine their potential moderating role in the associations between personality and social cognitive abilities.

Additional Implications for Individual Differences Research

Individual differences in social cognition are conceptually similar to several other constructs that exist in disparate lines of work—for instance emotional intelligence and Gardner's interpersonal intelligence (Gardner, 2011; McEnrue & Groves, 2006). In recent years, emotional intelligence has gained widespread popularity as a construct in both popular culture and among researchers, with some arguing emotional intelligence is predictive of broad positive life outcomes and should be used to inform hiring decisions (Bar-On, 2001; Emanuel & Gudbranson, 2018; Fox & Spector, 2000; Goleman, 1996; Stein & Book, 2011; Watkin, 2000; van der Linden et al., 2010; 2012; 2017). In contrast, others have argued that measures of emotional intelligence (particularly those relying only on self-report) provide little incremental validity over qualities such as general intelligence and Conscientiousness, when it comes to predicting important occupational and educational outcomes (Amelang & Steinmayr, 2006; Antonakis, 2004; Gottfredson, 1997; Landy, 2005; Ones et al., 2012; Schmidt et al., 2008; Thorndike & Stein, 1937Van Rooy & Viswesvaran, 2004; Willoughby & Boutwell, 2018).

Despite legitimate concerns regarding the incremental validity of social cognition or emotional intelligence assessments, even when they are highly correlated with measures of general intelligence, several task-based measures of social cognitive ability do appear to show unique associations with socially relevant personality traits and outcomes (Allen et al., 2017; Stiller & Dunbar, 2007). This appears particularly true for tasks that show acceptable internal consistency and interindividual performance variability—such as the eyes task and the mentalizing vignettes; other measures like the triangle task appear to have more attenuated associations with interpersonal traits and might be best used in younger populations or in conjunction with neuroimaging techniques like functional MRI. Given the results of the current study and presence of stronger effects compared to previous large sample research on ToM and associated outcomes, it also seems likely that utilization of latent variable modeling may be useful when evaluating associations between social cognitive ability and related traits or outcomes. Moving forward, continued attention to the psychometric properties of social cognitive ability measures and their discriminant validity with measures of general intelligence is essential.

The current study provides insights into the theory and measurement of social cognition and its association with Agreeableness-Antagonism subfactors, a key step in better characterizing possible mechanisms and risk factors related to social cognitive deficits. Because problems with ToM and related interpersonal outcomes are characteristic of multiple mental disorders and symptom domains, elucidating their association with normal-range personality traits and improving their measurement may eventually help facilitate more effective methods for assessment and treatment. Such an approach is in line with the Hierarchical Taxonomy of Psychopathology's conceptualization of psychiatric illness, the National Institute of Mental Health's (NIMH's) Research Domain Criteria (RDoC) initiative, and theories of psychopathology that emphasize continuity with normal personality variation and impairments in cybernetic functioning (DeYoung & Krueger, 2018; Insel et al., 2010; Kotov et al., 2017). Future work on this topic could incorporate additional tasks to span a range of social cognitive and interpersonal abilities, including more of those recommended by the NIMH's Workgroup on Tasks and Measures for RDoC (Barch et al., 2016). Furthermore, it remains to be seen whether the personality correlates of social cognitive abilities in populations with more extreme levels of Antagonism (i.e., criminal offenders or those diagnosed with antisocial or narcissistic personality disorder) would reflect the patterns observed in the general population. Finally, another topic worth exploring further is whether the personality correlates and mechanisms of social cognitive deficits are consistent or divergent across different disorders; in particular, research should explore whether ToM deficits seen in those with schizophrenia or autism are similar in etiology and mechanisms to those associated with Antagonism and related personality disorders.

Limitations

Though the current study had numerous strengths, there are a few limitations worth noting. First, the current sample had an overrepresentation of females and people of European and Asian ancestry; future work should attempt to collect more demographically representative samples. The current study's measures of Agreeableness-Antagonism were self-reported and could be usefully supplemented in future research by peer reports or clinician ratings. Also, this work would have further benefited from the inclusion of a general intelligence measure such as the Wechsler Adult Intelligence Scale or International Cognitive Ability Resource (Condon & Revelle, 2014; Wechsler, 2008), as including such measures would allow us to better parse associations between personality and social cognition without the influence of general cognitive ability; such measures could also allow researchers to directly test how individual differences in general cognitive ability might be associated with Agreeableness-Antagonism and its constituent subfactors. These limitations could be usefully addressed by recruiting additional large samples with extensive, high-quality measures of personality, social cognition, general cognition, and real-world social functioning including peer-report and experience sampling data.

Additionally, although the current set of questionnaires captured a broad range of Agreeableness-Antagonism facets, future research could also incorporate measures from the HEXACO and Dark Triad literatures (Collison et al., 2018; Jonason & Webster, 2010; Jones & Paulhus, 2014; Lee & Ashton, 2004; Ashton & Lee, 2007). This approach could directly test whether the constructs represented in those measures (i.e., Honesty-Humility and Machiavellianism) map onto the Honesty-Manipulativeness dimension revealed in the current study using measures based on the Five Factor Model, and whether Honesty-Humility and Machiavellianism show corresponding associations with ToM. Future work could also intentionally recruit participants with clinically significant levels of Antagonism and/or other traits related to social cognitive deficits; this could be a particularly fruitful line of work if the tools used to assess social cognitive ability and personality domains such as Agreeableness could eventually be used to foster more effective early detection and intervention for forms of psychopathology related to interpersonal dysfunction.

Despite the potential of latent variables comprising multiple task or questionnaire indicators to capture truer estimates of effect size, compared with models using only individual manifest variables, this method also runs the risk of inflating effect sizes beyond their population values, especially when indicators share a relatively small portion of variance. Thus, it is important to interpret results of the current study in terms of the general trends seen across models, indicators, and traits, rather than merely focusing on effect sizes from any particular structural model.

Chapter 1 Conclusion

Agreeableness-Antagonism is robustly related to life outcomes, including victimization, relationship satisfaction, aggression, and a variety of psychiatric disorders (Gore & Widiger, 2013; Lynam & Miller, 2019). Despite its enormous consequences however, Agreeableness-Antagonism is arguably the least studied dimension of the Big Five and their pathological counterparts (Gore & Widiger, 2013; Lynam & Miller, 2019). The current research improves the scientific understanding of Agreeableness-Antagonism, replicating and extending work that suggests differential relations of Agreeableness-Antagonism subfactors with social cognitive ability. My findings suggest ToM abilities might facilitate individual differences in most traits related to Agreeableness, with a distinctly negative association with specific honesty-related tendencies. This paradox adds to a set of interesting similar patterns where the correlates or outcomes of personality traits diverge at levels below the Big Five, further underscoring the importance of facet-level research and parsing the subfactors of broad personality domains.

CHAPTER 2:

Activation of the Default Network During a Theory of Mind Task Predicts Individual Differences in Agreeableness and Social Cognitive Ability

As highly social animals, humans are tasked daily with navigating complex social interactions. In order to succeed in these interactions, we often rely upon social cognitive processes that allow us to understand the targets of social interaction, including the ability to perceive and empathize with others' emotions (Barrett et al., 2011; Singer & Klimecki, 2014). In particular, *theory of mind* (ToM), or *mentalization*, describes peoples' ability to recognize and understand the mental states of other people (Gallagher & Frith, 2003). ToM has been positively correlated with social competence (Bosacki & Wilde Astington, 1999; Liddle & Nettle, 2006) and negatively associated with aggressive tendencies (Meier et al., 2006; Mohr et al., 2007). ToM is also positively correlated with social cooperativeness (Paal & Bereczkei, 2007) and social network size, such that individuals with better ToM abilities reported larger networks of friends compared to individuals with less-developed ToM abilities (Liddle & Nettle, 2006; Stiller & Dunbar, 2007). Collectively, these findings suggest the importance of ToM in facilitating successful social interactions.

Although nearly all people are capable of demonstrating ToM to some degree, research on individual differences in ToM can show how variations in ToM ability influence real-world social outcomes. For example, poor ToM performance has been a predictor in mental health research; people with autism (Baron-Cohen et al., 1985;

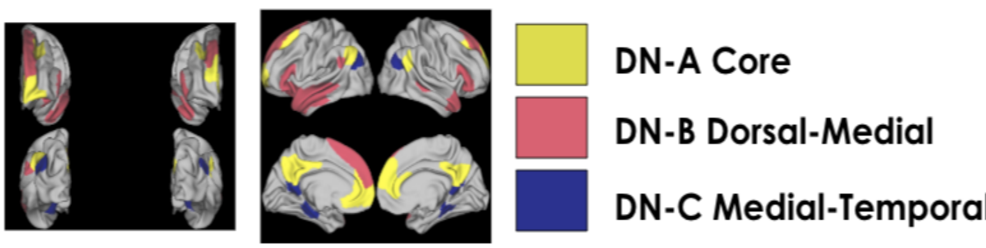
Pinkham et al., 2008) and schizophrenia (Harrington et al., 2005; Pedersen et al., 2012; Pinkham et al., 2008)—two disorders linked with social impairments such as the inability to form appropriate relationships with others (Sasamoto et al., 2011)—have been found to have poor ToM ability. Poor ToM ability has also been correlated with a lack of understanding of how one's actions affect other people, and with difficulty in accurately assessing others' intentions (Baker, 2003).

The current literature on individual differences in ToM ability is limited, and what does exist is largely siloed into subfields such as developmental or clinical psychology. A large portion of such work focuses on variations in the development of ToM among children (e.g., Bowman et al., 2017; Devine & Hughes, 2013; Wang et al., 2016) using false-belief tasks that do not capture the full complexity of ToM (Altschuler et al., 2018; Apperly, 2012; Tager-Flusberg, 2011). Much of the remaining research focuses on ToM deficits in those with mental illness or developmental disabilities (e.g., Baron-Cohen et al., 1985; Dahlgren et al., 2010; Kerr et al., 2003; Pedersen et al., 2012), but such studies are often limited in their sample size and methods of assessing ToM ability. Additionally, a majority of research on ToM uses only single tasks to measure this ability, likely failing to capture the breadth of skills and abilities encompassed by ToM (Altschuler et al., 2018; Apperly, 2012). Considering the scarce research on variation among healthy adults, further research on individual differences in ToM ability could help elucidate the underlying causes for variation in ToM and how this variation might relate to functional outcomes in daily life, eventually paving the way for more effective identification and intervention for those with deficits in these abilities (regardless of whether they meet criteria for specific psychological or developmental disorders).

Theory of Mind and the Default Network

Investigating the neural substrates of individual differences in ToM is one potentially useful approach to increasing scientific understanding of ToM. ToM has been consistently linked to brain regions such as the dorsal medial prefrontal cortex (dmPFC) and temporoparietal junction (TPJ; Allen et al., 2017; Amodio & Frith, 2006; Carrington & Bailey, 2009; Frith & Frith, 2006; Sabbagh et al., 2004; Saxe & Kanwisher, 2003; Saxe & Powell, 2006; Saxe & Wexler, 2005; Schurz et al., 2014; Spunt & Lieberman, 2012; Vogeley et al., 2001; Young et al., 2010). These regions of the brain are included in what is called the default network. Originally conceptualized as the brain's default mode because it was found to be more active when people were merely at rest in the MRI scanner rather than engaged in tasks involving outwardly-directed attention, the default network is now thought of as a network broadly involved in internal simulation, and tasks have been identified that activate it specifically (Andrews-Hanna et al., 2014; Smith et al., 2009). Functions of the default network appear to include simulating the mental states of others and simulating one's own experience during memory, prospection, or fantasy (Allen et al., 2017; Blain, Grazioplene, et al., 2020; Nettle & Liddle, 2008; Mars et al., 2012; Meyer, 2019; Schilbach et al., 2008; 2012; Schurz et al., 2014; Seitz et al., 2006; Tamir et al., 2016). It is worth noting that the so-called "social brain" encompasses additional regions beyond the default network that are involved in social cognition and social interaction (Adolphs, 2009; Brothers, 1990; Frith & Frith, 2010). Other structures, such as the anterior cingulate cortex and insula, appear to be more involved in affective empathy and processes that allow us to vicariously experience and detect the personal relevance of others' emotions (Jackson et al., 2006; Bernhardt & Singer, 2012). For the purposes of this study, however, I focused my investigations on the default network, which has been most widely implicated in research on the neural correlates of various kinds of social cognition, and especially theory of mind (Andrews-Hanna et al., 2014; Buckner et al., 2008; Mars et al., 2012; Meyer, 2019; Schilbach et al., 2008; 2012).

Figure 2.1.



Default Network Subsystems in the Human Connectome Project. DN = Default Network. The above figure displays the three default network subsystems identified by Yeo et al. (2011).

Three subsystems of the default network have been identified: The core, dorsal medial, and medial temporal subsystems (Figure 2.1; Andrews-Hanna et al., 2014; Yeo et al., 2011). Though integrated into a single larger functional network, these three subsystems show some degree of functional specialization, with the dorsal medial subsystem exemplifying the strongest specific associations with ToM tasks (Allen et al., 2017; Buckner et al., 2008; Spreng & Andrews-Hanna, 2015). Nonetheless, the broader default network appears to be important for social cognition and ToM. For instance, the core subsystem of the default network, though primarily active when dealing with personally relevant information and self-related processes, includes several brain regions that have been more specifically implicated in social cognitive processing, including the posterior cingulate cortex (PCC), anterior medial prefrontal cortex (amPFC), and the angular gyrus (Andrews-Hanna et al., 2014; Hyatt et al., 2015; Yeo et al., 2011). Perhaps least related to ToM and social cognitive processing, the default network's medial temporal subsystem is typically associated with autobiographical thoughts and memories, though it has also shown links to the overall default network functions of mental simulation and imagination (Spreng & Andrews-Hanna, 2015). As mentioned, the dorsal medial subsystem is most closely associated with social cognition and contains some of the regions most studied in research on the social brain (i.e., the dmPFC and TPJ), but it is worth mentioning that this subsystem also appears to play an important role in language comprehension generally (Spreng & Andrews-Hanna, 2015). Given existing research on the role of the default network and its subsystems, we can expect that the default network broadly, but the dorsal medial subsystem in particular, will be active during social processing tasks and that individual differences in the function of these networks might underpin individual differences in social cognitive abilities.

Theory of Mind and Agreeableness

When studying individual differences, it is useful to make connections with the broad personality models that attempt to identify the major domains of psychological variation, such as the Five Factor Model (John et al., 2008). One of the Big Five, Agreeableness, which describes traits related to altruism and cooperation, has been associated with variations in ToM ability (Allen et al., 2017; Nettle & Liddle, 2008) and thus provides a particularly useful context for understanding individual differences in social cognitive abilities and associated neural networks (Allen et al., 2017; Laursen et al., 2002; Krueger et al., 2012). Agreeableness has been shown to correlate positively with many of the same beneficial social outcomes as ToM ability (Allen et al., 2017; Ozer & Benet-Martinez, 2006), indicating the importance of further exploring the relation between ToM and Agreeableness.

As Agreeableness is one of the traits most strongly related to individual differences in interpersonal behavior (DeYoung et al., 2013; Graziano & Eisenberg, 1997; Koole et al., 2001), better elucidation of Agreeableness and its associated cognitive mechanisms (including social cognitive processes such as ToM) could allow us to better predict and understand variation in interpersonal behavior and relationship functioning. This research also has the potential to contribute to theoretical models of personality. Until recently, much more emphasis has been placed on the characterization, rather than explanation, of variation in personality, and this is particularly true for Agreeableness (Nettle & Liddle, 2008). Examining ToM and social cognitive ability as one potential correlate of variation in Agreeableness, and examining the relation of both constructs to underlying variation in the default network, would contribute to neurocognitive accounts attempting to explain individual differences in Agreeableness and associated interpersonal outcomes (e.g., Allen & DeYoung, 2017; DeYoung & Weisberg, 2018; Xiao et al., 2019).

Utility of Latent Variable Modeling

To accurately determine the associations among constructs such as Agreeableness, social cognition, and default network function, it is first essential that we can accurately and reliably measure each of these constructs individually. As discussed in Chapter 1 of this dissertation, one way to increase our ability to reliably measure and model individual difference constructs and estimate their associations with other variables by using latent variable methods, such as structural equation modeling (SEM) combined with multi-indicator designs. Single-task performance-based indicators are often limited in their scope and measure constructs narrower than those they purport to represent, but using multi-indicator designs and latent variable frameworks allows us to move toward measuring constructs more reliably as what is shared across multiple tasks, thereby avoiding underestimation of true effect sizes (Apperly, 2012; Blain, Longenecker et al., 2020; Campbell & Fiske, 1959; Eisenberg et al., 2019; Enkavi et al., 2019; Nosek & Smyth, 2007). For example, social cognitive ability can be modeled as the shared variance in performance across tasks, which can then give a more accurate estimate of how individual differences in these abilities are associated with constructs such as personality and brain function.

SEM has shown specific promise for analyzing brain function. Such analyses can be facilitated using atlases such as the cortical parcellation created by Schaefer et al. (2018), where each parcel (a functionally homogeneous region of the cortex) is assigned to one of the large-scale functional networks identified by Yeo et al. (2011). This local-global parcellation scheme provides an ideal opportunity for the implementation of SEM, as the activation of a given neural network can be modeled as the shared variance of activation scores for its constituent parcels. In SEM, parcels with variance more representative of the overall network receive higher weighting in the computation of a latent variable representing overall network activation. These latent variables, representing brain activity in a given network, can then be examined as predictors of various behavioral variables, such as personality or task performance. In the current research, I leverage the advantages of SEM to investigate brain-behavior associations.

The Current Study

Research on the default network and Agreeableness provides a promising avenue to improve understanding of individual differences in ToM ability. The current study used functional magnetic resonance imaging (fMRI) to investigate relations among these constructs. Specifically, I investigated whether individual differences in activity of the default network during a ToM task (Abell et al., 2000; Castelli et al., 2000) was related to individual differences in social cognitive ability and Agreeableness. I hypothesized that neural activity in the default network, and in particular its dorsal medial subsystem, would be greater when participants were engaged in ToM (social) vs. non-social animations (*Hypothesis 1*). Further, I hypothesized ToM-related activity in the default network would positively predict participants' ToM ability as indicated by accuracy on the triangles task (*Hypothesis 2a*) and by the shared variance of performance on multiple social cognitive tasks (*Hypothesis 2b*). Finally, I expected that the same ToM-related activity in the default network would be positively associated with the personality trait of Agreeableness (*Hypothesis 3*).

Method

Data and materials for the current study are available on the Human Connectome Project's website: https://www.humanconnectome.org/study/hcp-young-adult. Additionally, I have made scripts and model specifications available in an Open Science Framework repository: https://osf.io/tf5sh/?view_only=bbe63663daf6443493ab1b330bfd3f55.

Participants

The current study included 1050 participants (564 female) from the Human Connectome Project (HCP; Van Essen et al., 2013) young adult sample; specifically, the subsample was taken from the full 1206 participant HCP sample and contained all participants with fMRI data for the ToM task. This sample included individuals between the ages of 22 and 37 (*M* = 28.8, *SD* = 3.7). Exclusion criteria for the HCP included a history of severe psychiatric, neurological, or medical disorders; however, participants were not excluded on the basis of mild psychopathology (i.e., mental illness without active psychosis or mania, medication use, or treatment for a period longer than one year). Informed consent was obtained for all participants (consent procedure is further detailed in Van Essen et al., 2013) and all study protocols were approved by the Institutional Review Board of Washington University in St. Louis (IRB # 201204036; "Mapping the Human Connectome: Structure, Function, and Heritability"). Participants completed a large battery of self-report measures and behavioral tasks; however, only measures and tasks relevant to the current research questions (i.e., measures of the Big Five and social cognitive ability) are discussed in this paper.

Measures

NEO Five-Factor Inventory (FFI). The NEO-FFI is a measure of the Big Five personality traits—Conscientiousness, Agreeableness, Neuroticism, Openness to Experience, and Extraversion. It consists of 60 items taken from the longer NEO Personality Inventory, Revised (NEO PI-R; Costa & McCrae, 1992), and uses a five-point Likert scale ranging from 0 ("Strongly Disagree") to 4 ("Strongly Agree"). Examples of Agreeableness items included "I generally try to be thoughtful and considerate," "Most people I know like me," and "If I don't like people, I let them know it (*reversed*)." The other Big Five scales (Conscientiousness, Neuroticism, Openness, and Extraversion) were used for tests of discriminant validity.

Social cognition tasks. Based on examinations of the data available in the HCP and comparisons to the existing literature, I originally identified five behavioral tasks as relevant to social cognition: a triangles ToM task, a facial emotion recognition task, a face memory condition from a working memory task, an emotional face matching task, and a moral-of-the-story identification task. After examining accuracy scores from all five of these tasks, I eventually came to focus on the first three tests due to substantial ceiling effects for the face matching and stories tasks, with most participants receiving perfect or nearly perfect scores.

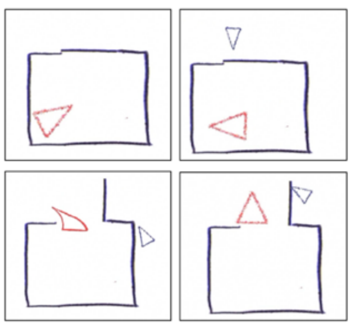
Tricky Triangles Task. While in a 3T fMRI scanner, participants were presented with a series of computerized animations of shapes interacting in either a random or social way (Castelli et al., 2000; Wheatley et al., 2007). Originally designed to assess ToM abilities in autism spectrum disorders (Abell et al., 2000), the task required participants to indicate whether each animation was random or social in nature after viewing each 20-second video clip. In the random condition, the shapes did not interact meaningfully with each other but rather moved around purposelessly. In the social condition, the shapes moved in ways that mimicked human behavior, including a variety of interaction types demonstrating particular social sequences such as coaxing, seducing, or mocking. Participants completed a total of 10 task blocks (two social and three random condition video blocks in the first run; three social and two random condition videos in the second run). Each task block was separated by a 15-second block in which participants observed a fixation cross (with 5 fixation blocks per run). Example stimuli are shown in Figure 2.2. Participants were asked to identify whether each video clip was random or social in nature and performance was scored for accuracy (i.e., whether participants correctly classified animations as random or social). Scores on the ToM triangles were negatively skewed, and a log transformation (in which scores were reversed before and after transformation to maintain scoring direction) was used to increase their normality.

Penn Emotion Recognition Task (ER40). The emotion recognition task was adapted from Gur et al. (2001). In this task, participants were presented with a series of 40 faces and were asked to identify what emotion each face expressed. Emotion options included “Happy,” “Sad,” “Angry,” “Scared,” and “No Feeling.” Eight faces were presented for each emotion, half of which were male and the other half female. See Figure 2.2 for example stimuli and answer choices. Participants’ accuracy and reaction times were recorded.

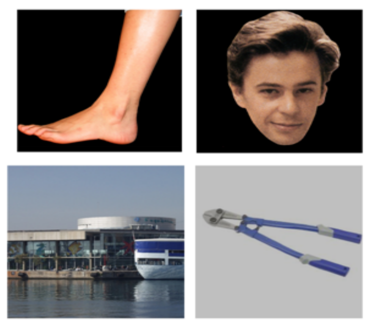
Two-back Task. In the two-back task, participants were presented with a series of stimuli from four categories: body parts, faces, places, and tools (Barch et al., 2013). See Figure 2.2 for example stimuli. In each of the conditions, participants were shown a series of objects and tasked to indicate by pressing a button whenever an object (i.e., face, body part, place, or tool) was presented that had been presented two trials previously. Each block consisted only of one stimulus type. Participants completed a total of 16 blocks (two runs of the two-back for each of the four stimulus types). Each block consisted of 10 trials, lasting 2.5 seconds each. Only the face condition was included in assessing social cognitive ability due its social relevance compared to the other conditions. Participants were scored for accuracy, separately for each of the object conditions.

Figure 2.2.

Penn Emotion Recognition (ER40)



Triangles Task



Two-back Conditions

Social cognition tasks from Chapter 2. The above figure shows images taken from the three social cognition tasks used in the study.

Task fMRI Data Acquisition and Processing

Data were obtained that had undergone preprocessing and preliminary analysis by researchers at the HCP. Specifically, I used results of the HCP's level-two, individual-subject, cortical-vertex-based analyses based on fMRI data acquired while participants completed random and social conditions of the tricky triangles task described above (Abell et al., 2000; Castelli et al., 2000). Specifics of the fMRI data acquisition are detailed in previous publications about the HCP (Ugurbil et al., 2013). In summary, whole-brain echo planar imaging acquisitions were acquired with a 32-channel head coil on a modified 3T Siemens Skyra used for all HCP data collection at Washington University in St. Louis (TR=720 ms, TE=33.1 ms, flip angle=52 deg, BW=2290 Hz/Px, in-plane FOV=208 × 180 mm, 72 axial slices, 2.0 mm isotropic voxels, with a multi-band acceleration factor of 8). One run of the ToM task used right-to-left phase encoding and the other utilized a left-to-right phase encoding.

Data analysis pipelines for the HCP were primarily built using tools adapted from FreeSurfer and FSL. The first step in processing included application of the HCP "fMRIVolume" pipeline. This process generates "minimally preprocessed" 4D time series data for each run and participant, and the pipeline steps include gradient unwarping, FLIRT-based motion correction, TOPUP-based field map preprocessing using a spin echo field map, brain-boundary-based registration of EPI to structural T1-weighted scan, non-linear (FNIRT) registration into MNI152 space, and grand-mean intensity normalization. The data were then transformed into grayordinate space, which allows for more efficient analysis of brain activation levels for components of the cortical surface. In this process, data from the cortical gray matter ribbon are projected onto the surface and then onto registered surface meshes with a standard number of vertices (in this case, approximately 30,000). Smoothing of the left and right hemisphere time series and autocorrelation estimates (from FILM) were done on the surface using a geodesic Gaussian algorithm.

Activity estimates were computed for the preprocessed functional time series from each run using a general linear model (GLM) implemented in FSL's FILM (FMRIB's Improved Linear Model) with autocorrelation correction. Predictors were convolved with a double gamma "canonical" hemodynamic response function to generate the main model regressors. To facilitate analyses of individual differences in response to given stimuli, GLM predictors were based on the category of each video clip rather than the rating of the individual (i.e., conditions were based on the appropriate response rather than each participant's actual response and accuracy was not considered in these GLMs). Each predictor covered the duration of a single video clip (20 s) and did not include time during fixation cross-viewing. To compensate for slice-timing differences and variability in the HRF delay across regions, temporal derivative terms derived from each predictor were added to each GLM and were treated as confounds of no interest. Subsequently, both the 4D time series and the GLM design were temporally filtered with a Gaussian-weighted linear highpass filter with a (soft) cutoff of 200 s.

Fixed-effects analyses were conducted using FEAT to estimate the average effects across runs within-participants. Cross-run statistical comparisons occurred in standard grayordinates space rather than volume space. As in the individual analysis, NIFTI-1 matrices were processed separately for left and right surface and subcortical volume data, and surface outputs were converted to GIFTI at the conclusion of analysis. Participant-level z-statistic maps (computed as z-transformed t-statistics) were combined from left and right hemisphere cortical and subcortical gray matter into the recently introduced CIFTI data format, with individual z-statistics for each condition output for each cortical vertex.

Group Prior Individualized Parcellation (GPIP)

Network activation was identified using group prior individualized parcellation (GPIP), an approach that begins with a standard atlas of parcels for all participants but adjusts the boundaries of each parcel for each individual to correspond to their unique cortical organziation. This is an effective solution to the problem that cortical functional organization is not identically related to anatomical landmarks in each person. For each participant, fMRI BOLD time-series acquired during the tricky triangles task in subject native space were resampled to the fsaverage5 cortical surface mesh (Dale et al., 1999) and normalized at each vertex. The resulting subject surface data were mapped onto a pre-defined group atlas with 400 functionally distinct regions (Schaefer et al., 2018) that align well with the 17-network atlas defined by Yeo et al. (2011). An iterative algorithm utilizing two Bayesian priors was applied to model connectivity between parcels and adjust parcel boundaries (Chong et al., 2017). Through this process, parcel boundaries were modified to reflect each participant’s unique patterns of functional connectivity. This method permits a more accurate approximation of individuals’ unique functional topography during the social cognition task, while maintaining correspondence of parcels across all participants and with the atlas. Previous research evaluating GPIP has demonstrated that individualized parcels exhibit greater network coherence and better segregation of task activation compared to the parcel locations from the initial group atlas (Chong et al., 2017), and a growing body of research has reported robust associations of the paramaters of individualized parcels with a variety of measures of individual differences (e.g., Anderson et al., 2021; Kong et al., 2019; Mwilambwe-Tshilobo et al., 2019).

For each participant, individualized parcels were resampled to grayordinate space to permit comparisons between parcel assignment and task activation values for each vertex. Following output of z-statistics for each cortical vertex for the social and random conditions for the ToM task (the processed data obtained from HCP's database) and generation of individualized parcellation mappings using GPIP, I mapped the vertex-wise individual participant data onto each participant's individualized parcels. I then computed parcel activation variables for each condition, for each cortical parcel associated with the default network, by averaging the z-statistics for vertices associated with each cortical parcel. The parcellation activation variables were sorted by default network subsystem and included 34 parcels associated with the core subsystem, 32 parcels associated with the dorsal medial subsystem, and 13 parcels associated with the medial temporal subsystem. I then reduced the number of variables for each of these subsystems, creating composite activation variables for cortical parcels that were anatomically adjacent. This left us with a total of 9 parcels (per condition) for the core subsystem, 9 for the dorsal medial subsystem, and 6 for the medial temporal subsystem. These parcels were used for subsequent analyses.

Statistical Analysis

Structural equation models (SEMs) were used to examine whether variation in brain activation during the social vs. random condition, for each subnetwork of the default network, predicted social cognitive ability and Agreeableness. Separate social- and random-activation latent variables were derived for each of the three subsystems, using all the corresponding cortical parcels for each subsystem as indicators. The latent variables produced by this procedure represent the shared variance among their indicators and thus can be interpreted as reflecting variation in the tendency toward activation, in a given condition, for each subnetwork as a whole. The core subsystem latent variables had a total of nine indicators: right temporal, right IPL, right PCC/Precuneus, right dPFC, right mPFC, left IPL, left PCC/Precuneus, left dPFC, and left mPFC. The dorsal medial subsystem latent variables had a total of nine indicators: right temporal, right anterior-temporal, right dPFC, right vPFC, left temporal, left IPL, left dPFC, left vPFC, and left lPFC. The medial temporal subsystem latent variables had a total of six indicators: right IPL, right parahippocampal cortex, right retrosplenial cortex, left IPL, left parahippocampal cortex, and left retrosplenial cortex.

Using the same approach to parcellation, I also created latent variables representing activation in the frontoparietal control network (FPCN), as identified by Yeo et al. (2011), in order to test for discriminant validity. The FPCN makes a good contrast to the default network because it is also involved in complex cognitive processes, like working memory and intelligence (Santarnecchi et al., 2017), but has not been strongly linked to social cognition. These frontoparietal variables were indicated by parcels located in the right PCC, right PFC, right temporal, right parietal, left PCC, left PFC, left orbitofrontal cortex, left temporal, and left parietal activation. (In Schaefer et al.'s parcellation scheme, each parcel is assigned to only one network of Yeo et al., so there was no overlap between indicators for FPCN and default network.)

In all of the models, residuals from anatomically identical manifest variables were allowed to correlate (e.g., the random and social manifest variables for right PFC activation in the dorsal medial models). Maximum likelihood estimation was used and common fit indices were computed, including the chi square, Tucker Lewis index, and root mean squared error of approximation (RMSEA). The Latent Variable Analysis (LAVAAN) package for R was used for estimating all models (Rosseel, 2012).

Based on these neural activation measurement models, I first tested for latent mean differences in activation for each of the latent variables representing the default network's three primary subsystems during the social vs. random animation conditions. Subsequently, additional SEMs were used to assess the relations among latent variables representing Agreeableness, social cognitive ability, and neural activation in the three default network subsystems and the FPCN, during the social and random animation conditions. More specifically, SEMs were conducted to test the effects of default network activation during the social animation condition on 1) accuracy on the triangles task, 2) latent social cognitive ability, representing accuracy on a variety of social cognitive tasks, and 3) Agreeableness. Separate models were computed for the core, dorsal medial, and medial temporal subsystems (and for subnetworks of the FPCN) because of a high degree of multicollinearity among the neural variables, leading to failures of model convergence if all networks were included at once. Post-hoc analyses were conducted separately for temporal and prefrontal components of the default network's dorsal medial subnetwork, given the particular relevance of these regions to social cognition and related personality traits and the fact that this network was shown to be significantly more active during social vs. random animations. Intelligence, sex (coded as 0 = female, 1 = male), age, and neural activation during the random condition (in the appropriate subnetwork) were included as covariates in all models.

In my analyses, including activation in the random condition as a covariate replaces using a contrast of the two conditions. Despite the ubiquity of contrast scores (differences in activation between two conditions) as variables of interest in neuroimaging research, this approach suffers from many of the problems that have been noted regarding the use of difference scores instead of including both conditions of interest in analyses (Allison, 1990; Edwards, 1994; Edwards, 1996; Wittenborn, 1951). When using a difference score, as is the case in a traditional fMRI contrast, variation in the effect of interest can either be due to the control condition (e.g., random animations) or the condition of interest (e.g., social animations). Difference scores do not capture any information about the association between scores on the two conditions of interest, instead imposing a linear restriction on their slopes when predicting outcome variables of interest (Allison, 1990; Edwards, 1994; Edwards, 1996; Wittenborn, 1951). Thus, if one were to use difference scores, they would not be able to identify the specific influence of social activation (vs. random activation) on the behavioral variables of interest. Including activation for both conditions as predictors allows us to partial out variance in activation that is shared between the conditions, such that we see the effect of the condition of interest after controlling for the baseline provided by the random condition. The effect of interest, therefore, indicates how much each subject's activation deviates in the condition of interest from the activation that would be expected based on the control condition. Thus, we can accurately estimate the unique associations of activation during the social condition (as well as activation during the random condition) with behavioral variables of interest.

Latent social cognitive ability was modeled using accuracy variables for the triangles task (correct vs. incorrect responses for the random and social animations), ER40, and the face memory condition from the two-back task. In addition to this latent social cognitive ability variable, I also conducted a test using accuracy on only the ToM triangles task as manifest criterion variable; this manifest outcome variable test was included because neural variables measured activation specifically during this task. For all relevant models, intelligence was modeled using tests of Picture Vocabulary, Matrix Reasoning, and English Reading, as well as a hierarchical working memory factor using the four two-back task conditions, onto which the face memory variable was allowed to cross load.

Results

Descriptive statistics for self-report measures and task performance are presented in Table 2.1. Pearson correlations among all variables are presented in Tables 2.2, 2.3, and 2.4. Weak positive zero-order correlations were observed between social cognition accuracy measures and NEO Agreeableness. Social cognition measures were also positively correlated with intelligence measures, with stronger magnitudes.

Table 2.1. *Descriptive statistics for Chapter 2 self-report and task measures*

	Mean (SD)	Skew	[Minimum, Maximum]
NEO Agreeableness	33.3 (5.8)	-0.3	[10, 48]
Triangles Task Accuracy	80.7 (12.1)	-1.8	[5, 100]
Emotion Recognition Accuracy	35.5 (2.6)	-1.0	[24, 40]
Two-back Overall Accuracy	89.3 (10.6)	-1.3	[25, 100]

Table 2.2. *Pearson correlations among Chapter 2 behavioral measures*

	Sex	Age	Vocab	Matrix	Read	Tool	Body	Place	Face	ER40	ToM	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
Sex	1.0																						
Age	-.21**	1.0																					
Vocab	.11**	.10**	1.0																				
Matrix	.14**	-.08**	.47**	1.0																			
Read	.12**	.03	.70**	.47**	1.0																		
Tool	.16**	-.13**	.25**	.36**	.28**	1.0																	
Body	.16**	-.11**	.29**	.35**	.30**	.55**	1.0																
Place	.15**	-.10**	.29**	.35**	.33**	.49**	.49**	1.0															
Face	.06*	-.09**	.33**	.38**	.38**	.53**	.46**	.49**	1.0														
ER40	-.05	-.01	.20**	.22**	.22**	.10**	.15**	.10**	.14**	1.0													
ToM	.01	-.01	.22**	.21**	.16**	.14**	.20**	.16**	.18**	.14**	1.0												
A1	-.04	.03	.03	.06*	.06*	-.03	-.03	-.02	.00	.07*	-.02	1.0											
A2	-.14**	.08**	.05	.09**	.09**	.01	.02	.03	.02	.08**	.08**	.17**	1.0										
A3	-.20**	.05	-.09**	-.02	-.07*	-.05	-.09**	-.05	-.04	.06*	-.02	.23**	.32**	1.0									
A4	-.11**	.02	.07*	.03	.05	-.06	-.03	-.04	-.04	.00	.07	.18**	.14**	.16**	1.0								
A5	-.12**	.01	-.01	-.03	-.01	-.04	-.03	-.02	-.01	.02	-.03	.12**	.19**	.31**	.13**	1.0							
A6	-.12**	.09**	.33**	.21**	.34**	.13**	.12**	.16**	.18**	.13**	.12**	.11**	.22**	.19**	.11**	.38**	1.0						
A7	-.03	-.03	-.15**	-.07*	-.11**	-.10**	-.08**	-.05	-.09**	-.07*	-.08*	.21**	.21**	.28**	.10**	.18**	.02	1.0					
A8	-.27**	.14**	-.08**	-.05	-.04	-.07*	-.06	-.03	-.06	.05	-.02	.21**	.31**	.49**	.11**	.34**	.24**	.27**	1.0				
A9	-.04	.02	.05	.02	.06*	-.04	-.01	-.02	.01	.01	.01	.09**	.27**	.26**	.18**	.25**	.22**	.13**	.27**	1.0			
A10	-.08**	.02	.12**	.12**	.14**	.00	.03	.03	.07*	.05	.08	.38**	.23**	.25**	.19**	.15**	.13**	.23**	.24**	.12**	1.0		
A11	-.10**	.06*	.20**	.22**	.22**	.09**	.08**	.11**	.12**	.07*	.14**	.18**	.26**	.19**	.17**	.20**	.32**	.01	.23**	.22**	.21**	1.0	
A12	-.19**	.10**	-.11**	-.07*	-.07*	-.03	-.08**	-.06	-.04	.00	-.02	.15**	.25**	.39**	.15**	.26**	.15**	.13**	.36**	.19**	.18**	.22**	1.0

Notes. *$p < 0.05$, **$p < 0.01$. Matrix = Penn Matrix Test, Vocab = Picture Vocabulary, Read = Reading English from the NIH Toolbox, Tool, Body, Place, and Face = Conditions on the 2-back task, ER40 = Accuracy on the Penn Emotion Recognition Task, ToM = Triangles task accuracy, A 1-12 = Agreeableness NEO-FFI items 1 through 12. Items A2, A3, A5, A6, A8, A9, A11, and A12 from the NEO-FFI were reverse scored.

Table 2.3. *Pearson correlations among Chapter 2 neural measures (for primary subnetwork of interest)*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Social_DN_B_RH_Temp	1.0																	
2. Social_DN_B_RH_AntTemp	.67	1.0																
3. Social_DN_B_RH_vPFC	.51	.37	1.0															
4. Social_DN_B_RH_dPFC	.54	.50	.53	1.0														
5. Social_DN_B_LH_Temp	.59	.51	.60	.64	1.0													
6. Social_DN_B_LH_IPL	.39	.41	.37	.49	.63	1.0												
7. Social_DN_B_LH_dPFC	.62	.61	.37	.55	.61	.55	1.0											
8. Social_DN_B_LH_lPFC	.63	.57	.40	.59	.65	.57	.60	1.0										
9. Social_DN_B_LH_vPFC	.66	.58	.40	.51	.57	.41	.76	.56	1.0									
10. Random_DN_B_RH_Temp	.66	.46	.41	.33	.41	.26	.38	.38	.44	1.0								
11. Random_DN_B_RH_AntTemp	.39	.63	.22	.28	.34	.26	.35	.35	.36	.65	1.0							
12. Random_DN_B_RH_vPFC	.39	.27	.84	.37	.46	.26	.25	.28	.30	.55	.33	1.0						
13. Random_DN_B_RH_dPFC	.35	.31	.39	.60	.42	.29	.31	.33	.32	.56	.48	.53	1.0					
14. Random_DN_B_LH_Temp	.42	.37	.48	.43	.71	.41	.38	.42	.39	.63	.55	.62	.65	1.0				
15. Random_DN_B_LH_IPL	.25	.29	.30	.33	.43	.68	.32	.33	.26	.40	.44	.40	.48	.64	1.0			
16. Random_DN_B_LH_dPFC	.33	.36	.17	.28	.36	.34	.57	.37	.44	.55	.60	.27	.48	.59	.56	1.0		
17. Random_DN_B_LH_lPFC	.40	.38	.35	.37	.43	.36	.33	.59	.34	.64	.59	.47	.61	.69	.57	.60	1.0	
18. Random_DN_B_LH_vPFC	.40	.36	.27	.30	.36	.27	.45	.34	.66	.65	.61	.39	.49	.57	.44	.74	.57	1.0

Notes. RH = right hemisphere, LH = left hemisphere, DN = default network, Temp = temporal, AntTemp = anterior temporal, IPL = inferior parietal lobule, PFC = prefrontal cortex, v = ventral, d = dorsal, l = lateral

Table 2.4. *Pearson correlations among Chapter 2 behavioral and neural measures (for primary subnetwork of interest)*

	Sex	Age	Vocab	Matrix	Read	Tool	Body	Place	Face	ER40	ToM	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
Social_DN_B_RH_Temp	.07*	-.02	.04	.09**	.06	.07*	.09**	.09**	.06	.05	.08*	.01	-.04	.03	-.02	.01	.02	-.02	-.01	-.05	-.04	.01	-.01
Social_DN_B_RH_AntTemp	.05	-.01	.03	.06	.05	.05	.09**	.09**	.05	.06	.04	.03	-.02	.05	.00	.11**	.06*	.03	.03	-.01	.01	.04	.02
Social_DN_B_RH_vPFC	-.05	.02	.18**	.15**	.12**	.15**	.16**	.17**	.18**	.13**	.12**	.00	.07*	.02	.02	.01	.09**	-.04	.01	.02	.05	.09**	.02
Social_DN_B_RH_dPFC	-.04	.02	.07*	.08*	.06	.12**	.10**	.14**	.12**	.07*	.07*	-.01	.03	.03	-.03	.03	.07*	-.03	.02	.04	-.03	.04	.01
Social_DN_B_LH_Temp	-.01	.01	.08*	.09**	.04	.09**	.10**	.13**	.14**	.07*	.11**	.01	.02	.04	.00	.04	.09**	-.05	.05	-.01	-.02	.04	.03
Social_DN_B_LH_IPL	-.10**	.09**	.03	.03	-.03	.06	.03	.07*	.10**	.01	.03	.05	.07*	.09**	.00	.08*	.13**	.00	.12**	.04	.00	.08*	.08*
Social_DN_B_LH_dPFC	-.03	.02	.04	.09**	.04	.05	.05	.06	.08*	.09**	.09**	.08**	.00	.11**	.02	.05	.05	.02	.08**	.02	.01	.07*	.05
Social_DN_B_LH_lPFC	.03	.03	-.01	.03	.00	.04	.02	.06	.04	.02	.00	.01	-.03	.04	-.03	.05	.02	.03	.05	-.01	-.02	-.01	.01
Social_DN_B_LH_vPFC	.11**	-.01	.06*	.10**	.09**	.09**	.09**	.11**	.09**	.06	.03	.05	-.05	.04	-.01	.01	.00	-.01	.01	.01	.00	-.01	.00
Random_DN_B_RH_Temp	.11**	-.04	.02	.09**	.04	.05	.03	.09**	.05	.06	-.02	-.02	-.02	-.01	-.07*	.02	.01	.00	-.04	-.07*	.00	.03	.00
Random_DN_B_RH_AntTemp	.04	-.05	.01	.01	.00	.00	-.01	.02	-.01	.03	-.05	-.01	.01	.01	-.06	.07*	.02	.04	.00	-.04	.02	.00	.00
Random_DN_B_RH_vPFC	-.04	.02	.11**	.13**	.10**	.12**	.14**	.17**	.16**	.11**	.04	.01	.09**	.01	.00	.02	.08*	.01	.03	-.01	.06	.10**	.05
Random_DN_B_RH_dPFC	.01	-.04	.00	.05	.02	.09**	.06*	.13**	.10**	.05	.01	-.03	.05	.03	-.04	.04	.05	-.01	-.01	.01	-.01	.03	.05
Random_DN_B_LH_Temp	.04	-.05	.03	.08*	.03	.08**	.08**	.13**	.12**	.06	.07*	-.03	.02	-.06	.03	.06	-.01	-.01	.01	-.06*	-.01	.02	.02
Random_DN_B_LH_IPL	-.04	.03	-.01	-.01	-.01	.06	.03	.08**	.10**	.01	.04	.00	.09**	.06*	-.04	.06*	.10**	.03	.08**	-.01	-.02	.07*	.05
Random_DN_B_LH_dPFC	.01	.01	-.02	-.02	-.01	.01	-.05	.01	.01	.02	.01	.01	.01	.06*	-.05	.02	.01	.02	.00	-.03	-.02	.00	.00
Random_DN_B_LH_lPFC	.07*	-.03	.01	.04	.03	.05	.02	.08*	.07*	.04	.01	-.04	.00	.02	-.05	.05	.02	.01	.01	-.02	-.01	-.02	.01
Random_DN_B_LH_vPFC	.10**	-.05	.01	.05	.03	.06*	.00	.09**	.05	.02	-.03	-.02	-.01	.00	-.04	.02	-.03	.00	-.04	-.04	-.01	-.02	.00

Notes. $*p < 0.05$, $**p < 0.01$. RH = right hemisphere, LH = left hemisphere, DN = default network, Temp = temporal, AntTemp = anterior temporal, IPL = inferior parietal lobule, PFC = prefrontal cortex, v = ventral, d = dorsal, l = lateral, Matrix = Penn Matrix Test, Vocab = Picture Vocabulary Test, Read = Reading English Test from the NIH Toolbox, Tool, Body, Place, and Face = Conditions on the 2-back working memory task, ER40 = Accuracy on the Penn Emotion Recognition Task, ToM = Accuracy on the triangles task, A 1-12 = Agreeableness NEO-FFI items 1 through 12. Items A2, A3, A5, A6, A8, A9, A11, and A12 from the NEO-FFI were reverse scored.

Hypothesis 1

SEM was used to test for mean differences in default network activation for the social vs. random conditions of the triangles task. Results are visualized in Figure 2.3. Fit statistics for all SEMs are presented in Table 2.5. For these models, criterion variables are calculated as latent variables representing overall activation across the regions of a given subnetwork and can be interpreted as reflecting activation (for a given condition) in that subnetwork as a whole. Activation of the default network's dorsal medial subsystem was significantly greater during the social condition vs. the random condition (Figure 2.3; z = 5.70, p < .001); this pattern of activation held both for regions centered on the temporal parietal junction and temporal pole (Figure 2.3; z = 6.26, p < .001) and for regions in the prefrontal cortex (Figure 2.3; z = 14.37, p < .001). Activation was significantly less for the social condition vs. the random condition in the medial temporal subsystem (Figure 2.3; z = -5.68, p < .001) and core subsystem (Figure 2.3; z = -9.42, p = .089).

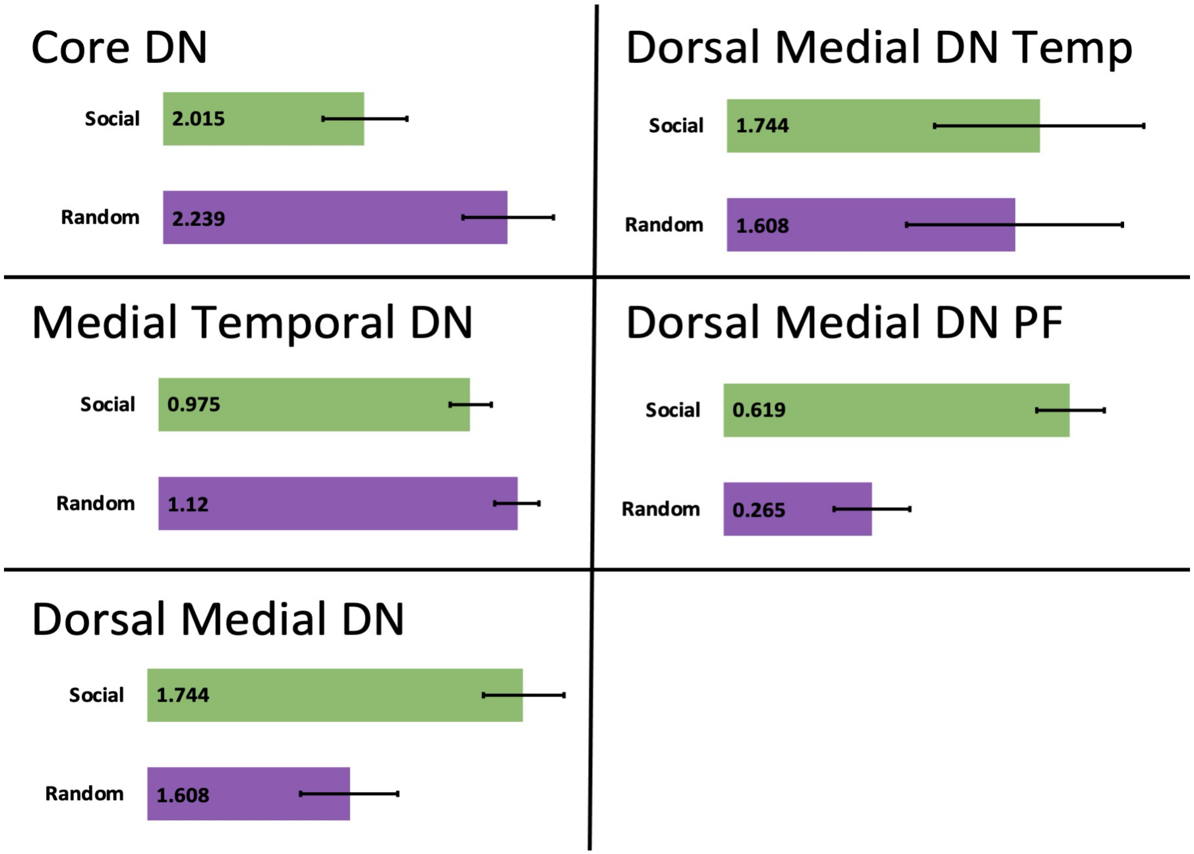
Table 2.5. *Fit statistics for Chapter 2 structural equation models*

Models	df	χ2	p	TLI	RMSEA	95% C.I.
Primary Models						
DN Overall — Social vs. Random	6	405.8	< .001	.911	.252	[.231, .273]
DN Core — Social vs. Random	126	1397.8	< .001	.913	.098	[.093, .103]
DN Dorsal Medial — Social vs. Random	96	1665.5	< .001	.876	.125	[.120, .130]
DN Dorsal Medial (Temporal) — Social vs. Random	6	23.5	.001	.990	.053	[.031, .076]
DN Dorsal Medial (Prefrontal) — Social vs. Random	30	818.8	< .001	.854	.158	[.149, .168]
DN Medial Temporal — Social vs. Random	48	704.1	< .001	.901	.114	[.107, .122]
DN Overall — Triangles Accuracy	88	363.1	< .001	.973	.055	[.049, .061]
DN Core — Triangles Accuracy	328	1899.5	< .001	.911	.068	[.065, .071]
DN Dorsal Medial — Triangles Accuracy	278	2087.0	< .001	.885	.079	[.076, .083]
DN Dorsal Medial (Temporal) — Triangles Accuracy	88	347.1	< .001	.948	.053	[.048, .059]

DN Dorsal Medial (Prefrontal) — Triangles Accuracy	152	1071.2	< .001	.893	.077	[.072, .081]
DN Medial Temporal — Triangles Accuracy	190	1117.8	< .001	.904	.069	[.065, .073]
DN Overall — Social Cognition Accuracy	101	361.6	< .001	.974	.050	[.044, .056]
DN Core — Social Cognition Accuracy	353	1913.1	< .001	.912	.065	[.063, .068]
DN Dorsal Medial — Social Cognition Accuracy	301	2098.8	< .001	.886	.076	[.073, .079]
DN Dorsal Medial (Temporal) — Social Cognition Accuracy	101	343.1	< .001	.953	.048	[.043, .054]
DN Dorsal Medial (Prefrontal) — Social Cognition Accuracy	169	1075.9	< .001	.896	.072	[.068, .076]
DN Medial Temporal — Social Cognition Accuracy	209	1131.3	< .001	.906	.065	[.062, .069]
DN Overall — Agreeableness	159	597.4	< .001	.961	.051	[.047, .056]
DN Core — Agreeableness	447	2161.9	< .001	.905	.061	[.058, .063]
DN Dorsal Medial — Agreeableness	389	2344.9	< .001	.879	.069	[.067, .072]
DN Dorsal Medial (Temporal) — Agreeableness	159	565.3	< .001	.924	.049	[.045, .054]
DN Dorsal Medial (Prefrontal) — Agreeableness	239	1292.5	< .001	.882	.065	[.061, .068]
DN Medial Temporal — Agreeableness	285	1271.4	< .001	.901	.057	[.054, .061]
Discriminant Validity Models						
FPCN — Triangles Accuracy	88	336.7	< .001	.973	.052	[.046, .058]
FPCN — Social Cognition Accuracy	101	334.6	< .001	.975	.047	[.042, .053]
FPCN — Agreeableness	159	568.8	< .001	.960	.050	[.045, .054]
FPCN A — Triangles Accuracy	382	2835.2	< .001	.863	.079	[.076, .082]
FPCN A — Social Cognition Accuracy	409	2860.3	< .001	.864	.076	[.074, .079]
FPCN A — Agreeableness	509	3065.2	< .001	.859	.069	[.067, .072]
FPCN B — Triangles Accuracy	382	2116.5	< .001	.911	.066	[.064, .069]
FPCN B — Social Cognition Accuracy	409	2133.3	< .001	.912	.064	[.061, .067]
FPCN B — Agreeableness	509	2383.9	< .001	.905	.059	[.057, .062]
FPCN C — Triangles Accuracy	118	419.5	< .001	.961	.050	[.045, .055]
FPCN C — Social Cognition Accuracy	133	420.1	< .001	.964	.046	[.041, .051]
FPCN C — Agreeableness	197	651.4	< .001	.945	.047	[.043, .051]
DN Dorsal Medial Prefrontal — Conscientiousness	239	1619.8	< .001	.861	.074	[.071, .078]
DN Dorsal Medial Prefrontal — Neuroticism	239	1344.8	< .001	.891	.066	[.063, .070]
DN Dorsal Medial Prefrontal — Openness	239	1745.8	< .001	.840	.078	[.074, .081]
DN Dorsal Medial Prefrontal — Extraversion	239	1628.1	< .001	.852	.075	[.071, .078]

Note. DN = default network, FPCN = frontoparietal control network

Figure 2.3.



Social vs. Random Brain Activation during the Triangles Task. DN = default network. Figure shows standardized latent means for activation in DN subsystems and the FPCN for the social vs. random conditions of the triangles task. Dependent variables were calculated (via structural equation modeling) as overall activation across the regions of a given subnetwork and can be interpreted as reflecting activation in that subnetwork as a whole (during each condition).

Hypotheses 2a, 2b, and 3

SEM was also used to test for associations of default network activation during the social condition of the ToM task with social cognitive ability and Agreeableness (Table 2.6). Activation of the dorsal medial, core, and medial temporal subsystems of the default network were positively associated with accuracy on the triangles task, controlling for age, sex, intelligence, and neural activation in the random condition (Table 2.6). Similarly, activation of all three default network subsystems was positively associated with shared variance in accuracy across various tests of social cognitive ability[1]. Across models, intelligence was a significant positive predictor of social cognitive ability. Further, sex was a significant predictor of social cognitive ability, with females performing better on social cognition tasks on average. The full measurement and structural model is presented for the prefrontal component of the dorsal medial subsystem in Figure 2.4.

Activation during the social condition significantly predicted Agreeableness for the medial temporal subsystem and for the prefrontal cortex component of the dorsal medial subsystem (Table 2.6). As with the social cognitive ability variable, sex was a significant predictor of Agreeableness, with females having higher levels of Agreeableness. A full measurement and structural model is presented for the prefrontal cortex component of the dorsal medial subsystem in Figure 2.5.

---

[1] At the request of a reviewer, I reran models without including the face condition of the two-back task, to ensure that including this variable and its cross-loading from the social cognition latent variable was not unduly influencing results. Findings were substantively equivalent across models when this variable was excluded.

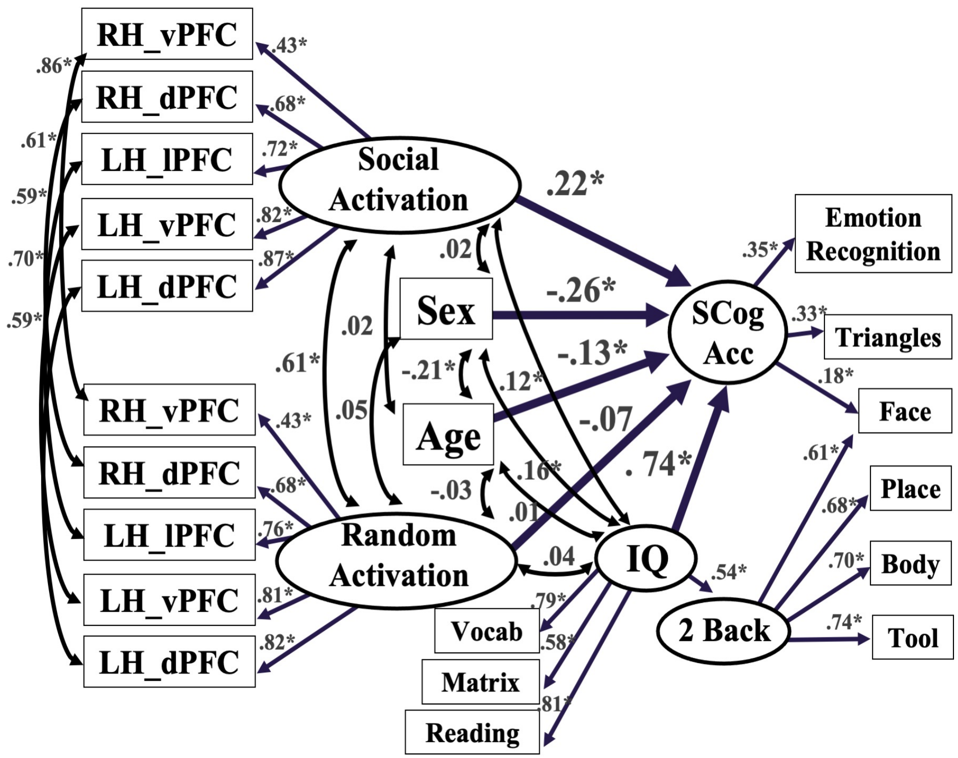
Table 2.6. *Results for Chapter 2 default network activation models*

Models	*z*	*β*	*p*
Triangles Accuracy			
DN Overall Social Activation	3.11	.13	.002
DN Overall Random Activation	-1.59	-.07	.113
DN Core Social Activation	2.69	.11	.007
DN Core Random Activation	-1.30	-.05	.193
DN Dorsal Medial Social Activation	2.50	.10	.013
DN Dorsal Medial Random Activation	-1.38	-.06	.168
DN Dorsal Medial (Temporal) Social Activation	2.22	.09	.026
DN Dorsal Medial (Temporal) Random Activation	-0.32	-.01	.751
DN Dorsal Medial (Prefrontal) Social Activation	2.52	.10	.012
DN Dorsal Medial (Prefrontal) Random Activation	-1.74	-.07	.083
DN Medial Temporal Social Activation	2.88	.12	.004
DN Medial Temporal Random Activation	-1.75	-.07	.080
Social Cognition Accuracy			
DN Overall Social Activation	2.83	.24	.005
DN Overall Random Activation	-0.39	-.03	.700
DN Core Social Activation	2.58	.22	.010
DN Core Random Activation	-0.26	-.02	.792
DN Dorsal Medial Social Activation	2.28	.19	.023
DN Dorsal Medial Random Activation	-0.14	-.01	.889
DN Dorsal Medial (Temporal) Social Activation	2.39	.20	.017
DN Dorsal Medial (Temporal) Random Activation	0.41	.03	.679
DN Dorsal Medial (Prefrontal) Social Activation	2.64	.22	.008
DN Dorsal Medial (Prefrontal) Random Activation	-0.80	-.07	.423
DN Medial Temporal Social Activation	2.62	.22	.009
DN Medial Temporal Random Activation	-1.05	-.08	.296
Agreeableness			
DN Overall Social Activation	1.38	.06	.168
DN Overall Random Activation	0.29	.01	.769
DN Core Social Activation	1.42	.07	.157
DN Core Random Activation	0.19	.01	.849

DN Dorsal Medial Social Activation	1.54	.07	.125
DN Dorsal Medial Random Activation	-0.76	.00	.939
DN Dorsal Medial (Temporal) Social Activation	1.28	.06	.201
DN Dorsal Medial (Temporal) Random Activation	-0.08	.00	935
DN Dorsal Medial (Prefrontal) Social Activation	1.99	.09	.046
DN Dorsal Medial (Prefrontal) Random Activation	-0.50	-.02	.621
DN Medial Temporal Social Activation	1.98	.09	.047
DN Medial Temporal Random Activation	-0.08	.00	.940

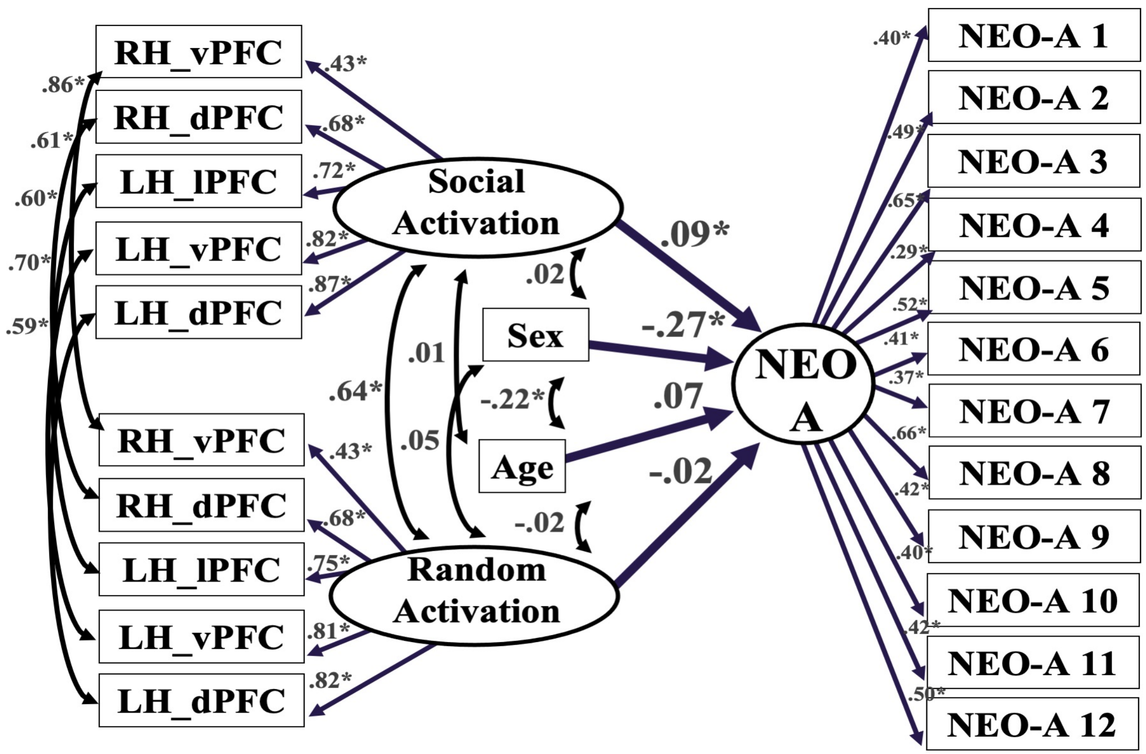
Note. DN = default network

Figure 2.4.



Relation between Dorsal Medial Prefrontal Activation and Social Cognitive Ability. RH = right hemisphere, LH = left hemisphere, PFC = prefrontal cortex, v = ventral, d = dorsal, l = lateral, SCog Acc = accuracy on the social cognition tasks. Structural equation modeling was used to test the association of dorsal medial default network activation during the social condition of the theory of mind task with social cognitive ability. Age, sex, intelligence, and neural activation in the random condition were included as covariates. Activation of the default network's dorsal medial subsystem (modeled as the shared variance among activation scores in spatially contiguous regions) was positively associated with shared variance in accuracy across various tests of social cognitive ability.

Figure 2.5.



Relation between Dorsal Medial Prefrontal Activation and Agreeableness. RH = right hemisphere, LH = left hemisphere, PFC = prefrontal cortex, v = ventral, d = dorsal, l = lateral, NEO A = Agreeableness. Structural equation modeling was used to test the association of dorsal medial default network activation during the social condition of the theory of mind task with Agreeableness. Age, sex, and neural activation in the random condition were included as covariates. Activation of the default network's dorsal medial subsystem (modeled as the shared variance among activation scores in spatially contiguous regions) was positively associated with shared variance in Agreeableness items from the NEO Five Factor Inventory.

A final set of analyses was conducted to examine discriminant validity. I found that activation in FPCN subnetworks during the social animations condition also significantly predicted accuracy on the triangles task and latent social cognitive ability, but not Agreeableness. Effects were generally, but not always, weaker compared to effects observed in the default network models (Table 2.7). Activation in the default network during the social condition blocks did not predict personality traits other than Agreeableness (i.e., Conscientiousness, Neuroticism, Extraversion, or Openness). Full results of these analyses are displayed in Table 2.7.

Table 2.7. *Results for Chapter 2 discriminant validity models*

Models	*z*	*β*	*p*
Triangles Accuracy			
FPCN Social Activation	2.63	.11	.009
FPCN Random Activation	-1.48	-.06	.140
Social Cognition Accuracy			
FPCN Social Activation	2.38	.20	.017
FPCN Random Activation	-0.29	-.02	.770
Agreeableness			
FPCN Social Activation	1.38	.07	.168
FPCN Random Activation	0.22	.01	.827
Triangles Accuracy			
FPCN A Social Activation	2.74	.12	.006
FPCN A Random Activation	-1.69	-.07	.090
Social Cognition Accuracy			
FPCN A Social Activation	2.31	.20	.021
FPCN A Random Activation	-0.60	-.05	.550
Agreeableness			
FPCN A Social Activation	1.29	.06	.198
FPCN A Random Activation	0.49	.02	.623

Triangles Accuracy			
FPCN B Social Activation	2.78	.12	.005
FPCN B Random Activation	-1.52	-.06	.129
Social Cognition Accuracy			
FPCN B Social Activation	2.35	.20	.019
FPCN B Random Activation	-0.23	-.02	.821
Agreeableness			
FPCN B Social Activation	0.99	.05	.323
FPCN B Random Activation	0.38	.02	.705
Triangles Accuracy			
FPCN C Social Activation	2.33	.11	.020
FPCN C Random Activation	-1.40	-.06	.163
Social Cognition Accuracy			
FPCN C Social Activation	1.90	.17	.057
FPCN C Random Activation	-0.08	-.01	.939
Agreeableness			
FPCN C Social Activation	1.11	.06	.265
FPCN C Random Activation	0.22	.01	.828
Conscientiousness			
DN Dorsal Medial - Prefrontal Social Activation	0.07	.00	.942
DN Dorsal Medial - Prefrontal Random Activation	-0.13	-.01	.897
Neuroticism			
DN Dorsal Medial - Prefrontal Social Activation	-0.90	-.04	.367
DN Dorsal Medial - Prefrontal Random Activation	-0.47	-.02	.639
Openness			
DN Dorsal Medial - Prefrontal Social Activation	1.12	.05	.264
DN Dorsal Medial - Prefrontal Random Activation	-0.89	-.04	.374
Extraversion			
DN Dorsal Medial - Prefrontal Social Activation	0.81	.04	.417
DN Dorsal Medial - Prefrontal Random Activation	-0.72	-.03	.471

Note. DN = default network, FPCN = frontoparietal control network

Discussion

The current study used a large fMRI sample, multiple behavioral tasks, and SEM to investigate how ToM-related activity in the default network and its subsystems was related to social cognitive ability and Agreeableness. Findings largely confirmed my three main hypotheses. Neural activity in the dorsal medial subsystem of the default network was significantly greater during the viewing of social animations compared to random animations (*Hypothesis 1*). Activity in the dorsal medial subsystem—while participants viewed the social animations—positively predicted performance on the triangles ToM task (*Hypothesis 2a*), and this was true with or without controlling for covariates such as intelligence, suggesting the association is robust. This positive association was also found for the default network's core and medial temporal subsystems, as well as components of the frontoparietal control network. Likewise, neural activity in these regions, during the social condition of the task, also positively predicted social cognitive ability, modeled using a latent variable indicated by accuracy scores on three different social cognition tasks (*Hypothesis 2b*). Finally, neural activity during the social animations positively predicted individual differences in the personality trait Agreeableness, for prefrontal regions of the dorsal medial subsystem and for the medial temporal subsystem (*Hypothesis 3*). Associations with Agreeableness were not seen for another neural network involved in complex cognition (the FPCN), nor were other personality traits associated with default network activation, suggesting specificity for the associations among default network activation and Agreeableness.

Collectively, the current results suggest that individual differences in both Agreeableness and ToM are related to variation in the same underlying neural network. Findings reinforce previous research tying the default network—and more specifically its dorsal medial subsystem—to ToM and the ability to understand the mental states and emotions of others (Allen et al., 2017; Amodio & Frith, 2006; Carrington & Bailey, 2009; Frith & Frith, 2006; Sabbagh et al., 2004; Saxe & Kanwisher, 2003; Saxe & Powell, 2006; Saxe & Wexler, 2005; Schurz et al., 2014; Spreng & Andrews-Hanna, 2015; Spunt & Lieberman, 2012; Vogeley et al., 2001; Young et al., 2010). The present study extends previous findings by demonstrating that activity in the dorsal medial subsystem not only predicted performance on a single task, but also performance on a variety of social cognitive tasks modeled as a latent variable, thereby providing evidence for a positive association between individual differences in broad social cognitive ability and default network function. This relation could provide insight into how both cognitive and neural variation contribute to individual differences in social functioning.

The current study demonstrated that ToM-related brain activity in prefrontal regions of the dorsal medial subsystem positively predicted individual differences in Agreeableness, a personality trait linked to social cognition and especially relevant for understanding social interactions. Previous research evaluating the association between personality and social functioning has linked both Extraversion and Agreeableness with interpersonal tendencies and Trait Affiliation (Côté & Moskowitz, 1998; DeYoung et al., 2013; DeYoung & Weisberg, 2018). Each of the Big Five traits can be thought of as relating to particular motivational, cognitive, and affective mechanisms (DeYoung, 2015;

DeYoung & Blain, 2020). For example, pattern detection and curiosity for Openness-Intellect (Bainbridge et al., 2019; Blain, Longenecker, et al., 2020; DeYoung et al., 2012; Silvia & Christensen, 2020) and reward sensitivity for Extraversion (Blain, Sassenberg, et al., 2020; Lucas et al., 2000; Smillie et al., 2012). Agreeableness appears to reflect tendencies related to navigating social norms and coordinating with the needs of others (DeYoung, 2015; DeYoung & Weisberg, 2018; Koole et al., 2001). Agreeableness in particular has been associated with prosociality (Caprara et al., 2010; Habashi et al., 2016), higher levels of satisfaction in relationships (Malouff et al., 2010; Weidmann et al., 2017), and less prejudicial behavior towards others (Sibley & Duckitt, 2008). Though Agreeableness has been positively associated with many desirable social outcomes, its underlying mechanisms remain understudied among the Big Five personality traits, with few studies having investigated its neurocognitive correlates (DeYoung & Blain, 2020). Social cognitive ability and default network function, however, appear to be promising candidates for understanding the substrates of individual differences in Agreeableness (Allen et al., 2017; Arbula et al., 2021).

Synthesizing Current Findings and Previous Work

The current findings were consistent with previous work, demonstrating positive associations of default network function with social cognitive ability and Agreeableness. For instance, previous studies with high statistical power have demonstrated positive associations of resting state functional connectivity within the default network and ToM ability and questionnaire measures of trait empathy and Compassion (Allen et al., 2017; Takeuchi et al., 2014). The current study extends this work to look at brain activity during a ToM task rather than just during rest, and suggests that ToM-related brain activity in the dorsal medial subsystem of the default network may be associated with both Agreeableness and ToM abilities.

Together with the previous work, the current findings suggest a possible explanation for why people high in Agreeableness tend to demonstrate better interpersonal outcomes than less agreeable people: highly agreeable people may have better social abilities because of differences in the function of specific brain networks including the default network and particularly its dorsal medial subsystem. This study also accounts for possible alternative explanations by including covariates such as intelligence, sex, and age. Though some covariates were also related to the variables of interest, controlling for them did not eliminate the hypothesized effects.

Given the current results, which suggest significant associations of social cognitive ability and Agreeableness with default network activation, it is worth mentioning contrasting findings from a recent study utilizing the same HCP dataset (Weiss et al., 2021). Weiss et al. found no meaningful relations between personality variables and neural activity during the same triangles task data I analyze in the current study. The authors attribute the lack of significant associations to methodological issues such as the questionable validity of the social cognition task and test–retest reliability of functional biomarkers. The utilization of better methods, such as the individualized parcellation approach of GPIP and latent variable modeling for behavioral and neural variables, can increase reliability and thus the ability to detect true associations among variables (Blain, Longenecker et al., 2020; Campbell & Fiske, 1959; Chong et al., 2017;

Keith, 2006; Kong et al., 2021; Eisenberg et al., 2019; Enkavi et al., 2019; Nosek & Smyth, 2007). This increased reliability of my variables of interest is a likely explanation for why I was able to detect significant associations among social cognition, personality, and default network activation in the current work, in contrast to the null effects observed by Weiss et al. (2021).

In the current study, although all default network subsystems showed robust relations to individual differences in social cognitive ability, only the dorsal medial subsystem was significantly more active during the social condition of the ToM task compared to the random condition, and only activity in the medial temporal subsystem and prefrontal regions of the dorsal medial subsystem significantly predicted individual differences in Agreeableness. Though we should avoid overinterpreting this specificity of the dorsal medial subsystem and its prefrontal regions, as effect sizes were fairly similar in magnitude across the subsystems, the current pattern of results is in line with research suggesting the dorsal medial subsystem may be more strongly linked to social cognition than the other two default network subsystems. The dorsal medial subsystem also appears to have broader functions in language processing, which can be argued to be inherently social (Spreng & Andrews-Hanna, 2015).

The core subsystem includes regions of the brain associated with social cognitive functions (Leech & Sharp, 2014; Spreng & Andrews-Hanna, 2015; Spreng et al., 2009), but also other cognitive functions that are not as specifically social in nature, such as the retrieval of autobiographical memory and personal knowledge (Moran et al., 2013; Spreng & Andrews-Hanna, 2015; Spreng et al., 2009). Similarly, the medial temporal subsystem is involved in more non-social cognitive functions, again including episodic memory (Buckner et al., 2008; Spreng & Andrews-Hanna, 2015). Future work should more specifically investigate the coordination of these three subsystems, as the joint activation of these subsystems and functional connectivity between the systems appears to be particularly relevant to social cognition and corresponding individual differences (Allen et al., 2017; Spreng & Andrews-Hanna, 2015; Takeuchi et al., 2014).

Role of General Cognitive Ability and Sex Differences

Social cognitive ability was significantly predicted by general cognitive ability and sex, as well as by Agreeableness and brain activity. The association between social cognitive ability and general cognitive ability is not surprising, as utilizing social cognitive processes likely also engages other general cognitive processes, such as working memory (Phillips et al., 2008; Spreng, 2013; Thornton & Conway, 2013), attentional processes (Holmes et al., 2003; Leslie et al., 2004; Schultebraucks et al., 2016), and nonverbal communication skills (Morrison et al., 2019). Indeed, previous research suggests strong positive correlations between general and social cognitive abilities (Allen et al., 2017; Landy, 2005; Thorndike & Stein, 1937).

What is perhaps more interesting is the potential role general intelligence might play in the association between sex and social cognitive ability, as well as how the current findings might provide some explanation for why previous research has shown mixed results for sex differences in social cognitive ability (DiTella et al., 2020). Without controlling for age or intelligence, associations between sex and the indicators for my latent social cognitive ability variable suggested little sex difference. Once general intelligence and age were introduced as covariates, however, a significant negative association appeared between sex and my latent social cognitive ability variable, indicating that females displayed higher social cognitive ability than males. Further, zero-order correlations between sex and all the indicators of the latent IQ variable (i.e., Picture Vocabulary, English Reading, Matrix Reasoning, and all four conditions of the two-back task) show that males significantly outperformed females in general cognitive ability in the current sample. Thus, it would make sense that females' greater ability in social cognitive tasks specifically might be suppressed when not controlling for general cognitive ability. Considering this possibility, and the fact that few studies looking at sex differences in social cognitive ability have controlled for general intelligence (e.g., Navarra-Ventura et al., 2018), the current study suggests that mixed results in previous research may stem from confounding sex differences in *general* cognitive ability with those specific to *social* cognitive ability. Current findings are consistent with the wealth of literature suggesting that females empathize with others more (Hoffman, 1977; Mestre et al., 2009) and are more accurate at interpreting the emotional states of others (Montagne et al., 2005; Nettle, 2007; Stiller & Dunbar, 2007; Wingenbach et al., 2018). This is also in line with research indicating that females are higher in Agreeableness (e.g., Costa et al., 2001; Weisberg et al., 2011).

Relevance to Psychopathology

Findings from the current study could potentially be extended in future research to benefit understanding of various forms of psychopathology. As previously mentioned, poor ToM performance has been associated with a variety of psychopathology dimensions and disorders such as schizophrenia (Abram et al., 2016; Pedersen et al., 2012), schizotypy (Blain et al., 2017; Bora, 2020), autism (Baron-Cohen et al., 1985; Brune & Brune-Cohrs, 2006), autistic traits (Best et al., 2008; Blain et al., 2017), and Williams syndrome (Tager-Flusberg & Sullivan, 2000). Likewise, low Agreeableness (i.e., Antagonism) has been associated with a host of personality disorders and negative real-life social outcomes (e.g., Anderson et al., 2018; Krueger et al., 2012).

The current results might serve to inform research done in these clinical populations and are consistent with recent dimensional and transdiagnostic frameworks for understanding psychopathology, such as the National Institute of Mental Health's Research Domain Criteria (RDoC; Insel et al., 2010) and the Hierarchical Taxonomy of Psychopathology (HiTOP; Kotov et al., 2017). These frameworks seek to understand psychopathology in terms of underlying dimensions rather than diagnostic categories. Agreeableness and ToM, especially when considering dysfunctionally low levels of functioning, are two such promising dimensions that could be useful in clinical research and practice. Future intervention research could explore ToM deficits and low Agreeableness as transdiagnostic targets for intervention. Likewise, as the default network and its dorsal medial subsystem appear to be implicated in ToM and Agreeableness, neurostimulation research could explore whether electrical or magnetic stimulation of brain regions such as the TPJ and dmPFC might lead to changes in social cognitive ability and relevant social outcomes (Johnson et al., 2013).

Methodological Considerations

Compared to much of the previous work done on the topic, the current study uses a large dataset conferring relatively high statistical power. Though a number of existing studies have investigated possible associations between individual differences in default network function and individual differences in social cognitive ability and related traits, the majority of these studies have used sample sizes in the range of 10 to 70 individuals (Hughes et al., 2019; Inagaki & Meyer, 2020; Kaplan & Iacoboni, 2006; Song et al., 2009; Tamir et al., 2016; Wagner et al., 2011; Waytz et al., 2012; Zhang et al., 2019), which are not optimal for detecting reliable estimates of between-subjects effects (Button et al., 2013). Considering this, the design of the present study should yield more reliable findings and contribute to the robustness of the field (while still keeping in mind the limitations of the current study detailed below).

The approach employed in this study to measure brain function could also benefit future research on the neurobiology of individual differences. I used a network-based approach by incorporating atlases based on patterns functional connectivity in large samples (Schaefer et al., 2018; Yeo et al., 2011). Each participant's data were individually mapped onto a 400-parcel atlas that aligned within 17 broader functional networks identified by Yeo et al. (2011), using GPIP to ensure that parcels were adjusted to the optimal location for each participant. Each of the networks described in Yeo et al. (2011) also summarize regions of the brain that tend to be synchronously active in patterns that can be consistently identified across samples. Understanding how specific cortical parcels map onto these broad networks can be used to unify and better understand previous findings for individual regions of interest from the social neuroscience literature (Tompson et al., 2018).

The current approach should be more effective for studying individual differences that will generalize across samples, compared with the typical use of contrasts for identifying brain regions on a voxel-wise or cluster-wise basis. By focusing on well-established, large-scale brain networks, the identified regions of interest represent a broad set of brain structures with *a priori* relevance for a given construct of interest (in this case, social cognition) rather than specific voxels or clusters that might be most strongly associated with that construct only by chance in any given sample (Vul et al., 2009; Yarkoni, 2009). Moreover, approaches that focus on broad networks may be more reflective of how the brain typically functions, relative to a more localized or modular approach. Brain-behavior associations appear to be more extensive than once believed, in contrast to the relatively small clusters or regions of the brain that are often reported in underpowered samples (Yarkoni et al., 2010). A majority of brain regions are involved in multiple psychological processes, and many psychological processes involve multiple different regions of the brain, not just in the case of social neuroscience (Poldrack, 2014; Yeo et al., 2011). A network-based approach allows researchers to capture a wider picture of brain function and its relation to behavioral constructs of interest; this approach, in conjunction with a large sample size, should lend itself to reproducibility and generalizability (Yarkoni, 2009).

Limitations

Despite the multiple strengths of the current study, there are some important limitations. First, although one advantage of the current study was using multiple tasks in defining the current social cognition accuracy variable, I still only utilized neuroimaging data from a single ToM task in computing neural activation variables. Future research could use an SEM approach to model how variance in brain activity during a variety of different ToM tasks completed in the scanner might predict social cognitive ability and personality. Moreover, although the current data provide evidence that individual differences in default network function are associated with social cognitive abilities and related personality traits, the causal direction and dynamics of these associations cannot be established with the current study design. Future work incorporating methods such as neuromodulation, dynamic causal modeling, experience sampling, and long-term longitudinal data collection could help to more clearly establish causal pathways involved in the neurobiology of personality and individual differences.

Finally, even though I found a significant correlation between ToM-related activation in the dorsal medial subsystem and Agreeableness, this correlation is likely attenuated in its effect size by the personality measure used. Although the NEO-FFI is a reasonably effective brief measure for evaluating the Big Five personality traits, it was not designed to allow for the assessment of personality at lower levels of the hierarchy, including personality aspects and facets. Given possible differential associations of subdimensions within Agreeableness with social cognition and default network function (Allen et al., 2017), a measure that can distinguish between subdimensions of

Agreeableness would be optimal. Future research examining the relation between these variables should include personality measures that can assess personality at multiple levels of the trait hierarchy in order to better discern which specific dimensions contribute to brain-behavior associations.

Chapter 2 Conclusion

Findings in a very large neuroimaging sample confirm and extend the current literature linking ToM, the default network, and Agreeableness. Given that ToM-related activation in prefrontal regions of the dorsal medial subsystem positively predicted both latent levels of Agreeableness and social cognitive ability, it appears that the functions of the default network may help account for the link between Agreeableness and ToM. These findings may inform future research that seeks to understand how normal functioning goes awry in psychopathology involving social deficits, and how individual differences in social cognition and related traits affect real-world relationship success, social network quality, and interpersonal functioning. In sum, the current research furthers work on the neural and personality correlates of individual differences in social cognition while demonstrating effective methods in social cognitive neuroscience research. I recommend that researchers consider using individualized parcellation methods, network-based hypotheses, and latent variable techniques such as SEM, rather than voxel-wise analyses, when designing future studies assessing individual differences.

CHAPTER 3:

Extraversion but not Depression Predicts Reward Sensitivity:

Revisiting the Measurement of Anhedonic Phenotypes

A longstanding debate exists regarding the utility of categorical versus dimensional frameworks for identifying and classifying psychopathology. Historically, the majority of research in psychiatry and clinical psychology has been framed around the categorical diagnoses set forth by the *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; *DSM-5*; American Psychiatric Association, 2013). Despite their influence and popularity, traditional DSM diagnoses have pervasive problems, including heterogeneity within and overlap among diagnostic categories, and much empirical research indicates that continuous, dimensional models of psychopathology are more reliable and valid than traditional categorical frameworks (Kotov et al., 2017; Markon, Chmielewski, & Miller, 2011; Wright et al., 2013).

In the case of depressive disorders, the problems with categorical diagnoses are evident in their high degree of comorbidity with other diagnoses (e.g., anxiety disorders) and the heterogeneity of symptoms used as criteria for diagnosis, which include contrasting symptoms such as hypersomnia and insomnia, or overeating and lack of appetite (Fried, 2017; Hasler et al., 2004; Hyman, 2002). These problems pose challenges not only in the clinic, where they raise barriers for using diagnosis to guide effective treatment, but also in research that attempts to clarify the psychological and biological mechanisms underlying psychiatric disorders. In response to these challenges, continuous, dimensional frameworks account for comorbidity and heterogeneity by using psychopathology-related traits and symptom dimensions that cut across diagnostic categories and are organized hierarchically (e.g., Krueger & Markon, 2014; Kotov et al., 2017). Narrower dimensions at lower levels account for heterogeneity, and they are grouped at higher levels according to their empirical patterns of covariance, which allows for the modeling of comorbidity.

In addition to focusing on transdiagnostic trait and symptom dimensions, another important approach to improving research on psychopathology is identifying and quantifying specific transdiagnostic affective and cognitive mechanisms that might contribute to downstream behavioral dysfunction and clinical symptomatology. This approach is exemplified by the National Institute of Mental Health's Research Domain Criteria (RDoC), which identifies transdiagnostic features of psychopathology at multiple levels of analysis, focusing on the use of behavioral paradigms and investigation of neural circuits (Insel et al., 2010). One example of an important transdiagnostic phenotype that has been researched both as a continuous trait or symptom dimension and in terms of underlying mechanisms is anhedonia, which is related to depression and a variety of other traditional clinical diagnoses. *Anhedonia* can be defined as a relative failure to obtain pleasure from activities, or stimuli, previously experienced as rewarding (Keedwell et al., 2005). Anhedonia has been demonstrated as a vulnerability factor for depressive symptomatology (Loas, 1996; Meehl, 1975) and is elevated among individuals diagnosed with depressive disorders (Snaith, 1993). Various attempts have been made to operationalize anhedonia, including tasks that assess responsiveness to rewarding stimuli

(Costello, 1972) and numerous self-report measures, including subscales derived from the Beck Depression Inventory (BDI; Beck et al., 1961; Beck et al., 1996; Joiner et al., 2003; Pizzagalli et al., 2005). Further, anhedonia has been conceptualized as part of a broader framework of dimensional psychopathology, as measured by instruments such as the Personality Inventory for the DSM-5 (PID-5; Krueger et al., 2012). In addition to being a key feature of depressive disorders, anhedonia is also present in other DSM diagnoses, from bipolar disorder and schizophrenia to various personality disorders (Andreasen et al., 2012; Di Nicola, 2013; Kwapil & Barrantes-Vidal, 2015).

One task developed to operationalize and measure anhedonia is a probabilistic reward task, introduced by Tripp and Alsop (1999) and popularized by Pizzagalli et al. (2005). This task has been referred to both as the Probabilistic Reward Task (PRT; Pizzagalli et al., 2008a, Pizzagalli et al., 2008b) and the Implicit Probabilistic Incentive Learning Task (IPILT; Barch et al., 2017); throughout the rest of this paper, I will use "PRT" to refer to the task. In the PRT, participants are rewarded at a differential frequency for discriminating between long and short mouth stimuli presented on a cartoon face, resulting in a systematic (but not typically conscious) preference for one stimulus over the other. Participants' response bias toward the more frequently rewarded or "rich" stimulus is an index of reward sensitivity, and participants' change in response bias from the beginning to the middle of the PRT is often used as an index of reward learning. Change in response bias was found to correlate negatively with scores on the BDI in an undergraduate convenience sample ($N = 61$) and to differ significantly in that sample between individuals with BDI scores high enough to indicate "mild depression"

($\geq$16, N = 15) and those with low scores (0–6, N = 21) (Pizzagalli et al., 2005). In another study, individuals diagnosed with depressive disorders demonstrated lower reward responsiveness than healthy controls (Pizzagalli et al., 2008b), and first-degree relatives of those with major depression also show lower reward responsiveness (Liu et al., 2015). PRT performance has also been found to predict perceived stress (Pizzagalli et al., 2007) and performance responds to acute stress (Bogdan et al., 2006). Other research has highlighted associations between performance on the PRT and relevant brain systems, such as resting electrical activity in the orbitofrontal cortex (Webb et al., 2016), feedback-related electrical potentials (Bogdan et al., 2011; Bress & Hajcak, 2013; Whitton et al., 2016), reward-related response in the anterior cingulate and basal ganglia (Santesso et al., 2008; Whitton et al., 2016), and dopaminergic functioning (Kaiser et al., 2018; Santesso et al., 2009).

Despite the breadth and apparent consistency of findings, many studies using the PRT have had serious statistical limitations such as small sample sizes and use of dichotomized scores (e.g., Pizzagalli et al., 2005; 2008), both of which reduce statistical power and thus increase the proportion of significant results that are false positives (Cohen, 1983; MacCallum et al., 2002). Furthermore, one of the highest-powered studies published using the PRT failed to show an effect of depression on response bias, comparing patients with major depressive disorder to healthy controls (N = 294; Lawlor et al., 2019) and other recent data fail to support a correlation between response bias and measures of depression, anhedonia, and Neuroticism—a personality trait related to depression (N = 216; Webb et al., 2020). The current study attempted to replicate the correlation between BDI scores and reward sensitivity, reported by Pizzagalli et al. (2005), and to examine potential associations with another personality trait related to depression—low Extraversion (Allen et al., 2017; Kotov et al., 2010). I used a large sample with adequate power to detect correlations in the expected range of effect sizes and to estimate such correlations with reasonable precision (Gignac & Szoderai, 2016; Hemphill, 2003; Richard et al., 2003; Schönbrodt & Perugini, 2013).

In addition to attempting a direct replication of the association between reward sensitivity and depression, I also extended my analyses to consider the likely association of reward sensitivity with relevant personality variables, namely Extraversion and its pathological variants. Research on dimensional approaches to psychopathology suggests that psychiatric symptoms can be described as risky or maladaptive variants of behaviors described by normal personality variation (DeYoung & Krueger, 2018). Most major dimensions of risk for psychopathology appear to reflect the same latent variables as the major dimensions of personality. For instance, maladaptively low Extraversion has been labeled "Detachment." Detachment is a core feature of depressive disorders and is also largely analogous to the negative symptoms of schizophrenia and negative schizotypy (Cicero et al., 2019; Kotov et al., 2016; Kotov et al., 2017).

Depression has been shown to be related to both low Extraversion and high Neuroticism (Allen et al., 2017; Kotov et al., 2010). This combination of low Extraversion and high Neuroticism is reflected in the item content of various self-report measures of depression, particularly for heterogeneous scales such as the BDI. Neuroticism is strongly related to depression and most other forms of psychopathology

(Widiger, 2011), and the negative affect and sensitivity to punishment characterizing Neuroticism is a key component of depression. Nonetheless, depression is also related to lack of reward responsiveness, reduced positive emotionality, and social withdrawal, all of which are components of low Extraversion (DeYoung, 2015; Lucas et al., 2000). Though Extraversion is often considered colloquially as primarily related to sociability, a large body of evidence suggests that the defining characteristic of Extraversion is reward sensitivity generally, not mere sociability, such that extraverts typically have more energy and positive affect than introverts even in nonsocial situations (Corr, 2008; Lucas et al., 2000; Smillie, 2013; Smillie et al., 2007; 2011a; 2011b; 2012; 2019). Thus, measures of Extraversion may show stronger effects than measures of depression when assessing relations with reward sensitivity, anhedonia, and other related variables. In other words, to the extent that depression is associated with reduced reward sensitivity, I hypothesize that this is because depression involves low Extraversion, which is, in theory, the primary manifestation in personality of variation in reward sensitivity.

The current study attempted to replicate findings from an investigation of the association between depression and reward sensitivity in a nonclinical population, while also conducting follow-up and extension analyses testing a more complete model of depression, Extraversion, Neuroticism, and reward sensitivity. The specific findings of Pizzagalli et al. (2005) that I attempted to replicate include (1) that reward learning was evident in the PRT for the sample as a whole, (2) the significant difference in response bias between participants meeting the threshold for "mild depression" in their BDI scores and those with low BDI scores, (3) the significant correlation between a measure of

Melancholic depression derived from the BDI and changes in response bias from Block 1 to 3 of the PRT, and (4) the significant difference in Melancholic depression scores between subjects with negative and positive changes in response bias (Pizzagalli et al., 2005).

First, I hypothesized I would replicate the results of Pizzagalli et al. (2005), in that participants would show a reward-learning effect and depression would be negatively associated with response bias across all blocks of the PRT and with response bias later in the PRT, relative to baseline. I also attempted a conceptual replication of Pizzagalli's depression findings by examining additional measures of Depressivity and Anhedonia from the PID-5. Second, I hypothesized that response bias and response bias relative to baseline would be positively related to Extraversion; I anticipated these effects would be stronger than those for depression, which reflects a combination of Detachment (low Extraversion) and Neuroticism. Consequently, my final hypothesis was that the effects of Extraversion and associated variance in depression on response bias would be more apparent when controlling for variance in Neuroticism.

Method

Participants

A total of 333 participants completed the PRT. Exclusion criteria based on performance were identical to those used by Pizzagalli et al. (2005). Thirty participants were excluded from further analyses because of a high prevalence of reaction time outliers (a total of more than 40 outliers across the task, with outliers being identified as individual trials with a reaction time that did not fall within a range of ±3 SD of a given participant's mean reaction time). Four additional participants were excluded from the analyses for having below-chance accuracy. The final sample consisted of 299 people (148 females) between the ages of 20 and 41 (M = 26.37, SD = 5.12). Participants were recruited from the community surrounding Minneapolis, MN, primarily through online advertisements, and individuals represented a variety of professions, with relatively few students.

All participants completed informed consent and all protocols were approved by the University of Minnesota Twin Cities Institutional Review Board (IRB# 1002M78152, "Neural Mechanisms of Personality in Decision Making"). The current article uses data from a large-scale study on the neurocognitive mechanisms of personality and decision making, conducted at the University of Minnesota Twin Cities. This paper includes the first analyses using a reward sensitivity task from this dataset and is my group's first publication involving results from this task. However, the personality and psychopathology data from this sample have been used in multiple articles that are already published. A list of those articles is publicly available at the Open Science Framework (https://osf.io/qf63r/).[2]

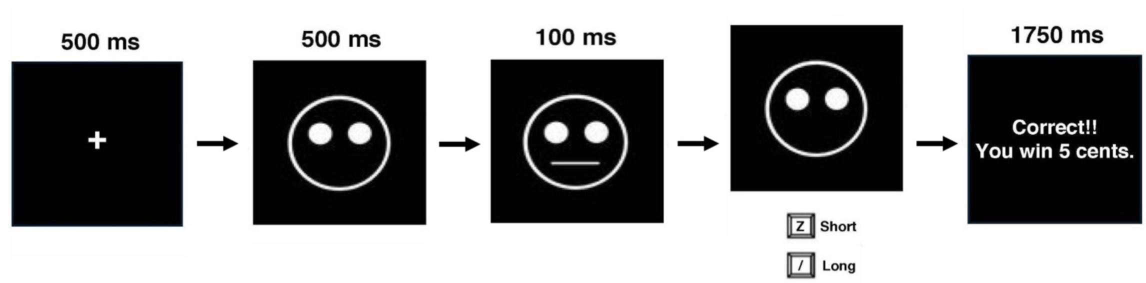
Probabilistic Reward Task (PRT)

Task Description. The PRT is a 25-minute signal detection task that has been validated in multiple previous studies and was designed to assess individuals' implicit

[2] I am unable to provide open access to the data used in this study because consent forms assured participants that their data would not be shared outside of the research team. Analytical scripts, a list of all procedures and measures included in this study, and a list of references also utilizing the same broader dataset are publicly available at the Open Science Framework (https://osf.io/qf63r/).

responsiveness to monetary reinforcements (Bogdan & Pizzagalli, 2006; Pizzagalli et al., 2005; 2008). For each trial, participants were presented with a fixation cross for 500ms, followed by a mouthless cartoon face (for 500ms). Then, either a long (13mm) or short (11.5mm) mouth was presented on the face, for 100ms, and then disappeared. For each trial, participants were then asked to determine whether the mouth presented was short or long. The cartoon face was then presented without the mouth, until the participant used the keyboard to make a response. The long and short mouths were presented equally as often, in a random order, with no more than three sequential presentations of a given mouth stimulus. A total of 300 trials were presented, split into three blocks of 100 trials (hereafter referred to as Block 1 for the first 100 trials and Blocks 2 and 3 for the following sets of 100 trials). A random selection of correct responses received positive feedback, for each of which participants were rewarded with 5 cents. Feedback was always accompanied by the monetary reward and no negative feedback was given. Figure 3.1 illustrates the PRT stimuli and procedure.

Figure 3.1.



Probabilistic Reward Task (PRT) stimuli and procedure

Across all three blocks of the PRT, an asymmetrical reinforcer ratio was used to encourage a response bias toward one of the two mouth stimuli (McCarthy & Davison, 1979; Tripp & Alsop, 1999). Correct identification of the long mouth (rich stimulus) was rewarded three times as often, compared to the short mouth (lean stimulus). In the current study, the long mouth was always rewarded more. In each block, a total of 40 trials received reward feedback, 30 trials for the rich stimulus (long mouth) and 10 trials for the lean stimulus (short mouth). Prior to the administration of the PRT, participants were told to win as much money as possible and that they would not be rewarded for every correct response they made. However, they were not told that there would be a disproportionate ratio of rewards between the two stimuli. The participants' performances were analyzed with respect to their accuracy, discriminability, and degree of response bias formed throughout the PRT.

Some previous research (albeit limited in scope and statistical power) has demonstrated the reliability of the PRT. Split-half reliability calculated using the Spearman-Brown prophecy formula for even and odd trials on the task has been reported at a coefficient of 0.71, in a sample of 294 individuals, demonstrating acceptable internal consistency (Lawlor et al., 2019). Test-retest reliability (i.e., the correlation between participants' performance scores at two different time points, such as test sessions that are a month apart), however, has been evaluated only in very small samples, showing test-retest correlations for response bias of $r = .57$ (N = 25; Pizzagalli et al., 2005) and $r = .62$ (N = 16; Santesso et al., 2008).

Data Collection and Reduction. Primary variables of interest for the PRT included accuracy, discriminability, response bias, and change in response bias. Accuracy was calculated as the percentage of stimuli correctly labeled as long or short for each of the three blocks. Discriminability—log(d)—and response bias—log(b)—variables were calculated using signal detection principles (McCarthy & Davison, 1979; Tripp & Alsop, 1999). Discriminability represents participants' tendency to correctly distinguish between stimuli after controlling for bias, whereas response bias represents a tendency to select one stimulus over the other (Pizzagalli et al., 2005).[3]

$$\boldsymbol{log(d) = \frac{1}{2}log(\frac{Long_{correct} \times Short_{correct}}{Long_{incorrect} \times Short_{incorrect}})}$$

$$\boldsymbol{log(b) = \frac{1}{2}log(\frac{Long_{correct} \times Short_{incorrect}}{Long_{incorrect} \times Short_{correct}})}$$

Discriminability and response-bias variables were calculated for each block, after which, main variables for discriminability and response bias were calculated by averaging values from Blocks 2 and 3, given that Block 1 was considered a primary learning phase. Finally, two change-in-response-bias variables were calculated as the difference in participants' response bias from Block 1 to Block 2 and from Block 1 to Block 3. Descriptive statistics for these variables are shown in Table 3.1.

Importantly, when using the performance measures derived from PRT, there are three primary ways to operationalize reward sensitivity: (1) response bias throughout all three blocks of the task, (2) difference scores for response bias from Block 1 to Blocks 2

---

[3] As is standard when using log transformations—because logarithmic functions are undefined for values of zero—a constant of .5 was added to all variables before they were entered into the log(d) and log(b) formulas (Brown & White, 2005). It is also worth noting that these formulas differ slightly from the traditional signal detection measures of d' and c (Green & Swets, 1966; Stanislaw & Todorov, 1999), as I use logit transformations rather than inverse normal transformations (McCarthy & Davison, 1979); values obtained, however, yield similar results and are related almost linearly (Brown & White, 2005; Brown & White, 2009). The current formulas were used to facilitate consistency with previous research using the PRT (Pizzagalli et al., 2005; Tripp & Alsop, 1999).

or 3, and (3) response bias in Blocks 2 and 3 of the task, controlling for Block 1 response bias (i.e., baseline). As response bias in the PRT is a measure of how differential reward frequencies influence participants' tendency to select one stimulus over the other, levels of response bias throughout the PRT can be interpreted as an index of reward sensitivity (i.e., participants with a greater tendency to select the more-rewarded stimulus have a higher response bias, which is interpreted as higher reward sensitivity). In addition to showing greater bias throughout the entire task, I expect individuals more sensitive to reward cues to develop a greater response bias over the course of the task (i.e., to show greater levels of reward learning). This reward learning effect is another way to define reward sensitivity in the PRT and can be operationalized using either difference scores (differences in response bias from Block 1 to Block 2 or 3) or by controlling for bias in Block 1 when predicting bias in Blocks 2 and 3.

Controlling for baseline is often preferable to using difference scores, when examining associations with individual-difference variables. Difference scores do not capture any information about the association between baseline scores and later scores, instead imposing a linear restriction on their slopes when predicting outcome variables; thus, the specific effect for baseline scores vs. scores at a second time point cannot be identified when using difference scores (Allison, 1990; Edwards, 1996; Whittenborn, 1951). In other words, if one finds an association with a difference score, one cannot tell whether the effect is due to variation in the baseline condition or to variation in the condition of interest. Controlling for baseline by partialling out variance in baseline performance from variance in the condition of interest is useful because difference scores are typically dependent on and correlated with baseline scores (Allison, 1990; Edwards, 1994; Edwards, 1996; Whittenborn, 1951). In the current design, controlling for Block 1 values of response bias when examining associations of a given variable with bias in Blocks 2 and 3 allows us to examine participants' deviation from their expected level of response bias relative to other participants. Such models yield outcome variables that are often more meaningful and informative than simple difference scores (Edwards, 1994; Edwards, 1996; Whittenborn, 1951) and essentially involve examining associations with rank-order change, rather than absolute change. In the current dataset, response bias in Block 1 was highly negatively correlated with difference scores in bias from Block 1 to 2 ($r = -.52$, $p < .001$), suggesting that controlling for baseline might be a better approach than using absolute difference scores. Nonetheless, I also present analyses using difference scores for my direct replication aims, in an effort to mirror Pizzagalli's original study (2005).

Questionnaire Measures

Participants were administered a variety of questionnaires to measure psychopathology and personality. Multiple measures of depression and of the Big Five were administered to facilitate the creation of latent variables. Peer reports were also collected for Big Five measures, from people who knew participants well, and at least one peer report was available for 236 participants. When multiple peers provided ratings for a given participant, they were averaged to create a single peer-report score.

Beck Depression Inventory (BDI-II). The BDI-II is a 21-item, 4-point Likert-format (0 for *symptom absent* and 3 for *severe symptoms*) self-report inventory used to assess presence and severity of depressive symptoms (Beck et al., 1996). The BDI-II exhibits high internal consistency, as well as external validity in predicting clinician ratings and scores on other validated depression measures, in both clinical and general population samples. In addition to overall scores, two sub-scores were created from the BDI data, in order to replicate analyses conducted by Pizzagalli et al. (2005). First, a score was computed for BDI items associated with anhedonic symptoms ("BDI anhedonic sub-score"): loss of pleasure (item #4), loss of interest (item #12), loss of energy (item #15), and loss of interest in sex (item #21) (Joiner et al., 2003). An additional sub-score was computed for *melancholic depression* (Pizzagalli et al., 2004)—a subtype of major depressive disorder characterized by pervasive anhedonia (Rush & Weissenburger, 1994)—by summing scores of BDI items that map onto the DSM-IV criteria for melancholia: loss of pleasure (item #4), guilty feelings (item #5), agitation (item #11), loss of interest (item #12), early morning awakening (item #16b), and loss of interest in sex (item #21) (Pizzagalli et al., 2004). All BDI scores were logarithmically transformed to approximate normality, as they showed original skew values greater than 1.0 (Table 3.2). Results of all analyses were, however, substantively equivalent, whether or not BDI scores were log transformed.

Personality Inventory for DSM-5 (PID-5). The PID-5 (Krueger et al., 2012) questionnaire includes 220 items rated on a 4-point Likert scale (between 0 for *very false or often false* and 3 for *very true or often true*). This inventory was designed to measure maladaptive traits that are symptoms of personality disorder in the alternative model of personality disorder for the DSM-5. The PID-5 comprises 25 primary trait scales that are grouped into five higher-order dimensions of Negative Affectivity, Detachment, Psychoticism, Antagonism, and Disinhibition (Krueger et al., 2012). For the present study, I used the Depressivity and Anhedonia scales, and scores were logarithmically transformed to approximate normality (original skewness = 1.43 & 1.65, respectively).

Big Five Aspect Scales (BFAS). The BFAS (DeYoung, Quilty, & Peterson, 2007) consists of 100 items that require response ratings based on a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The questionnaire subdivides each of the Big Five into two component aspects (DeYoung et al., 2007), each assessed by 10 items, which can be averaged to generate 20-item Big Five scores. In the present study I used scores for Extraversion and Neuroticism.

Big Five Inventory (BFI). The BFI (John, Naumann, & Soto, 2008) was used to evaluate participants based on the Big Five factors of personality—Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience. This measure consists of 44 items scored on a 5-point Likert scale (1 for *disagree strongly* and 5 for *agree strongly*). I used scores for Extraversion and Neuroticism.

Analyses

Statistical Power. Pizzagalli et al. (2005) found a negative correlation between change in response bias and melancholic depression (r = -.28, p = .035, N = 61). They also found a range of associations between their response-bias variables and self-report measures of depression and anhedonia, including measures taken at a follow-up visit. One effect they reported between response bias and total depression at follow-up (r = -.46, p < .025, N = 25) is surprisingly large, considering that few variables that do not share method variance are correlated at this magnitude (Hemphill, 2003). Such large effects are likely to be inflated due to sampling variability in small samples. Given the current sample of 299 and an alpha threshold of .05, I had 90% statistical power to detect a correlation of ± .19 or stronger, and 80% power to detect a correlation of ± .16 or stronger.

Effects of Task Manipulation. A two-way repeated-measures ANOVA was conducted on accuracy, using block and type of stimulus as within-subject factors. Additionally, one-way repeated-measures ANOVAs were conducted separately for discriminability and response bias across each of the three blocks. In each instance, these ANOVA models were followed by dependent-samples t-tests. Task performance variables, across the three blocks, are visualized using bar plots with error bars representing standard error of the mean (Figure 3.2).

Direct Replication of Depression Associations. Repeated measures ANOVA models were conducted to test the interaction of depressivity-by-block on response bias. Two models were created using BDI as either a categorical variable with two levels or a continuous variable. Matching the criteria used by Pizzagalli et al. (2005), the level of high BDI consisted of cases with total BDI scores greater than or equal to 16, and low BDI consisted of cases with a total score less than or equal to 6. My second model, using continuous BDI scores, was incorporated to avoid the loss of statistical power using the extreme groups ANOVA as employed in Pizzagalli's original analysis. Correlations were computed for each task performance variable and total BDI scores, as well as the two BDI subscales for Melancholic and Anhedonic depression. Finally, an independent-samples t-test was used to test the difference in BDI scores for participants with positive vs. negative response biases.

Follow-up Depression Analyses. In addition to my direct replications of the association of BDI scores with response bias, I used similar repeated-measures ANOVA models, t-tests, and bivariate correlation analyses in conjunction with the Anhedonia and Depressivity scales of the PID-5. Following these tests, structural equation modeling was used to examine relations between latent factors for depression and response bias, thereby removing error variance associated with individual scales and allowing a more powerful test of the association between depression and reward sensitivity. Scores for BDI, PID-5 Anhedonia, and PID-5 Depressivity were used as indicators of a latent variable for depression, while response-bias values for each of the three blocks were used as indicators of a latent response-bias variable. Two models were fit to examine the prediction of (1) response bias across all three blocks, by latent depression, and (2) response bias in Blocks 2 and 3, by latent depression, controlling for Block 1 response bias. These two models allow us to examine two different operationalizations of reward sensitivity, one focusing on participants' general tendency to select the more frequently reward stimulus across all blocks and the other focusing on how this tendency develops throughout the task, relative to participants' baseline levels. Common fit indices were computed for all structural equation models.

Extension Analyses of Personality and Task Performance. To assess the relations of PRT performance and latent personality factors, an additional series of structural equation models was computed. All models were constructed using full information maximum likelihood estimation to allow use of peer ratings despite missing data for some participants. Personality measures were allowed to load on latent factors of Extraversion and Neuroticism, and additional latent variables were used to model peer-report method effects (for Extraversion and Neuroticism)[4]. For each of my models, I ran two versions of the model (mirroring the above models for depression), one in which the criterion latent variable was made up response bias of indicators from each of the three blocks and one in which the criterion latent variable was made up of response bias indicators from only Blocks 2 and 3 and response bias from Block 1 was included as a predictor variable (allowing me to control for participants' baseline response bias). The first set of models examine the effects of Extraversion. Then, I examine the effects of Extraversion and Neuroticism, while modeling their associated variance in depression using a hierarchically nested latent variable, indicated by BDI and PID-5 Depressivity and Anhedonia and loading onto both Neuroticism and Extraversion. In the construction of models including both Extraversion and Neuroticism, the residual variances of peer and self-report versions of the same scales, for Extraversion and Neuroticism, were allowed to correlate. The latent variables of Extraversion and Neuroticism were allowed to correlate, as were the methods factors of Peer-Extraversion and Peer-Neuroticism. Finally, I ran supplemental versions of the Extraversion and Neuroticism models that did

[4] These peer-report factors are not factors of substantive theoretical interest to be included as predictors of reward sensitivity but rather are methods factors to account for shared method variance among the given peer report measures. Thus, their correlations with the predictor variables of interest were set to zero and they were not used as predictors of response bias in the models. The shared variance of self- and peer reports is already captured by the primary E (and N) latent variables, as the peer-report measures have loadings on both the peer methods factors and the primary personality latent variables.

not include a hierarchically nested Depression latent variable, which appear in the appendix.

Results

Effects of Task Manipulation

Descriptive statistics for PRT performance variables are reported in Table 3.1. Discriminability was not correlated with response bias ($r = -.06$, $p = .31$) or change in response bias ($r = .07$, $p = .20$).

Table 3.1. *Chapter 3 performance variables by block (means and standard deviations)*

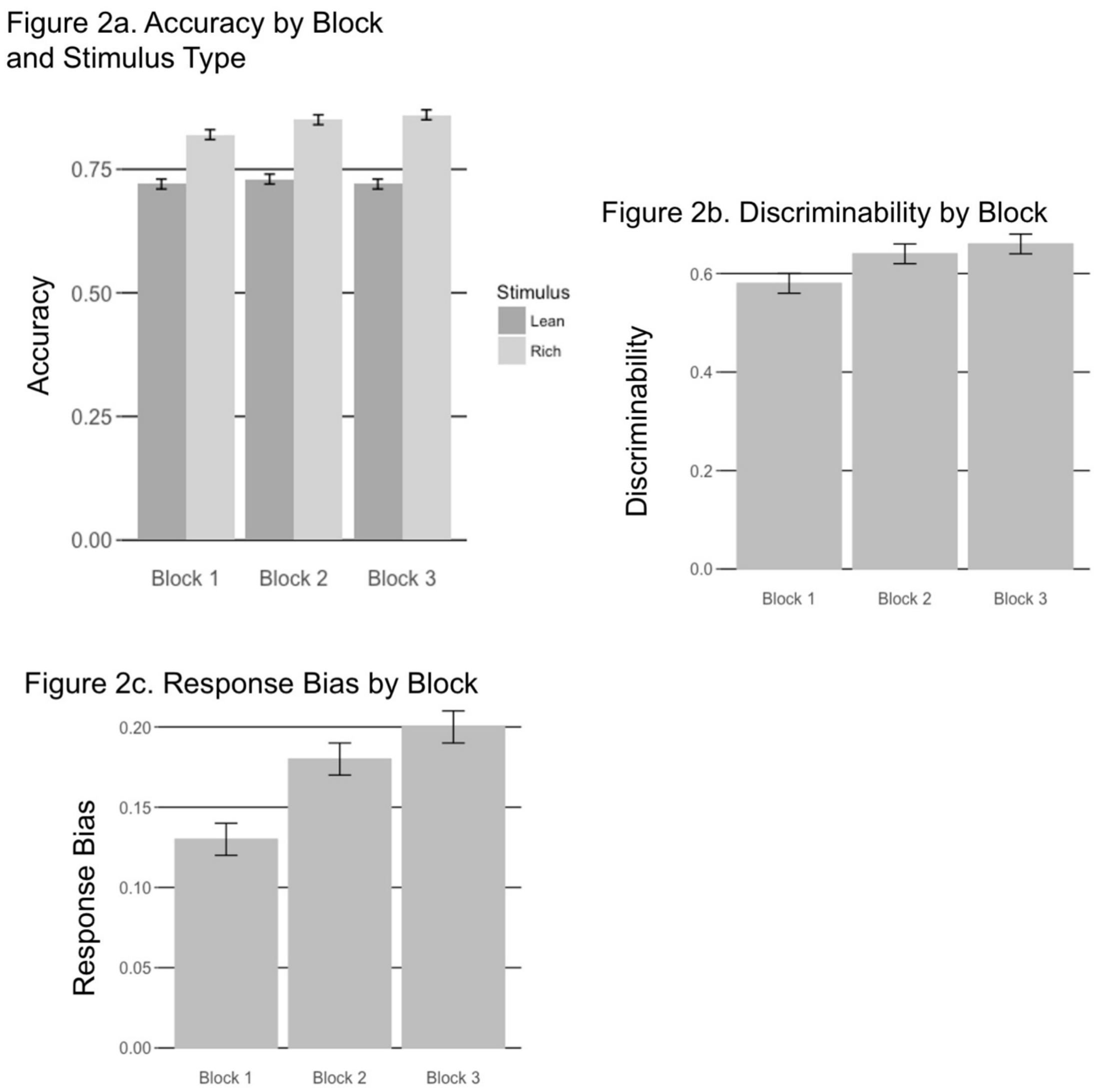
Variable	Block 1	Block 2	Block 3
Rich Accuracy	0.82 (0.11)	0.85 (.09)	0.86 (0.08)
Lean Accuracy	0.72 (0.13)	0.73 (0.13)	0.72 (0.15)
Accuracy	0.77 (0.10)	0.79 (0.09)	0.79 (0.10)
Discriminability	0.58 (0.27)	0.64 (0.28)	0.66 (0.30)
Response Bias	0.13 (0.19)	0.18 (0.19)	0.20 (0.21)
Rich RT (ms)	546 (155)	519 (151)	521 (150)
Lean RT (ms)	577 (166)	560 (166)	567 (172)
RT (ms)	562 (158)	539 (156)	544 (157)

Accuracy (Figure 3.2a). A two-way repeated measures ANOVA was computed for accuracy, using block and stimulus type as a within-subjects factor. There were significant main effects of block ($F_{(2, 297)} = 8.18$, $p < .001$) and stimulus type ($F_{(1, 298)} = 391.72$, $p < .001$), as well as a significant block-by-stimulus interaction ($F_{(2, 297)} = 10.61$, $p < .001$). Compared to accuracy in Block 1, overall accuracy was higher in Blocks 2 ($t_{(298)} = -3.71$, $p < .001$) and 3 ($t_{(298)} = -3.88$, $p < .001$). Accuracy did not significantly increase from Block 2 to 3 ($t_{(298)} = -0.60$, $p = .55$). Across blocks, accuracy was higher for the rich stimulus compared to the lean stimulus ($t_{(298)} = 19.25$, $p < .001$).

Discriminability (Figure 3.2b). A repeated measures ANOVA was computed for discriminability, using block as a within-subjects factor. There was a significant effect of block on discriminability ($F_{(2, 297)} = 13.83$, $p < .001$), with discriminability increasing significantly from Block 1 to Blocks 2 ($t_{(298)} = -2.63$, $p = .009$) and 3 ($t_{(298)} = -3.40$, $p < .001$), but not from Block 2 to Block 3 ($t_{(298)} = -0.84$, $p = .40$).

Response Bias (Figure 3.2c). A repeated measures ANOVA was computed for response bias, using block as a within-subjects factor. There was a significant main effect of block on response bias ($F_{(2, 297)} = 12.15$, $p < .001$), namely, response bias increased significantly from Block 1 to Blocks 2 ($t_{(298)} = -2.88$, $p = .004$) and 3 ($t_{(298)} = -4.03$, $p < .001$), but not from Block 2 to Block 3 ($t_{(298)} = -1.26$, $p = .21$).

Figure 3.2.



Effects of reward task manipulation on performance.

Attempted Direct Replication of Depression Associations

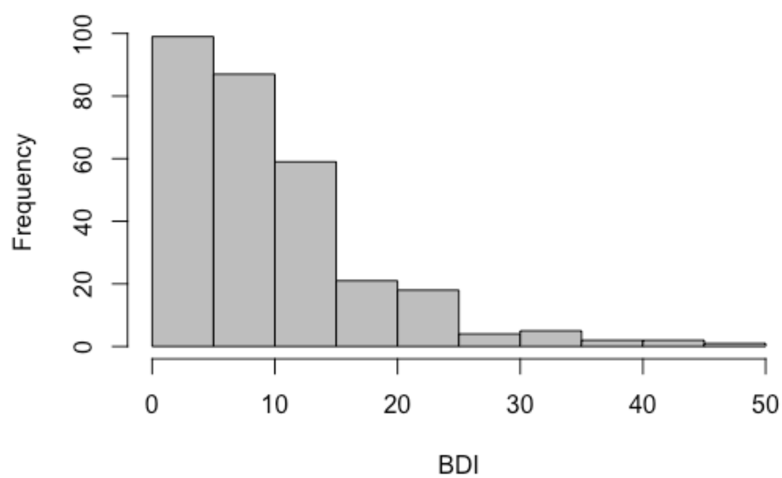
Descriptive statistics and measures of internal consistency reliability—Cronbach's α (Cronbach, 1951) and ω_t (McDonald, 1999; Revell & Condon, 2019)—for self- and peer-report measures are reported in Table 3.2.

Table 3.2. *Descriptive statistics for Chapter 3 questionnaire measures*

Measure	Mean (SD)	Skew	Range	Cronbach's α	ω_t
Self-Report					
BDI Total	12.00 (8.08)	1.56	46.00	.90	.90
BDI Anhedonic	1.65 (1.76)	1.42	9.00	.66	.67
BDI Melancholic	4.14 (2.49)	1.13	13.00	.61	.71
PID-5 Depressivity	1.43 (0.50)	1.60	3.00	.91	.92
PID-5 Anhedonia	1.65 (0.54)	1.19	2.88	.84	.86
BFAS Neuroticism	2.59 (0.67)	.26	3.55	.90	.90
BFAS Withdrawal	2.70 (0.71)	.23	3.70	.81	.82
BFAS Volatility	2.49 (0.80)	.44	3.90	.89	.90
BFI Neuroticism	2.56 (0.78)	.15	3.75	.85	.85
BFAS Extraversion	3.69 (0.57)	-.42	3.40	.86	.87
BFAS Assertiveness	3.63 (0.65)	-.43	3.30	.83	.84
BFAS Enthusiasm	3.74 (0.73)	-.51	3.70	.84	.85
BFI Extraversion	3.30 (0.82)	-.13	4.00	.87	.88
Peer-Report					
BFAS Neuroticism	2.61 (0.61)	.21	2.82		
BFAS Withdrawal	2.58 (0.59)	.31	2.73		
BFAS Volatility	2.63 (0.73)	.26	3.35		
BFI Neuroticism	2.63 (0.71)	.23	3.71		
BFAS Extraversion	3.75 (0.47)	-.42	2.62		
BFAS Assertiveness	3.67 (0.52)	-.20	2.90		
BFAS Enthusiasm	3.83 (0.58)	-.71	3.30		
BFI Extraversion	3.56 (0.68)	-.09	3.46		

The distribution of BDI scores is shown in Figure 3.3. Of note, 61 of the 299 participants (20%) had 'elevated BDI scores' using the criterion of total score ≥ 16, used by Pizzagalli et al. (2005); the original study had 15 of 62 participants (24%) in this BDI score range. A total of 77 participants had BDI scores that were indicative of at least mild depression, using Beck's original suggested cut-off ranges (Beck et al., 1996).

Figure 3.3.



Beck Depression Inventory (BDI-II) score distribution.

Repeated measures ANOVA models were computed to determine whether there was a block-by-depressivity interaction on response bias. No significant interaction was found between block and BDI group ($F_{(2, 161)} = 0.47$, $p = .63$). Because this test reduces power relative to treating BDI score as a continuous variable and using the whole sample, I also conducted a repeated measures ANOVA including block and continuous BDI scores as predictors of response bias, for which there was also no significant interaction ($F_{(2, 296)} = .16$, $p = .69$). Pearson correlations were used to further investigate associations between PRT performance variables and BDI. No correlations were significant at an alpha level of .05 (Table 3.3).

Table 3.3. *Pearson correlations between Chapter 3 depression-related variables and task performance*

Measure	Accuracy	Discriminability	Response Bias	Δ RB1-2	Δ RB1-3
BDI	-.07	-.03	-.07	.01	-.02
BDI Anhedonic	-.05	.04	-.05	.04	-.02
BDI Melancholic	-.08	-.02	-.09	.02	.00
PID-5 Depressivity	-.05	-.07	-.05	-.02	-.09
PID-5 Anhedonia	-.07	-.05	-.07	-.02	-.06

Of particular note, the 95% confidence interval around the correlation between change in response bias and melancholic depression in the current sample, [-.13, .13], did not contain -.28, the correlation detected and presented as a key finding in Pizzagalli et al.'s (2005) original study. Finally, contrary to the original findings, individuals with positive vs. negative changes response-bias showed no significant differences in overall BDI ($t_{(296)} = 0.18$, $p = .85$) or its Anhedonic ($t_{(296)} = -0.35$, $p = .72$) or Melancholic sub-scores ($t_{(296)} = -0.06$, $p = .95$).

Follow-up Depression Analyses

After testing for direct replication of the associations between PRT performance and BDI, I followed up these analyses with similar tests for the Depressivity and Anhedonia scales of the PID-5. There were no significant associations between these PID-5 scales and PRT performance (Table 3.3).

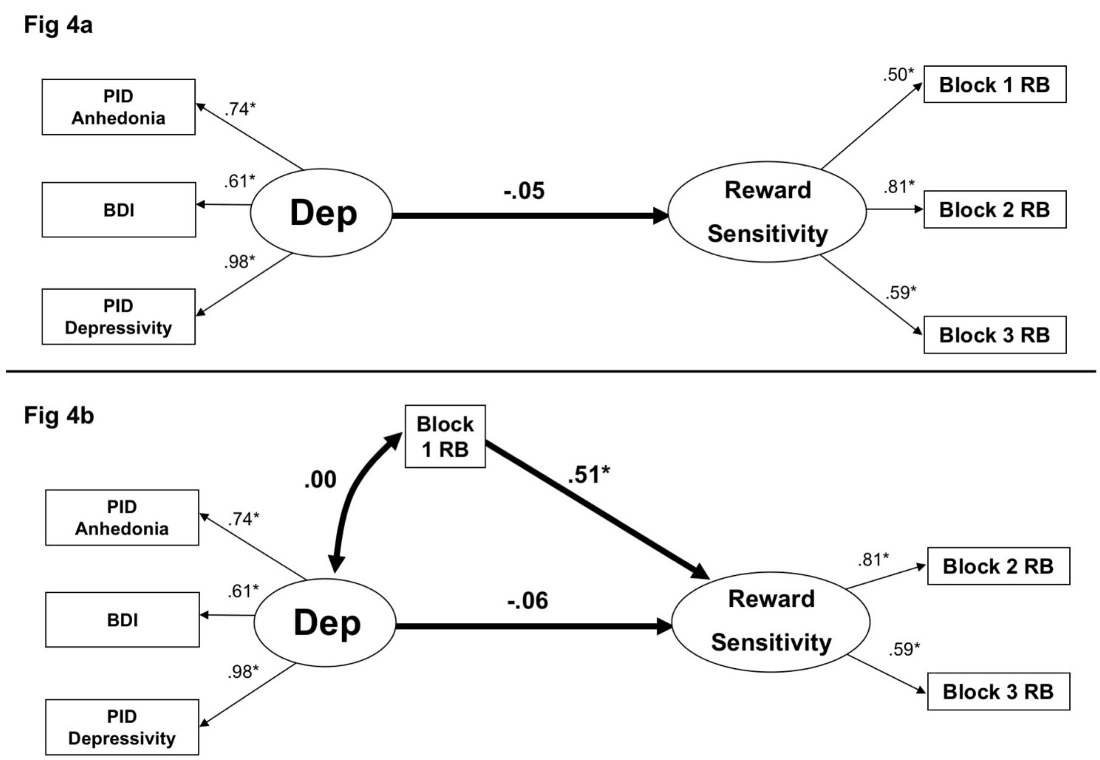
Similarly, there was no significant interaction of task block with PID-5 Depressivity ($F_{(2, 292)} = 1.51$, $p = .22$) or PID-5 Anhedonia ($F_{(2, 292)} = 0.63$, $p = .43$), in predicting response bias. Finally, individuals with positive vs. negative response-bias values showed no significant differences in PID-5 Depressivity ($t_{(292)} = 0.68$, $p = .50$) or PID-5 Anhedonia ($t_{(292)} = 1.00$, $p = .32$).

Table 3.4. *Fit statistics for Chapter 3 structural equation models*

Model	RMSEA	95% C.I.	χ^2	p	TLI	CFI
Depression and Response Bias	.000	[.000, .020]	3.1	.927	1.0	1.0
Depression and Response Bias Controlling for Baseline	.000	[.000, .031]	2.8	.900	1.0	1.0
E and Response Bias	.078	[.046, .111]	31.0	.001	.957	.978
E and Response Bias Controlling for Baseline	.081	[.048, .115]	910.9	.001	.954	.978
E, N, Depression, and Response Bias	.061	[.046, .075]	134.3	< .000	.959	.971
E, N, Depression, and Response Bias Controlling for Baseline	.062	[.047, .076]	132.9	< .001	.957	.971
E, N, and Response Bias	.028	[.000, .053]	40.6	.169	.993	.996
E, N, and Response Bias Controlling for Baseline	.030	[.000, .055]	39.2	.148	.992	.995

Next, I used structural equation models to extend the analyses from Pizzagalli's original study. All structural equation models had acceptable fit, as indicated by RMSEA values less than .085 and TLI values greater than .950 (Table 3.4). Results of a structural equation model predicting reward sensitivity, modeled as the shared variance of response bias across the three blocks from shared variance in BDI, PID-5 Depressivity, and PID-5 Anhedonia, are displayed in Figure 3.4a. Latent depression was not a significant predictor of Reward Sensitivity, as modeled using shared variance in response bias across all three blocks (β = -.05, 95% CI [-.19, .09]; Figure 3.4a). Depression also did not significantly predict the shared variance of response bias in Blocks 2 and 3, when controlling for response bias from Block 1 (β = -.06, 95% CI [-.19, .07]; Figure 3.4b).

Figure 3.4.



Structural equation models of depression predicting response bias (a) and response bias controlling for baseline (b).

Extension Analyses of Personality and Task Performance

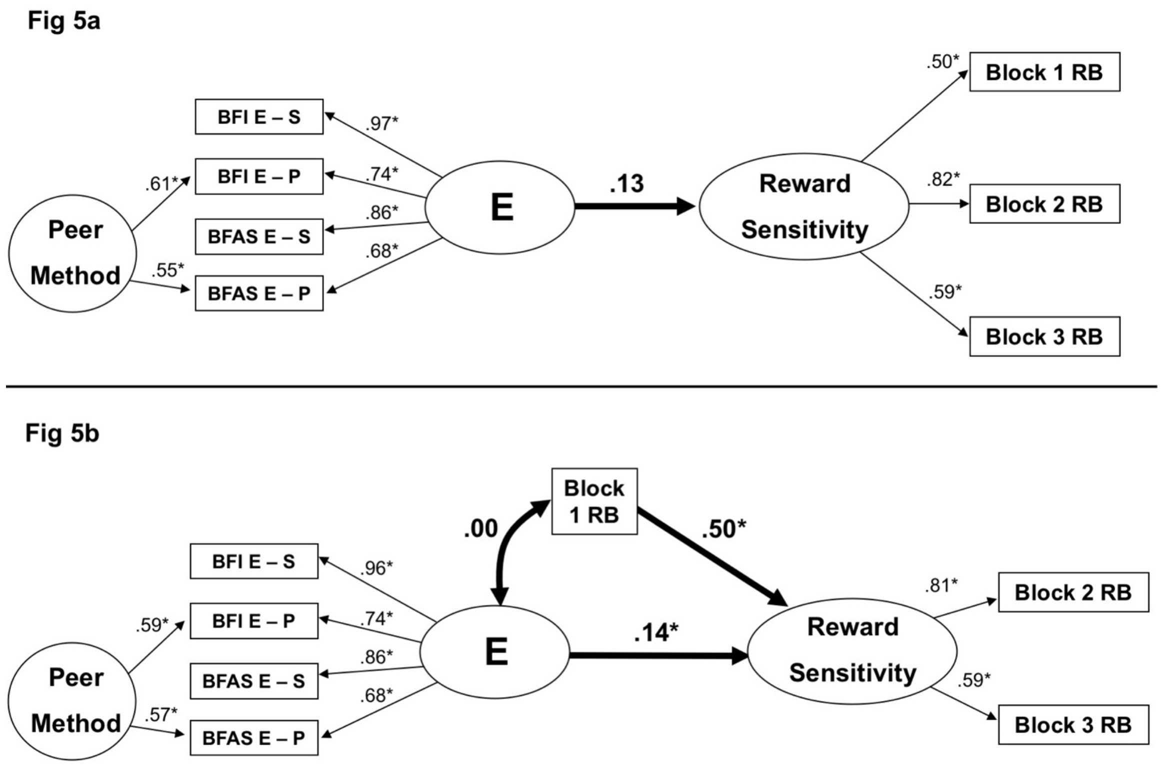
Pearson correlations between measures of personality and PRT performance variables are presented in Table 3.5. Several Extraversion variables showed significant positive correlations with change in response bias and with response bias aggregated over Blocks 2 and 3 (Table 3.5). Next, I used structural equation models to examine the effects of latent Extraversion on Reward Sensitivity as operationalized using response bias aggregated across all blocks (Figure 3.5a) and bias in Blocks 2 and 3 controlling for Block 1 (Figure 3.5b). Extraversion was a significant positive predictor of response bias in Blocks 2 and 3, controlling for bias in Block 1 ($\beta = .14$, 95% CI [.02, .27]; Figure 3.5b), but the association between Extraversion and response bias across all three blocks did not reach statistical significance ($\beta = .13$, 95% CI [-.01, .26]; Figure 3.5a); nonetheless, the associations for both models were both in the positive direction and were nearly identical in their magnitude, as evidenced by highly overlapping confidence intervals.

Table 3.5. *Pearson correlations of Chapter 3 task performance with Extraversion and Neuroticism*

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	RB	Accuracy	Discriminability	ΔRB 1-2
1. BFAS Extraversion	1														.09	-.03	.00	.06
2. BFI Extraversion	.82	1													.11	-.07	-.07	.12
3. BFAS Extraversion - Peer	.62	.62	1												.06	.10	.11	.16
4. BFI Extraversion - Peer	.58	.69	.83	1											.11	.00	.03	.19
5. BFAS Neuroticism	-.39	-.30	-.20	-.17	1										.02	-.07	-.05	.01
6. BFI Neuroticism	-.37	-.31	-.22	-.19	.81	1									.02	-.05	-.05	-.01
7. BFAS Neuroticism - Peer	-.19	-.13	-.34	-.26	.55	.57	1								.10	-.13	-.12	.02
8. BFI Neuroticism - Peer	-.26	-.24	-.39	-.33	.55	.63	.91	1							.03	-.09	-.10	-.04
9. BDI	-.22	-.24	-.25	-.16	.47	.47	.29	.35	1						-.07	-.08	-.08	-.07
10. PID-5 Depressivity	-.44	-.37	-.33	-.20	.59	.63	.37	.40	.59	1					-.05	-.06	-.06	-.05
11. PID-5 Anhedonia	-.59	-.53	-.45	-.35	.53	.52	.38	.41	.45	.73	1				-.07	-.06	-.06	-.02
12. RB Block 1	.00	-.02	-.08	-.08	.11	.12	.09	.09	-.03	.00	-.01	1				-.10	-.07	-.54
13. RB Block 2	.07	.10	.09	.12	.13	.11	.11	.04	-.02	-.02	-.03	.41	1			.01	.14	.55
14. RB Block 3	.08	.03	.00	.04	.04	.07	.11	.03	-.05	-.10	-.08	.30	.48	1		-.18	-.07	.13
Mean	3.69	3.30	3.75	3.56	2.59	2.56	2.61	2.63	12.0	1.43	1.65	.13	.18	.20	.19	.79	.65	.18
SD	.57	.82	.47	.68	.67	.78	.61	.71	8.08	.50	.54	.19	.19	.21	.17	.09	.27	.19

Notes. N = 299 (236 for peer-report measures). Correlations of r > .11 are significant (at an α of .05) for all variable pairs not including peer reports, and variable pairs including peer reports are significant when r > .12. BFAS = Big Five Aspect Scales, BFI = Big Five Inventory, BDI = Beck Depression Inventory, PID-5 = Personality Inventory for DSM-5, RB = response bias, ΔRB = change in response bias.

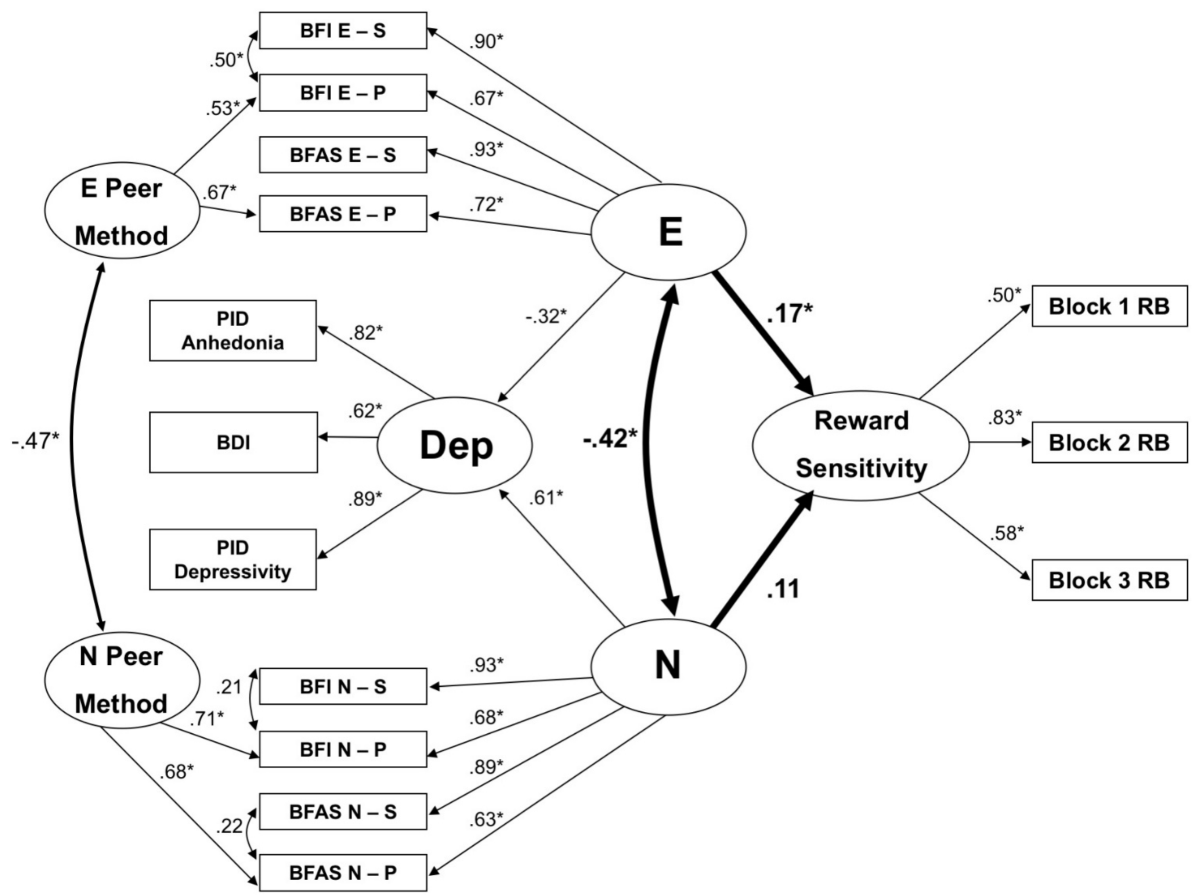
Figure 3.5.



Structural equation models of Extraversion predicting response bias (a) and response bias controlling for baseline (b).

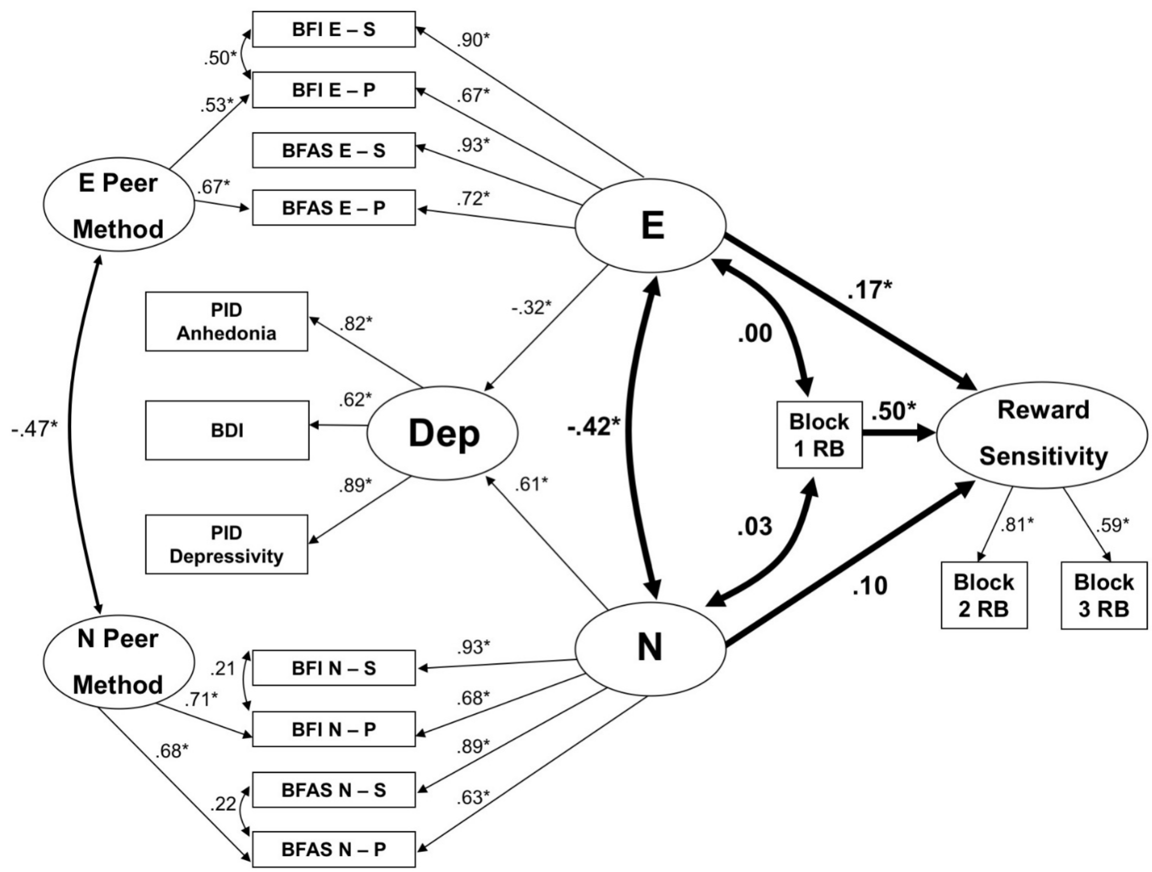
Next, I ran models predicting Reward Sensitivity from both Extraversion and Neuroticism, including Depression as a hierarchically nested latent variable loading onto both personality traits (Figures 3.6 and 3.7). Latent Depression had a negative loading from Extraversion (λ = -.32, 95% CI [-.43, -.22]) and a positive loading from Neuroticism (λ = .61, 95% CI [.52, .70]). Extraversion (and associated variance in depression) positively predicted Reward Sensitivity modeled both as the shared variance of response bias across all three blocks (β = .17, 95% CI [.01, .32]; Figure 3.6) and as response bias in Blocks 2 and 3, using Block 1 bias as a covariate (β = .17, 95% CI [.03, .32]; Figure 3.7). Neuroticism (and associated variance in depression) did not significantly predict response bias in either model. Results were substantively similar using models that did not include a hierarchically nested Depression latent variable, with Extraversion showing positive associations with response bias, whether or not Block 1 bias was controlled for.

Figure 3.6.



Model of Extraversion, Neuroticism, and response bias.

Figure 3.7.



Model of Extraversion, Neuroticism, and response bias controlling for baseline.

Discussion

The first aim of the current study was to replicate previous work linking performance on the PRT with depressive symptomatology, as measured by the Beck Depression Inventory (Pizzagalli et al., 2005). Specifically, I hypothesized that BDI scores would be negatively associated with response bias, change in response bias, and response bias controlling for baseline. However, I found no associations between response-bias variables and BDI and replicated none of the significant effects from the original study. Neither were there significant associations with the Depressivity or

Anhedonia scales of the PID-5, which I used to test a conceptual replication. The structural equation models testing the effects of depression on PRT performance also showed no effects. It is worth noting, however, that the current participants' group-level response to the PRT manipulation did replicate previous findings, as participants, on average, did develop a response bias toward the more frequently rewarded stimulus, and the strength of this bias increased across the three blocks (Pizzagalli et al., 2005).

Given the much greater statistical power of the current study than the original, these null findings suggest that the original results may have been false positives. The finding from the original study that the Melancholic subscale of the BDI was significantly correlated with response bias was based on a sample size of 61 individuals (Pizzagalli et al., 2005), a sample with low statistical power for detecting all but the largest effect sizes regularly observed in individual differences research (Gignac & Szodorai, 2016; Richard et al., 2003). Moreover, some findings presented in the original study were detected after dichotomizing key variables, which can lead to severe reduction of statistical power (Cohen, 1983; MacCallum et al., 2002), thereby increasing the likelihood that any significant results are false positives.

The Role of Extraversion in Reward Sensitivity and Depression

Although the association of reward sensitivity with depressive symptomatology did not replicate in the current study, I did find support for my hypothesis that reward sensitivity would be associated with Extraversion. In an effort to extend the results of previous work by Pizzagalli et al. (2005) and to integrate them with current theory and research on personality and dimensional models of psychopathology, I tested associations of PRT performance with personality dimensions related to depression, hypothesizing that reward sensitivity—modeled both as response bias and as response bias controlling for baseline—would be positively associated with levels of Extraversion but not with levels of Neuroticism. My hypotheses were largely confirmed (though one of the six SEMs testing the relation between Extraversion and reward sensitivity did not quite reach statistical significance) consistent with research showing associations between Extraversion and reward responsiveness as measured in other behavioral tasks (Ávila & Parcet, 2002; Robinson, Moeller, & Ode, 2010). The association of Extraversion with reward sensitivity evident in behavioral tasks aligns well with evidence that variation in the brain's reward system is a major neural correlate of Extraversion (Allen & DeYoung, 2017; Smillie & Wacker, 2014; Smillie et al., 2019). Theories of the biological basis of Extraversion emphasize the role of the dopaminergic incentive reward system (DeYoung, 2013), which is involved in the kind of reward learning that occurs in the PRT (Depue & Collins, 1999).

Findings suggest that levels of Extraversion, not depressivity, are associated with reward sensitivity. More precisely, any association that depressivity has with reward sensitivity is likely to be due to its association with Extraversion. The models in Figures 6 and 7 do imply at least a weak association between depression and reward sensitivity because depression is an indicator of Extraversion (though not as strongly as it is an indicator of Neuroticism). However, it also suggests that investigations into reward-function deficits as a transdiagnostic factor underlying depression would be better off investigating the Detachment (low Extraversion) symptom dimension specifically, rather than focusing on depression symptoms more broadly. Indeed, the association between Extraversion and reward sensitivity becomes stronger after controlling for variance in Neuroticism, suggesting that reward sensitivity is related to Extraversion and associated variance in depression, rather than to the depression variance associated with Neuroticism. Findings support the value of research on psychopathology that is theoretically driven and focuses on empirically validated dimensional constructs that bridge the gap between psychopathology and personality (DeYoung & Krueger, 2018).

In addition to corroborating research on the role of reward sensitivity in Extraversion, findings support research suggesting depression is related to both negative affectivity and lack of reward responsiveness, with these symptoms related to Neuroticism and Extraversion, respectively (Kotov et al., 2010). This conceptualization is in line with RDoC's distinction between positive and negative valence systems and with many traditional models of depressive symptomatology (Barch et al., 2016; Clark & Watson, 1991). Findings are also interesting when interpreted in conjunction with recent evidence that performance on the PRT predicts positive response of depression patients to bupropion but not sertraline (Ang et al., 2020); this is relevant to possible distinct mechanisms of depression related to Neuroticism vs. Extraversion, as bupropion is thought to act on dopamine (though also norepinephrine), a neurotransmitter theoretically and empirically linked to Extraversion (Ascher et al., 1995; Depue & Collins, 1999), while sertraline acts primarily on serotonin, which is related to Neuroticism (De Vane et al., 2002; Wright et al., 2019). Future studies investigating cognitive and affective mechanisms of depression—and their potential amelioration through psychopharmacological and behavioral interventions—might benefit from incorporating measures of Extraversion and Detachment, in addition to measures more closely related to Neuroticism and Negative Affect.

Reliability and Validity Considerations

When using behavioral tasks to assess affective and cognitive mechanisms, such as reward sensitivity, it is important to ensure one's tasks are both reliable in their measurement of individual differences and valid in measuring the constructs they seek to represent. Merely having face validity does not make a measure reliable or valid. Moreover, even tasks able to detect robust effects at the group level can fail to produce reliable measurement of individual differences (Dang et al., 2020; Hedge et al., 2018; Enkavi et al., 2019a; 2019b; Schnabel et al., 2008). Many of the most frequently used behavioral tasks emerge from the cognitive and social psychology literatures, where there is a focus on reducing individual differences in task performance in efforts to reduce measurement error; hence, interindividual variability on task performance is often seen as an obstacle to be overcome, rather than a substantive variable to be tested (Cronbach, 1957). This leads to problems when these tasks are then adopted for individual differences research because low between-subject variability inherently reduces the reliability and, in turn, the validity of these measures (Hedge et al., 2018). Compared to tasks from social and cognitive psychology, measures emerging from clinical psychology, where the PRT originates, often fare better in evaluations of their psychometric properties because individual differences are of primary interest in that field (Barch et al., 2016; Pinkham et al., 2018). Regarding Pizzagalli's PRT in particular, there is some work—albeit limited in scope and hindered by low statistical power—establishing the task's reliability and validity.

The two main ways to quantify the reliability of a task are internal consistency and test-retest reliability. The fact that all three response bias variables (for Blocks 1 through 3) are correlated with one another and load significantly onto a single latent variable in the current sample provides evidence of internal consistency. As mentioned in my methods section, there is also evidence that the PRT has acceptable split-half reliability, providing further evidence for the task's internal consistency (Lawlor et al., 2019). In comparison, the evidence for adequate test-retest reliability is limited to two studies with very small samples. Notably, however, the PRT is currently under evaluation as part of two major research efforts using large clinical samples: the Cognitive Neuroscience Test Reliability and Clinical Applications for Schizophrenia (CNTRACs) Consortium (Barch et al., 2017; Gold et al., 2012) and the Establishing Moderators and Biosignatures of Antidepressant Response for Clinical Care for Depression (EMBARC) Study (Trivedi et al., 2016; Webb et al., 2016; Webb et al., 2020). These efforts should help to further establish (or refute) the PRT's reliability as a useful measure for personality and psychopathology research.

It is also important to consider criterion validity of the PRT in relation to other tasks and variables it would be expected to predict. Many researchers using the PRT—as well as workgroups evaluating this and similar measures (e.g., the National Advisory Mental Health Council Workgroup on Tasks and Measures for RDoC)—have framed its published associations with clinical constructs of interest such as depression and anhedonia, supplemented with neuroimaging and candidate gene research relating performance to brain regions and neurochemicals related to reward processing, as evidence of convergent validity (Delgado et al., 2016). Nonetheless, a majority of these studies (particularly those with neuroimaging or genetic components) are lacking in statistical power (Bogdan et al., 2006; Bogdan et al., 2011; Bress & Hajcak, 2013; Liu et al., 2015; Pizzagalli et al., 2007; Pizzagalli et al., 2008a; Pizzagalli et al., 2008b; Santesso et al., 2008; Webb et al., 2016; Whitton et al., 2016). Thus, further research establishing the reliability and validity of the PRT is essential.

Limitations

In addition to these broader issues of reliability and validity, there are a few other limitations worth discussing. Although the current findings suggest that performance on the PRT is not correlated with depression in a community sample, I did not investigate participants with severe levels of anhedonia or depression, as was done in studies using the task in clinical populations (e.g., Pizzagalli et al., 2008a; Pizzagalli et al., 2008b; Vrieze et al., 2013). Response bias on the PRT may, in fact, be reduced among depressed individuals with more extreme levels of anhedonia, even if the association is not strong enough to be detected in the general population, though other recent studies in clinical samples also call this association into question (Lawlor et al., 2019). This possibility could not explain, however, the failure to replicate the findings of Pizzagalli et al. (2005), given that BDI scores in the current sample were comparable to or higher than those in the original undergraduate sample. Additionally, the lack of association between PRT performance and questionnaire measures of anhedonia and depression does not speak to the question of whether these characteristics might be related to reward sensitivity assessed using different behavioral tasks. Future research in this area could address these limitations by incorporating additional measures of reward sensitivity and recruiting additional participants in the clinical range of depressive symptomatology. In particular, using latent variable frameworks to assess the relation between joint personality-psychopathology dimensions and reward sensitivity modeled as shared variance of performance on multiple tasks could be particularly useful for addressing several of the limitations noted here (Blain et al., 2020a; Blain et al., 2020b; Campbell & Fiske, 1959; Nosek & Smythe, 2007).

Chapter 3 Conclusion

In summary, results of the current study failed to replicate previous findings and suggest that reduced reward response observed in previous studies may have been driven by low-levels of Extraversion or by the presence of Detachment-related psychopathology, rather than by depressive symptoms more generally. Thus, these findings emphasize the importance of transdiagnostic research and the conceptualization of depression as related to both high Neuroticism and low Extraversion. They also provide support for the theory that reward sensitivity is a core mechanism of Extraversion. Finally, they underscore the importance of replication with adequate sample sizes in moving toward reproducibility in psychological research.

CHAPTER 4:

Affiliation: A Consequential, Interstitial Trait

Interpersonal behaviors and relationships have long been a topic of interest for psychologists, philosophers, and artists alike. The ability to develop and maintain healthy relationships is an essential part of well-being, and better understanding these topics is essential to psychological science. Situational factors may influence one's decision to engage in affiliative behavior, but there also appears to be an underlying trait representing the tendency toward affiliation, which can be usefully examined through personality psychology. Affiliation seems to require two distinct psychological processes: finding interactions with other people rewarding, but also empathizing with and caring about them. These tendencies toward reward sensitivity and empathizing can be understood through the framework of personality psychology, and more specifically, in relation to the Big Five traits of Extraversion and Agreeableness.

The Big Five and Social Behavior

The core traits in modern personality psychology are the Big Five (Costa & McCrae, 1992; John et al., 2008), of which two traits are known to be particularly related to social behavior and interpersonal functioning: Extraversion and Agreeableness. Recent research has further explored the component parts of these traits, revealing that each of the Big Five can be reliably decomposed into two distinct aspects (DeYoung et al., 2007). The two aspects within Extraversion are labeled Assertiveness and Enthusiasm, and the two aspects within Agreeableness are labeled Compassion and Politeness. Each of these aspects can be further decomposed into additional finer-grained traits, typically referred to as facets. For example, Compassion can be broken into facets such as Empathy and Altruism. Together, the traits of Extraversion and Agreeableness, along with their component aspects, can capture a broad array of individual differences when it comes to social behavior and interpersonal functioning.

There are also promising mechanistic frameworks that can help us further understand the underlying processes and functions associated with these traits. Though a majority of personality research has focused on the description and measurement of traits in terms of taxonomies such as the Big Five, more recently, comprehensive explanatory frameworks have begun to emerge. For instance, Cybernetic Big Five Theory (CB5T) was developed as the first biologically grounded theory of personality explicitly designed to explain the Big Five (DeYoung, 2015; Allen & DeYoung, 2017). CB5T identifies the major psychological functions that underlie each of the Big Five and begins to identify the complex neurobiological systems that instantiate those functions. Cybernetics is the study of principles governing goal-directed, self-regulating systems, and CB5T is based on the premise that the Big Five represent variation in universal human mechanisms that evolved to enable people to pursue their goals. Thus, traits are conceived as dimensions that can be used to describe psychological variation in any human population across human history, reflecting variation in parameters of biological mechanisms that all people share.

CB5T makes use of the empirical demonstration that each of the Big Five has two major subfactors (i.e., aspects) that appear to reflect the most important distinctions for discriminant validity within each of the five domains, in terms of genetic variance, relations to psychopathology, and predictive ability (DeYoung et al., 2007; DeYoung et al., 2016; Jang et al., 2002). When describing interpersonal functioning in relation to the Big Five, it is sometimes useful to identify relations that are specific to one of the ten lower-level aspects, rather than being characteristic of the broader Big Five dimensions. CB5T has implications for understanding social behavior (DeYoung & Weisberg, 2019), particularly as it can help us understand the functions of traits such as Extraversion and Agreeableness, as well as their component aspects and their intersection.

Extraversion describes one's tendency to be approach-oriented, sociable, and expressive of positive emotions, all of which seem to reflect an underlying sensitivity to reward (Chapter 3 of this dissertation; Corr, 2008; DeYoung, 2015; DeYoung & Weisberg, 2018; Lucas et al., 2000; Smillie, 2013; Smillie et al., 2007; 2011a; 2011b; 2012; 2019). Rewards can be understood through two main subtypes: 1) incentive rewards, which are cues that one is getting closer to achieving a goal and 2) hedonic rewards, which correspond to the enjoyment experienced once a goal is achieved. These two forms of reward can be described as "wanting" and "liking", respectively, and CB5T proposes that the aspects of Assertiveness and Enthusiasm differentiate between sensitivity to these two forms of reward (DeYoung et al., 2015; DeYoung & Weisberg, 2018). Though both aspects of Extraversion are related to social functioning, Enthusiasm and associated experiences of interpersonal pleasure may be particularly important to the formation and maintenance of close relationships.

Agreeableness describes one's tendency to be altruistic and prosocial, rather than callous and exploitative. As humans are social mammals, we must be able to coordinate our goals, interpretations, and strategies with our conspecifics; individual differences in Agreeableness are associated with our ability to do so successfully (DeYoung et al., 2015; DeYoung & Weisberg, 2018; Graziano & Tobin, 2013; Van Egeren, 2009). Mechanisms underlying cooperation, altruism, and associated variation in Agreeableness appear to include the ability to perceive the emotions and interpret the mental states of others (Allen et al., 2017; Chapters 1 and 2 of this dissertation; Graziano et al., 2007; Mayer et al., 2008; Nettle & Liddle, 2008; Wilkowski et al., 2006), as well as the ability to suppress aggressive impulsive and destructive emotions (Meier et al., 2006). Though both aspects of Agreeableness are important for social functioning, Compassion may be particularly important when it comes to close relationships.

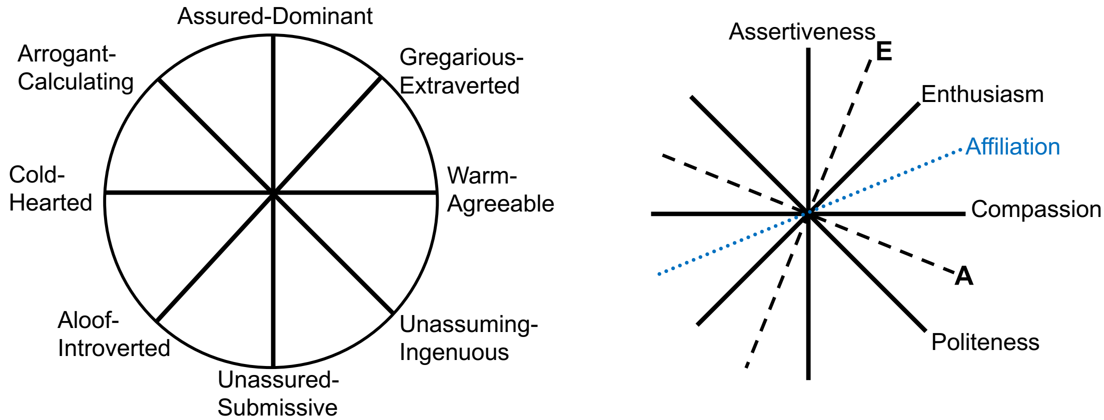
To understand individuals' interest in and desire for close relationships as a personality trait, a combination of these previously identified traits may be helpful. In particular, the Enthusiasm aspect of Extraversion and the Compassion aspect of Agreeableness correlate strongly with one another and are both associated with tendencies to form and maintain relationships. This specific trait blend, called Trait Affiliation, is situated halfway between the two aspects (as measured in vector space, implying equal relations to the two traits) and is easily integrated into other psychological models of interpersonal behavior, such as the Interpersonal Circumplex or IPC (Barford et al., 2015; DeYoung et al., 2013; DeYoung & Weisberg, 2018).

The Interpersonal Circumplex

As discussed in the general introduction to this dissertation, the IPC is a widely used structural model of interpersonal traits and behaviors. Although early interpersonal theory stemmed from the work of Harry Stack Sullivan (1953, 1964), the IPC of today emerged from contributions of Timothy Leary (1957). As a circumplex (a model in which variables are arranged in a circle and appear at regular intervals around the circumference), the IPC represents interpersonal variables using two orthogonal dimensions or axes: 1) Status and 2) Love (Gurtman, 2009; Wiggins, 1979). Locations of variables in the IPC provide a quick way to visualize how the variables relate to each axis and to one another. Variables close together on the IPC are highly positively associated, those on opposite sides of the circumference are negatively associated, and those at right angles to one another are uncorrelated.

In addition to the two orthogonal dimensions of the IPC, additional locations on the circumplex can be specified (Figure 4.1). These are typically specified using degrees, with 0° representing the Love axis and other values proceeding counterclockwise. The IPC is frequently discussed in terms of eight subdivisions, known as octants, which can be defined by the high and low poles of the Status and Love axes, as well as their 45° rotations or diagonals. Various questionnaires have been created to assess how well individuals can be described by each of the eight octants, in terms of trait tendencies or behaviors in a specific situational context (e.g., Markey & Markey, 2009).

Figure 4.1



Integration of the Big Five, Interpersonal Circumplex, and Trait Affiliation.

Since the IPC and Big Five are two of the most used and influential models for understanding individual differences in personality and social behavior, substantial efforts have been taken to unify these systems (Barford et al., 2015; DeYoung et al., 2013; McCrae & Costa, 1989; Pincus, 2002; Wiggins & Pincus, 1994). Interestingly, relative to most other Big Five domain pairs, Extraversion and Agreeableness show a high degree of circumplexity. This means the various aspects and facets of these two traits show a consistent density of variables around the circle that represents their variance in two orthogonal dimensions, rather than clustering predominantly on the major axes (Gurtman, 2009; Saucier, 1992). Consequently, facets of Extraversion and Agreeableness will sometimes group together in factor analyses, particularly when there is a disproportionately high number of markers of one domain or the other (e.g., Church, 1994; Church & Burke, 1994).

Several studies show that Extraversion and Agreeableness can be represented as rotational variants of the IPC axes, falling near 60° and 330º, respectively (DeYoung et al., 2013; McCrae & Costa, 1989; Pincus, 2002; Wiggins & Pincus, 1994). A full integration of the IPC and relevant Big Five traits was pictured in Figure 4.1. Lower-order factors of Agreeableness and Extraversion, including both facets and aspects, have also been mapped onto the IPC. For instance, previous research suggests the facets of Extraversion tend to cluster into groups around the 45° and 90° positions on the circumplex, whereas facets of Agreeableness are spread out from roughly 0º to 300º, moving clockwise (Barford et al., 2015; DeYoung et al., 2013; McCrae & Costa, 1989; Pincus, 2002). Aspects of Extraversion and Agreeableness also map onto the IPC, with Compassion falling at 0º, Enthusiasm at 45º, Assertiveness at 90º, and Politeness at 315º (Barford et al., 2015; DeYoung et al., 2013). As mentioned previously, though, another trait particularly relevant to individual differences in social behavior and the tendency to form close relationships may be represented by the blend of Compassion and Enthusiasm, which would be predicted to fall at 22.5º on the IPC. Already, a few personality questionnaires have been shown to approximate this position in the IPC between Compassion and Enthusiasm—namely, scales assessing warmth and social closeness (DeYoung et al., 2013).

The "Warmth" Facet, Social Closeness, and Trait Affiliation

One motive for more closely examining the interstitial space between Agreeableness and Extraversion (i.e., Trait Affiliation) is to help better define the already fuzzy line between some facets of these domains. As an illustration of this ambiguity, one of the most widely used measures of the Big Five, the NEO PI-R (Costa & McCrae, 1992), includes a "Warmth" facet scale in Extraversion, but another popular Big Five measure, the Abridged Big Five Circumplex scales for the IPIP (AB5C-IPIP; Goldberg, 1999), includes "Warmth" under Agreeableness. Although this facet assignment scheme may at first seem contradictory, it is consistent with how the scales empirically function. The AB5C-IPIP Warmth scale shows a primary loading on Agreeableness with a secondary loading on Extraversion, whereas the NEO PI-R Warmth scale shows the opposite pattern (Goldberg, 1999; Johnson, 1994). In terms of content, both scales appear to reflect affiliative tendencies, but AB5C-IPIP Warmth focuses on altruism and prosociality (i.e., Compassion-related tendencies) whereas NEO PI-R Warmth focuses on gregariousness (i.e., Enthusiasm-related tendencies). This is consistent with results showing both scales fall between the 0º and 45º angles when examined in the IPC factor space, with NEO PI-R Warmth falling closer to 45º and AB5C-IPIP closer to 0º (DeYoung et al., 2013). Thus, Warmth as a concept seems to already characterize the interstitial space between Agreeableness and Extraversion, or more specifically, the space between Compassion and Enthusiasm (DeYoung et al., 2013; Saucier et al., 1992). The exact wording of Warmth items on any given measure will determine whether they fall closer to Agreeableness, closer to Extraversion, or equidistant between Compassion and Enthusiasm.

Another personality construct relevant to the space between Compassion and Enthusiasm is Social Closeness, as measured by the Multidimensional Personality Questionnaire (MPQ). Those high in Social Closeness frequently experience warmth and affection, like to spend time with friends and family, turn to others for comfort, and value close personal relationships (Tellegen & Waller, 2008). The MPQ Social Closeness scale has already been used specifically in research on Trait Affiliation (Depue & Morrone-Strupinsky, 2005; Morrone-Strupinsky & Depue, 2004; Morrone et al., 2000; Moore et al., 2014). Scores on MPQ Social Closeness have been shown to predict film-induced increases in warm and affectionate states (Morrone-Strupinsky & Depue, 2004), statistically moderate attentional response to affiliative cues (Moore et al., 2014), and statistically moderate response to the opiate antagonist naltrexone (Depue & Morrone-Strupinsky, 2005). Finally, when examined in relation to the broader factor space of the IPC, MPQ Social Closeness tends to fall between the 0º and 45º angles, where Trait Affiliation would be expected to fall if it indeed represents a blend of Compassion and Enthusiasm; nonetheless, MPQ Social Closeness is slightly closer to Enthusiasm than Compassion—both theoretically and empirically—falling at 31.5º on the IPC in a recent study (DeYoung et al., 2013). Thus, an even more precise targeting of Trait Affiliation at the 22.5º angle might be possible.

Despite considerable research on the topic, the precise location of Trait Affiliation within the broader personality factor space has remained a topic of debate. Some research groups have used MPQ social closeness as a marker of Affiliation, whereas many researchers using the IPC model refer to the Love axis as “Affiliation.” Neither of these operationalizations represents a precise blend of Compassion and Enthusiasm falling at the 22.5º angle, so each of these versions of Affiliation may fail to directly capture important variance central to the ability to form and maintain close relationships.

According to the work of Depue & Morrone-Strupinsky (2005), warmth and affection are the states most associated with Affiliation as a construct (Morrone-Strupinsky & Depue, 2004). Thus, my collaborators' previous work has attempted to locate where individual differences in participants' self-reported tendency to feel affectionate falls along the IPC. This work showed that, as predicted, a single item measure, the adjective "affectionate," fell very close to the 22.5° angle, or more precisely at 19.8º in one sample and at 23.0º in another (DeYoung et al., 2013).

When taken in conjunction with CB5T and neurobiological models of Affiliation, work modeling Trait Affiliation in relation to Extraversion, Agreeableness, and the IPC suggests that individual differences in Trait Affiliation may stem from variation in both Agreeableness-related processes of empathy, cooperation, and social cognition, as well as Extraversion-related processes involving sensitivity to reward (Depue & Collins, 1999; Depue & Morrone-Strupinsky, 2005; DeYoung, 2015; DeYoung et al., 2013; DeYoung & Weisberg, 2018). Because this specific interstitial personality dimension—Trait Affiliation—may be particularly important in influencing and predicting life outcomes related to interpersonal functioning and relationship success, additional research to better assess, characterize, and explain the trait is imperative.

The Current Research

Previous psychometric work suggests that Trait Affiliation represents an interstitial personality facet, blending the Compassion aspect from Agreeableness and the Enthusiasm aspect from Extraversion (DeYoung et al., 2013). Trait Affiliation can be incorporated into other frameworks such as the IPC, and relevant neurobiological models have emphasized the importance of investigating specific mechanisms and consequences of this trait (Depue & Morrone-Strupinsky, 2005; DeYoung et al., 2013; DeYoung & Weisberg, 2018). Despite the obvious importance of Affiliation, there is currently a lack of psychometrically validated questionnaires that specifically measure this trait.

Although certain existing measures—such as the MPQ's Social Closeness scale and Warmth facet scales from certain Big Five questionnaires—approximate a blend of Compassion and Enthusiasm, these scales tend to lean more heavily toward one aspect or the other and they were not empirically derived to reliably measure the intersection of these Big Five two aspects. I believe a scale that precisely captures variance in affiliative tendencies as an equally balanced blend of Compassion and Enthusiasm would be a worthwhile contribution. Thus, my research group has developed a new self-report measure of Trait Affiliation. This questionnaire assesses individuals' tendencies to seek out, develop, and maintain close relationships.

Developed from an initial set of 24 candidate items taken from the International Personality Item Pool, the Trait Affiliation Scale in its current form consists of ten items. In this chapter, I document the construction and validation of this scale. Data for this validation effort are compiled from a collection of six samples (total $N = 27{,}047$) and organized into six studies. A summary of the datasets and studies is shown in Table 4.1.

Table 4.1. *Summary of datasets used in Chapter 4*

Label	***N***	**Studies Including this Dataset**	**Purpose/ Scale Properties Tested**
ESCS	409	4a, 4c	Item Selection Convergent Validity
SAPA	25732	4b, 4c	Scale Refinement Convergent, Discriminant, and Structural Validity
PPA	259	4c, 4d	Convergent, Discriminant, and Structural Validity Test-retest Reliability
YW	280	4c, 4e	Convergent, Discriminant, and Structural Validity Criterion/Incremental Validity
SBSC	335	4c, 4e	Convergent, Discriminant, and Structural Validity Criterion/Incremental Validity
SBAV	195	4c, 4f	Convergent, Discriminant, and Structural Validity Criterion/Incremental Validity

Study 4a focuses on the item selection and scale construction process, as well as examining initial construct validity. Study 4b focuses on scale refinement, including the application of item response theory to evaluate item information and create a ten-item scale from the initial set of 24 candidate items; Study 4b also provides further evidence of convergent and discriminant validity. Study 4c provides evidence of internal consistency and structural, convergent, and discriminant validity in terms of the scale's final ten-item version, using a set of six samples. Study 4d provides evidence of test-retest reliability in a four-wave longitudinal dataset. Study 4e examines evidence of incremental validity for self-reported, real-world social outcomes, testing associations of Trait Affiliation with relevant interpersonal variables (e.g., social goals, social behaviors, social network size, and social cognitive ability) above and beyond Compassion and Enthusiasm alone. Finally, Study 4f assesses the incremental validity of the Trait Affiliation Scale in predicting response to laboratory paradigms designed to induce affiliative states. After presenting my results, I discuss the theoretical and practical importance of Affiliation as a trait, while providing recommendations for use of the scale in future research.

Study 4a: Item Selection and Scale Creation

To create the Trait Affiliation Scale, I utilized the International Personality Item Pool (IPIP), which contains more than 2,500 public-domain personality items, each of which was completed by participants in the Eugene Springfield Community Sample (ESCS). Many previous research efforts have used the IPIP to create new scales, based on the associations of IPIP items with other validated personality measures or specified criteria of interest (e.g., DeYoung et al., 2007; Markey & Markey, 2009; Goldberg, 1999).

Deriving test items empirically rather than theoretically has several advantages. For instance, using the IPIP allows us to select items for the new Trait Affiliation Scale that are specifically associated with the intersection of Compassion and Enthusiasm, rather than relying on existing measures that approximate the content of Affiliation and its theorized location in the IPC and Big Five factor space (e.g., the MPQ or various Warmth facet scales). Using the IPIP to create new measures has numerous advantages, not the least of which is the fact that the items are in the public domain so that measures developed using the IPIP can help facilitate reproducibility and open science.

In Study 4a, I examined associations of IPIP items with the mean of BFAS Enthusiasm and Compassion, to select items specifically tapping into Trait Affiliation. (I planned to subsequently use the selected IPIP variables as candidate items in follow-up scale validation and refinement analyses.) After I selected a pool of items specifically correlated with the average of Enthusiasm and Compassion, I computed participants' average scores on these items and used principal axis factoring with a Procrustes rotation to examine where this variable falls in the factor space of the IPC and Big Five. Based on preliminary work using a single-item Trait Affiliation variable, I hypothesized the new Trait Affiliation Scale variable would fall near the 22.5º angle of the interpersonal circumplex, when modeled alongside a variety of other relevant variables from the Big Five and IPC (DeYoung et al., 2013).

Method

Participants

Sample 1 included 409 individuals (166 males and 243 females) from the Eugene-Springfield Community Sample (ESCS). Participants' ages ranged from 22 to 85 years (M = 52.8, SD = 12.5). The sample spanned all levels of educational attainment, with an average of approximately 2 years of postsecondary schooling. Most participants identified as White (97%), and 1% or less (for each category) identified as Hispanic, Asian American, Native American, or did not report their ethnicity. The sample used in the current study was a subsample of the ESCS dataset, with participants selected who had completed Affiliation-related measures of interest.

Procedure and Materials

Participants in the ESCS were recruited by mail from lists of homeowners and agreed to complete questionnaires, delivered by mail, for pay, over a period of many years, beginning in 1994. Participants completed a variety of measures related to personality and individual differences. Specific measures used in the current study are described below (see previous research for a discussion of the broader IPIP; Goldberg et al., 2006):

Big Five Aspect Scales (BFAS; DeYoung et al., 2007). The BFAS consists of 100 items that use a five-point Likert scale. The questionnaire is based on the five-factor model and breaks down each of the factors into two aspects. In the current study, the Agreeableness and Extraversion aspect scales (i.e., Compassion, Enthusiasm, Politeness, and Assertiveness) were used.

Big Five Inventory (BFI; John et al., 2008). The BFI is a well validated and widely utilized measure of the Big Five, consisting of 44 items using a 5-point Likert scale. In the current study, the Agreeableness and Extraversion scales were used.

IPIP-IPC Scales (Markey & Markey, 2009). This measure uses 32 IPIP items describing interpersonal behaviors and has been validated using observations of social behavior in addition to correlations with other established IPC measures (Markey, Anderson, & Markey, 2012). Participants rate how accurately each phrase describes themselves on a scale of 1 (very inaccurate) to 5 (very accurate). Four items constitute the scale for each octant of the IPC. Scores were ipsatized because all items were positively keyed. Both the BFAS and the IPIP-IPC scales utilize IPIP items, and I identified two items that were included in both instruments. To avoid inflating correlations by including redundant data in two of the measures, in analyses using both the IPIP-IPC and BFAS, I removed two items from the BFAS Compassion scale that were also included in the IPIP-IPC. (I chose to shorten the BFAS because its scales are longer than those for the IPIP-IPC.)

NEO Personality Inventory-Revised (NEO PI-R; Costa & McCrae, 1992). The NEO PI-R is a well-validated, proprietary measure of the Big Five domains and their facets. For the current study, I included scores on the Warmth facet scale from the NEO PI-R. Warmth is considered a part of the Extraversion domain in the NEO PI-R.

Abridged Big 5 Circumplex IPIP scales (AB5C-IPIP; Goldberg, 1999). Scores on the Warmth facet scale were included, which is considered a part of the Agreeableness domain in the AB5C-IPIP. Two items were excluded because they were identical to BFAS items, leaving 9 items.

Multidimensional Personality Questionnaire (MPQ; Tellegen & Waller, 2008). Three MPQ scales related to interpersonal functioning and social behavior were included. First, I used the 21-item Social Closeness scale which has been used as a measure of Trait Affiliation in previous research (Depue & Morrone-Strupinsky, 2005; DeYoung et al., 2013; Morrone-Strupinsky & Depue, 2004). Other scales included were Social Potency (25 items) and Aggression (19 items), which describe traits similar to Assertiveness and low Politeness, respectively. These scales are included in the current analyses because a large amount of previous work on interpersonal behavior has administered the MPQ.

Statistical Approach

Item Selection. To select items for the Trait Affiliation Scale, I computed Pearson correlations of 2552 items from the IPIP with a BFAS Affiliation proxy variable, as well as examining correlations with all the individual BFAS subscales. The Affiliation proxy variable was computed for each participant as the average of their respective Compassion and Enthusiasm scores. Candidate items for the scale were selected based on the criteria of 1) having a correlation with the BFAS Affiliation proxy variable greater than or equal to r = .30 in absolute value, 2) having a correlation with the BFAS Affiliation proxy variable greater than either of the individual correlations with Compassion and Enthusiasm, and 3) having correlations with Compassion and Enthusiasm that were at least .10 larger than correlations with any the other eight BFAS aspects.

Initial Test of Construct Validity. In order to assess the structural, convergent, and discriminant validity of the identified candidate items, I computed a Trait Affiliation score for each participant based on the average of all candidate items and then examined how these scores were associated with other variables from the Big Five and IPC; specific variables from the identified candidate items that overlapped with IPC or Big Five measures also included in these analyses were removed before computing the Trait Affiliation scores. To examine associations with these other constructs, I followed a method virtually identical to that reported in previous work (DeYoung et al., 2013), designed to capture the factor space described by interpersonal theory.

To examine how the Trait Affiliation variable related to the established IPC factor space, first, I generated a target matrix with two factors, based on the IPC. This represented a circular structure where the positive pole of Status loaded 0 on the first factor and .8 on the second. (.8 was chosen instead of 1.0 to account for measurement error; no variable is likely to have a perfect loading on either factor.) Other variables were also assigned target loadings based on their hypothesized IPC locations; for example, Trait Affiliation was predicted to fall at the 22.5 angle. I then extracted two factors from the interpersonal variables using principal axis factoring, applying a Procrustes rotation to align the solution to the target matrix (Schönemann, 1966). The target loading matrix is shown beside the rotated observed loadings in Table 4.3. Angular projections for each variable were computed as the arctangent of the quotient of each variable's pair of factor loadings.

To assess the circumplex structure of the interpersonal variables, I utilized Tucker's congruence coefficients as described in previous work (DeYoung et al., 2013; Terracciano et al., 2003). This method computes congruence coefficients for each variable based on the correspondence of target and observed factor loadings. These coefficients are computed as the cosine of the angle between target and observed loadings (represented as vectors) and are analogous to correlations—ranging from -1 to 1, with higher magnitude values indicating greater similarity. Coefficients greater than .95 are considered evidence of replication, whereas those greater than .85 are evidence of similarity (Lorenzo-Seva & Ten Berge, 2006).

Results and Discussion

Item Selection

The initial selection of Trait Affiliation Scale items is presented in Table 4.2, along with the correlations of each item with the BFAS Affiliation proxy variable, Compassion, and Enthusiasm. This selection process yielded a total of 23 items. Correlations of the selected Trait Affiliation candidate items with the BFAS Affiliation proxy variable ranged in absolute value from .31 to .66. Complete documentation for all 2552 tested IPIP items (including correlations of each item with the 10 BFAS aspects and BFAS Affiliation proxy variable, as well as an item-level explanation of selection and exclusion criteria) is available in supplementary materials.

Table 4.2. *Trait Affiliation Scale candidate items with item parameters and IPIP/SAPA codes*

IPIP Code	SAPA Code	Item Text	*r Ac-Ee ESCS*	*r Ee ESCS*	*r Ac ESCS*	Final Scale ?	a_i SAPA	AUC SAPA
H1151	q_505	Cheer people up.	.62	.52	.47	Y	1.31	.77
X161	q_748	Enjoy bringing people together.	.45	.37	.37	Y	1.02	.67
A4	q_899	Find it difficult showing people that I care about them. (R)	-.47	-.41	-.40	Y	-0.89	.61
H100	q_1000	Give compliments.	.48	.37	.43	Y	0.97	.60
H26	q_1418	Make others feel good.	.60	.41	.42	Y	1.27	.70
V181	q_2881	Don't feel the need to be close to others. (R)	-.47	-.42	-.42	Y	-0.94	.62
V230	q_2970	Have no need for close friendships. (R)	-.48	-.41	-.36	Y	-0.81	.50
V129	q_3030	Love to make other people happy.	.52	.39	.45	Y	1.19	.68
(NA)	q_3893	Feel affectionate towards people.	—	—	—	Y	1.20	.77
V299	q_2900	Don't think it's important to socialize with others. (R)	-.46	-.39	-.30	Y	-1.09	.73

D42	q_854	Feel that having close friends is not especially important to me. (R)	-.40	-.34	-.34	N	-0.78	.48
A132	q_1086	Have difficulty showing affection. (R)	-.55	-.50	-.49	N	-0.76	.49
A79	q_1154	Hug my close friends.	.49	.37	.43	N	0.81	.54
A84	q_1189	Keep my happy feelings to myself. (R)	-.47	-.47	-.45	N	-0.62	.35
D6	q_1471	Often do nice things for people.	.44	.30	.35	N	0.90	.53
H682	q_1535	Prefer to do things by myself. (R)	-.31	-.29	-.28	N	-0.35	.13
H660	q_1923	Want to be left alone. (R)	-.42	-.46	-.37	N	-0.70	.42
V218	q_2764	Am good at understanding others' feelings.	.49	.41	.42	N	0.89	.54
V285	q_2869	Do not go out of my way to make others smile or laugh. (R)	-.34	-.30	-.29	N	-0.85	.54
V295	q_3006	Know what to say to make people feel good.	.47	.35	.33	N	0.75	.45
M44	q_3697	Help my friends.	.41	.33	.38	N	0.88	.47
H21	q_150	Am interested in people.	.59	.45	.45	N	0.90	.55
H107	q_1419	Make people feel at ease.	.66	.46	.47	N	0.87	.53
D82	q_1575	Rarely enjoy being with people.	-.43	-.40	-.34	N	-0.69	.41

In addition to the 23 items taken from the IPIP, I added one additional item to the candidate item pool based on previous research on Affiliation and its relation to other traits. This 24th item asks whether participants “Feel affectionate towards people” and is based on research indicating that warmth and affection are the emotional states most associated with Trait Affiliation (Depue & Morrone-Strupinsky, 2005) and work suggesting scores on this item closely match the predicted 22.5^{o} angular position of Trait Affiliation in relation to variables from the Big Five and IPC (DeYoung et al., 2013).

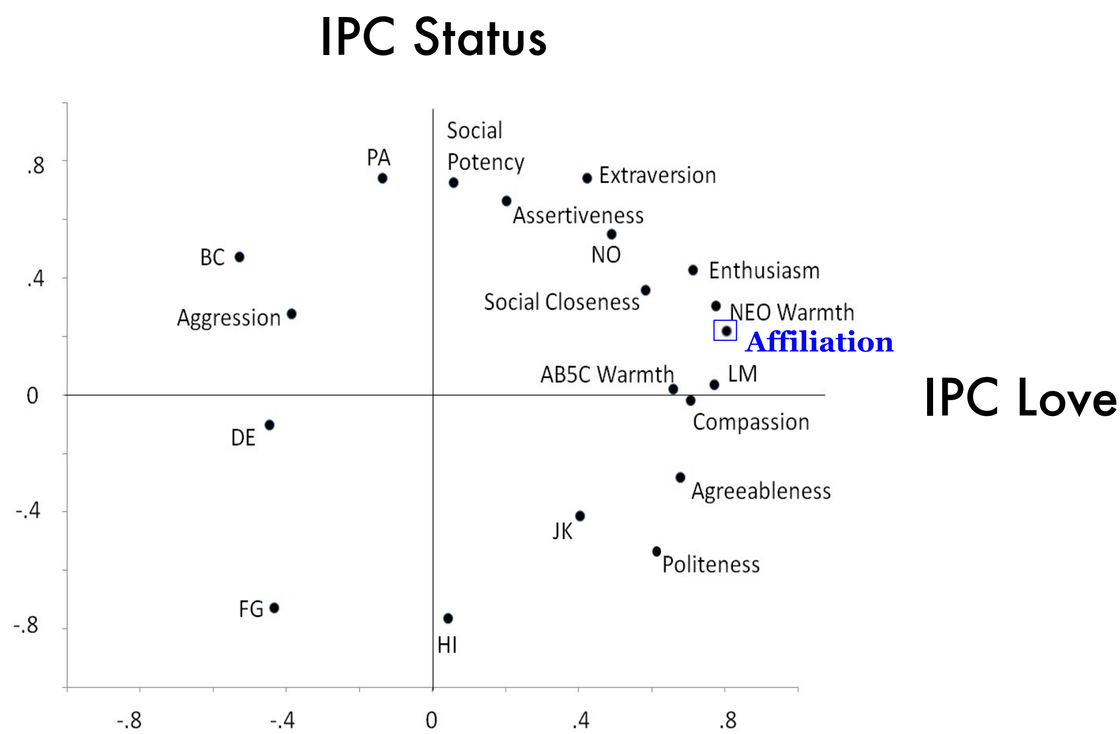
Initial Test of Construct Validity

Target and observed (rotated) factor loadings for Trait Affiliation (as measured by the average of candidate items) and other interpersonal variables are shown in Table 4.3, along with predicted and observed angular projections for each variable and corresponding Tucker’s congruence coefficients. Factor loading plots are visualized in Figure 4.2. Results in this sample for variables other than the new Trait Affiliation items are discussed in greater detail in previous work (DeYoung et al., 2013).

Table 4.3. *Target and rotated factor matrices with corresponding IPC angles and Tucker's congruence coefficients*

Variable	*Target Matrix*			*Rotated Matrix*			Congruence Coefficient
	F1	F2	θ	F1	F2	θ	
Trait Affiliation	.74	.31	22.5	.80	.35	23.26	1.00
BFAS Assertiveness	.00	.80	90	.20	.66	73.09	.96
BFAS Enthusiasm	.57	.57	45	.71	.43	31.16	.97
BFAS Compassion	.80	.00	0	.73	-.02	-1.32	1.00
BFAS Politeness	.57	-.57	315	.61	-.54	-41.42	1.00
BFI Extraversion	.31	.74	67.5	.41	.74	61.00	.99
BFI Agreeableness	.74	-.31	337.5	.67	-.28	337.33	1.00
IPIP-IPC PA (Assured-Dominant)	.00	.80	90	-.14	.74	100.43	.98
IPIP-IPC NO (Gregarious-Extraverted)	.57	.57	45	.48	.55	48.73	1.00
IPIP-IPC LM (Warm-Agreeable)	.80	.00	0	.77	.04	2.79	1.00
IPIP-IPC JK (Unassuming-Ingenuous)	.57	-.57	315	.40	-.41	313.99	1.00
IPIP-IPC HI (Unassured-Submissive)	.00	.80	270	.05	-.76	273.61	1.00
IPIP-IPC FG (Aloof-Introverted)	-.57	-.57	225	-.42	-.72	239.82	.97
IPIP-IPC DE (Cold-Hearted)	.80	.00	180	-.46	-.11	192.91	.97
IPIP-IPC BC (Arrogant-Calculating)	-.57	.57	135	-.51	.47	137.29	1.00
NEO Warmth	.74	.31	22.5	.78	.31	21.49	1.00
AB5C-IPIC Warmth	.74	.31	22.5	.87	.08	5.11	.95
MPQ Social Closeness	.74	.31	22.5	.60	.36	31.39	.99
MPQ Social Potency	.00	.80	90	.05	.73	85.90	1.00
MPQ Aggression	-.57	.57	135	-.38	.28	143.79	.99

Figure 4.2.



Trait Affiliation mapped onto the circumplex model with other interpersonal variables. PA = Assured-Dominant, NO = Gregarious-Extraverted, LM = Warm-Agreeable, JK = Unassuming-Ingenuous, HI = Unassured-Submissive, FG = Aloof-Introverted, DE = Cold-Hearted, BC = Arrogant-Calculatin›g.

Consistent with how the Affiliation items were derived (using correlations with Compassion and Enthusiasm as criteria) and the hypothesized location of Trait Affiliation in the broader IPC factor space, the Trait Affiliation variable fell in the first quadrant of the circumplex between Enthusiasm and Compassion and did not significantly deviate from its hypothesized angular projection of 22.5º. Compared to previous work that found a similar angular position for the single item "affectionate" variables in two other samples (DeYoung et al., 2013), the new Trait Affiliation variable was located further from the origin, suggesting higher factor loadings and correspondingly higher reliability in relation to other IPC variables. Moreover, the pilot Trait Affiliation Scale variable more closely matched the theorized 22.5º angle, compared to scales measuring similar constructs such as MPQ Social Closeness and Warmth facet scales from the NEO PI-R and AB5C-IPIP (though NEO Warmth was nearly as close in its approximation of this angular position).

Study 4b: Scale Refinement

In Study 4a, I selected candidate items for the Trait Affiliation Scale based on their associations with the average of Compassion and Enthusiasm; analyses using a circumplex approach show that the mean of these items fits well into existing models, with the Trait Affiliation Scale falling near the predicted 22.5º angle in the IPC factor space. Nonetheless, further refinement of scales after initial item selection can be useful to ensure all items are adequately contributing to measurement of the underlying latent trait of interest, especially when these analyses are conducted in an independent cross-validation sample. Useful statistical approaches for scale refinement include factor analytic methods applied in the framework of classical test theory, as well as item response theory (IRT) methods that can examine how items differentially function across varying levels of a given underlying trait (Fan, 1998; De Ayala, 2009; Steinberg & Thissen, 1995).

IRT consists of a family of statistical methods that model the probability of given response(s) to an item (e.g., *Strongly Agree*) as a function of various person and item parameters. Persons are placed on latent trait (θ) continua, and their locations are trait-level estimates based on their responses to an entire scale. Unidimensional IRT models place persons on one θ dimension at a time, whereas multidimensional IRT models simultaneously place persons on two or more θ dimensions (e.g., Compassion and Politeness, for an Agreeableness scale). Typically, θ is standardized so the mean is 0 and the standard deviation is 1.

In most IRT models, items can exhibit varying levels of discrimination and difficulty. A highly discriminating artist is sensitive to subtle differences among different colors, however slight; similarly, a highly discriminating item is sensitive to changes in θ, such that increases in θ sharply increase the probability of responding *Agree* or *Strongly Agree* (and vice versa for the probability of *Disagree* or *Strongly Disagree* responses). In contrast, an item with no discriminating ability (a = 0) is akin to a color-blind person; responses to the item reveal nothing about individuals' levels of the underlying trait. "Difficulty", originating from the education literature, describes the θ value at which the probability of a given response to an item is exactly 0.5. For polytomous items like the Trait Affiliation Scale's Likert-scale items, each response category has a corresponding difficulty parameter in the most common IRT models. In the current study, I used IRT to examine the functioning of candidate items and select an optimal group of items for discriminating amongst individuals across a range of underlying latent Trait Affiliation levels.

Though traditional data collection techniques are certainly useful for scale creation and validation, a promising alternative to supplement traditional methods is Synthetic Aperture Personality Assessment (SAPA). SAPA uses planned missing data to administer a selection of personality items from a large pool of potential items to a large sample of participants (Revelle et al., 2021). In SAPA, each participant completes a subset of items from given scales, rather than completing entire scales (as in traditional data collection). Covariance matrices at the inter-scale level can then be recreated using item-level correlations. The SAPA method is well-suited for assessment across multiple domains of personality and for constructing and validating new scales, as SAPA circumvents the need to administer an impractically large quantity of items to any given participant (Condon & Revelle, 2015; Condon et al., 2017; Revelle et al., 2016).

In Study 4b, I used data collected using the SAPA approach and conducted analyses (i.e., factor analysis and IRT) to refine the initial item selection for the Trait Affiliation Scale. I also examined scale-level associations between Trait Affiliation Scale candidate items and related variables of interest (e.g., the Big Five, demographic characteristics, and facets of Agreeableness and Extraversion).

Method

Participants

Sample 2 included 25,732 individuals who provided data through SAPA. Participants included 9924 males (38.6%) and 15,808 females (61.4%), who ranged in age from 14 to 90 years (*M* = 26.0, *SD* = 10.4). Participants included individuals from a variety of different countries, with 18,891 individuals (77.4% of those who listed a nationality) being from the United States. Of the participants who indicated an ethnicity, 11,984 identified as White or Caucasian (67.6%), 1691 as Black or African American

(9.5%), 1685 as Latino or Hispanic (9.5%), 894 as Asian or Pacific Islander (5.0%), 139 as Native American or Alaskan Native (0.8%) and 1344 as Multiracial or Other (7.6%).

Procedure and Materials

Participants provided data through the SAPA website (https://www.sapa-project.org/survey/start.php), where they completed a brief demographic questionnaire and items from a variety of questionnaires of personality and tests of cognitive ability. Data for the current project were collected between September 11, 2013, and October 15, 2013. Measures with items included in the current study included the Trait Affiliation Scale, BFAS, IPIP-HEXACO scales, and Questionnaire Bix Six scales. Participants also completed cognitive ability items from the International Cognitive Ability Resource. Study 4b measures not previously introduced in Study 4a are further described below:

IPIP-HEXACO (Ashton et al., 2007). The IPIP equivalent of the HEXACO PI-R includes a set of 240 items, which measure the six personality domains of Honesty-Humility, Emotionality, Extraversion, Agreeableness, Conscientiousness, and Openness to Experience. Each of these domains has several facets including Expressiveness, Social Boldness, Sociability, and Liveliness for Extraversion, Forgiveness, Gentleness, Flexibility, and Patience for Agreeableness, and Sincerity, Fairness, Greed Avoidance, and Modesty for Honesty-Humility.

Questionnaire Bix Six (QB6; Thalmayer et al., 2011). Items from the 48-item Questionnaire Big Six scales were administered, including variables representing the personality domains of Conscientiousness, Agreeableness, Extraversion, Originality, Resiliency, and Honesty/Propriety.

International Cognitive Ability Resource (ICAR; Condon & Revelle, 2014). The ICAR is a set of publicly available items that measure various domains of cognitive ability. ICAR items included in the current study were sampled from the domains of Letter and Number Series Completion, Matrix Reasoning, Verbal Reasoning, and Three-Dimensional Rotation. Each item uses a multiple-choice format with a single correct answer.

Statistical Approach

Tests of Construct Validity. To assess the convergent and discriminant validity of the candidate Trait Affiliation Scale items, I computed scale-level correlations between Trait Affiliation and a variety of other personality variables. First, the mixedCor function from the *psych* package for R was used to compute item-level correlations; then, the scoreOverlap function was used to compute scale-level correlations based on the item-level correlation matrix, corrected for potential item overlap within any given pair of scales (Revelle, 2021). Finally, heat maps were used to visualize scale-level correlations.

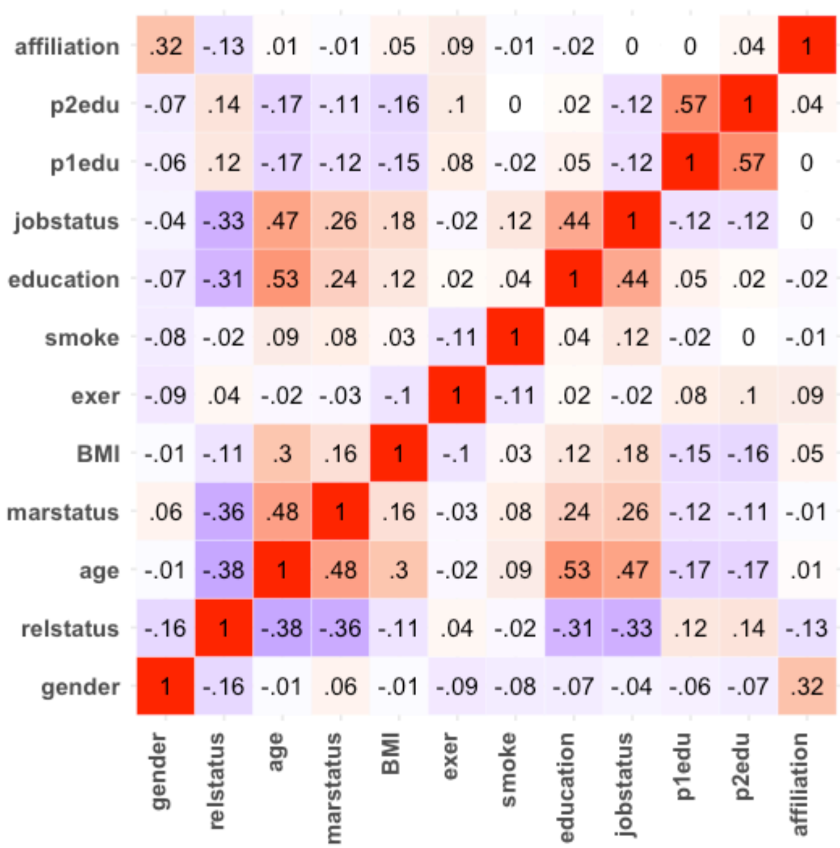
Scale Refinement. First, model-based reliability was tested using Omega (total and hierarchical). Then, IRT was used to examine item information and refine item selection. Though a variety of estimation techniques exist for fitting IRT models in data similar to those from the current study—for example the unidimensional graded response model (Samejima, 1969)—I chose to use Item Response Analysis by Exploratory Factor Analysis of polychoric correlations, as implemented in the *psych* package for R (Revelle, 2021). Using minimum residual factoring with a single factor, item-level polychoric correlations were analyzed and resultant loadings were transformed into item discrimination parameters. I also generated item-information plots and computed area under the item information curve for each candidate item.

Results and Discussion

Tests of Construct Validity

Associations of Trait Affiliation with various demographic variables are visualized using heat maps in Figure 4.3.

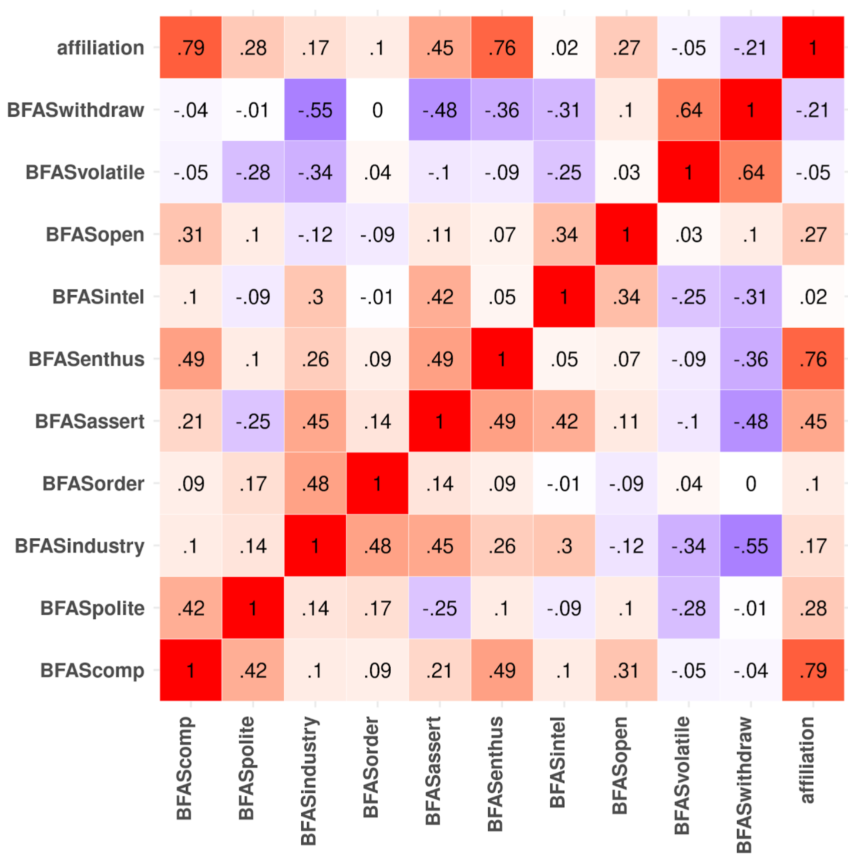
Figure 4.3.



Associations of Affiliation and demographic variables. p2edu = second parent education level, p1edu = first parent education level, exer = frequency of exercise, BMI = body mass index, marstatus = marital status, relstatus = relationship status.

Consistent with previous research on gender differences in Affiliation-related traits (Costa et al., 2001; Feingold, 1994; Weisberg et al., 2011), women in the current dataset tended to have higher scores on Trait Affiliation than men ($r = .32$). Higher Trait Affiliation was associated with slightly higher likelihood of being in a committed relationship ($r = -.13$) but not with marital status ($r = -.01$). Associations with other demographic variables (e.g., age, educational attainment, body mass index, and exercise frequency) were negligible (r's $< .10$).

Figure 4.4.

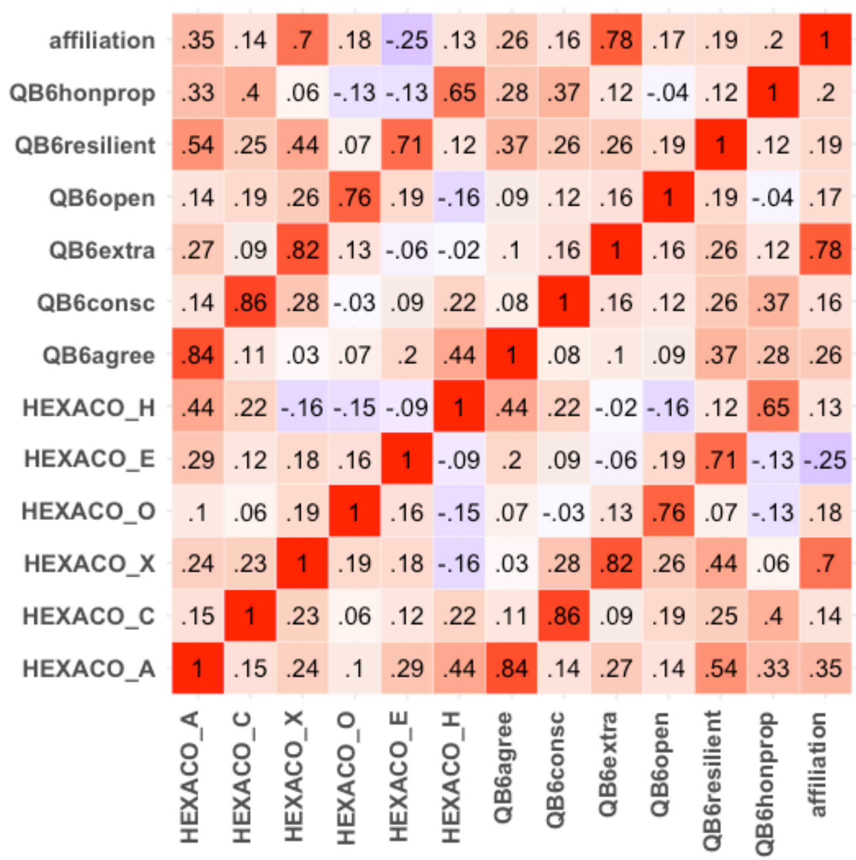


Associations of Affiliation and Big Five. BFAS = Big Five Aspect Scales, open = BFAS Openness, intel = BFAS Intellect, enthus = BFAS Enthusiasm, assert = BFAS assertiveness, order = BFAS Orderliness, industry = BFAS Industriousness, polite = BFAS Politeness, comp = BFAS Compassion.

As expected, based on how the scale items were selected, Trait Affiliation was highly positively correlated with BFAS Compassion (*r* = .79) and Enthusiasm (*r* = .76). The next highest correlations for Trait Affiliation were for BFAS Assertiveness (*r* = .45) and Politeness (*r* = .28), followed by BFAS Openness (*r* = .27) and Withdrawal (*r* = -.21).

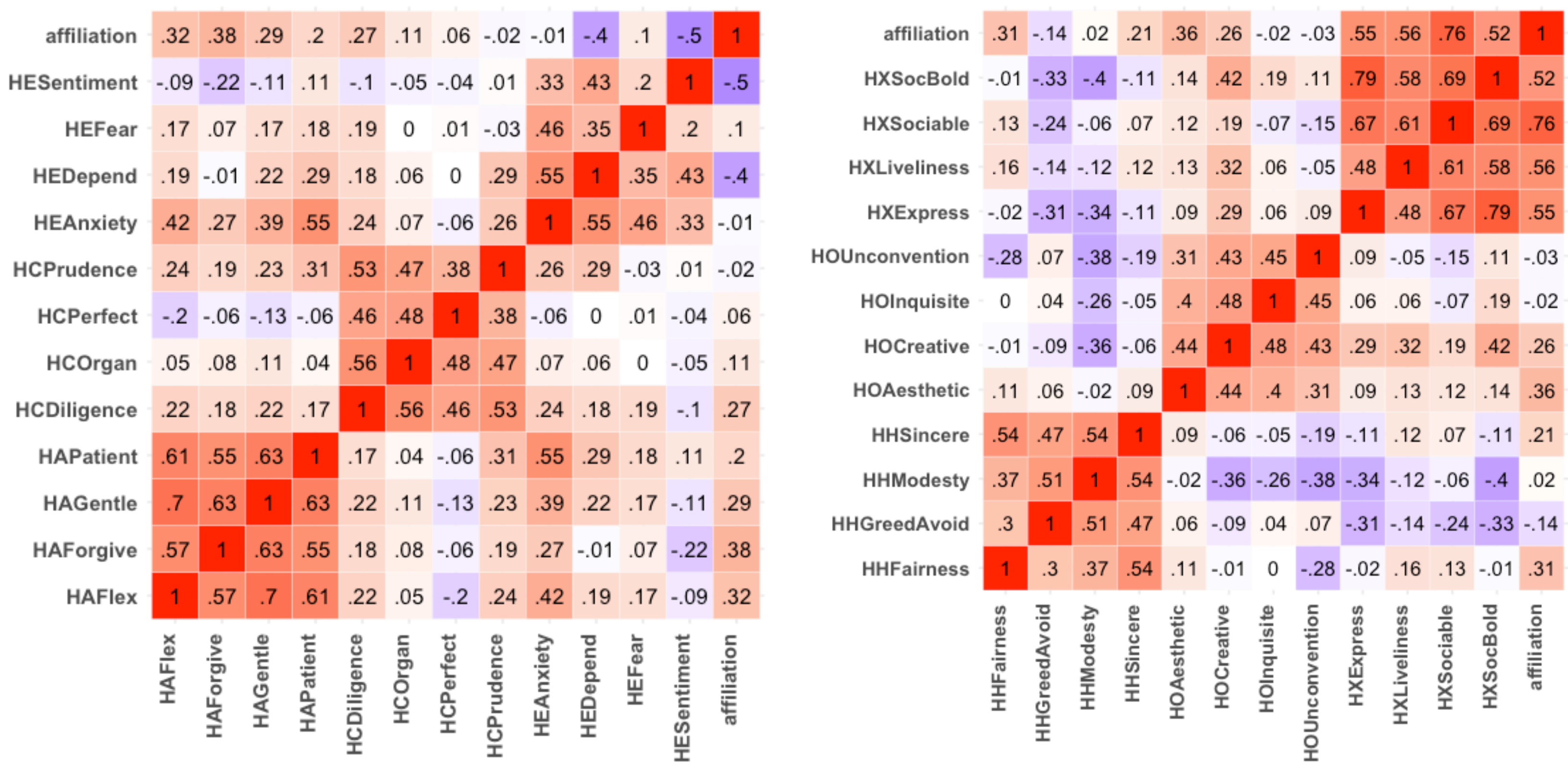
Scale-level correlations of Trait Affiliation with HEXACO/Big Six domains and facets are shown in Figures 4.5 and 4.6. The strongest associations at the domain level were positive associations of Trait Affiliation with Extraversion scales from the IPIP-HEXACO (*r* = .70) and QB6 (*r* = .76). Correlations of Trait Affiliation with Agreeableness were markedly smaller for both the IPIP-HEXACO (*r* = .35) and QB6 (*r* = .26), compared to Big Five Agreeableness. Attenuated associations compared to Big Five Agreeableness are consistent with known differences in how Agreeableness is operationalized in HEXACO vs. the Big Five; in six-factor personality models, Agreeableness is typically akin to a combination of high Politeness and low Volatility, whereas Compassion becomes a part of the Emotionality factor, along with high Withdrawal (Anglim & O'Connor, 2019; Ashton et al., 2014; Ludeke et al., 2019). In terms of associations at the facet level, the highest correlations with Trait Affiliation were seen for IPIP-HEXACO Sociability (*r* = .76), Liveliness (*r* = .56), Expressivity (*r* = .55), Social Boldness (*r* = .52), Entitlement (*r* = -.50), and Dependability (*r* = .40).

Figure 4.5.



Associations of Affiliation, Big Six, and HEXACO domains. QB6 = Big Six questionnaire, hon prop = Honesty-Propriety, resilient = Resiliency, open = Openness, extra = Extraversion, consc = Conscientiousness, agree = Agreeableness, H = Honesty-Humility, E = Emotionality, O = Openness to Ideas, X = Extraversion, C = Conscientiousness, A = Agreeableness.

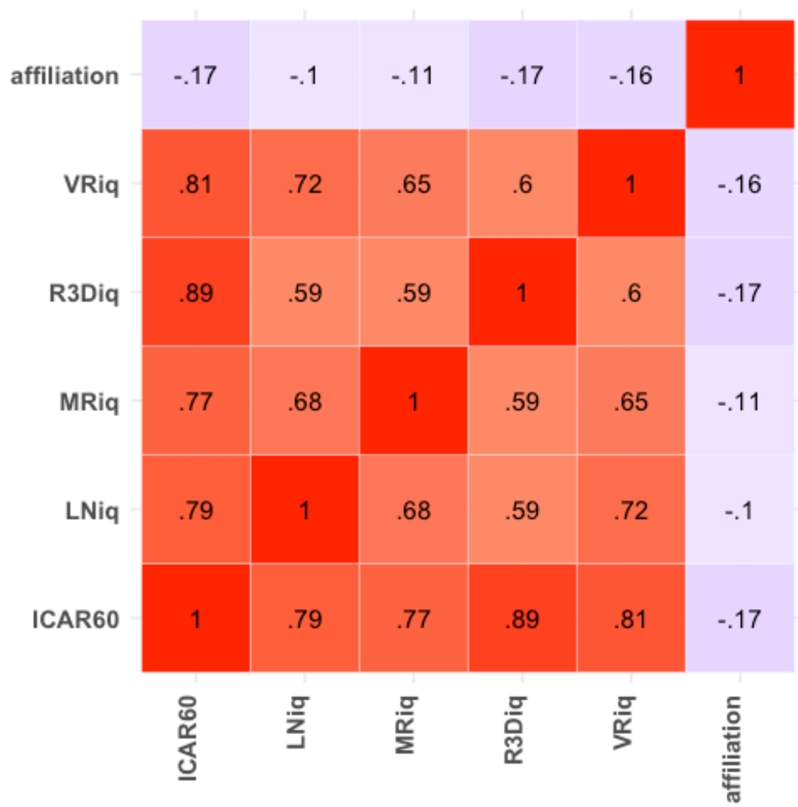
Figure 4.6.



Associations of Affiliation and HEXACO facets. HE = Emotionality, HC = Conscientiousness, HA = Agreeableness, HX = Extraversion, HO = Openness to Ideas, HH = Honesty-Humility, Sentiment = Sentimentality, Depend = Dependence, Perfect = Perfectionism, Organ = Organization, Patient = Patience, Gentle = Gentleness, Forgive = Forgiveness, Flex = Flexibility, SocBold = Social Boldness, Express = Expressivity, Unconvention = Unconventionality, Inquisite = Inquisitiveness, Creative = Creativity, Aesthetic = Aesthetic Appreciation, Sincere = Sincerity, GreedAvoid = Greed Avoidance.

Finally, scale-level correlations of Affiliation with cognitive ability, as measured by items from the ICAR, are visualized in Figure 4.7. Relatively small negative correlations were observed for Trait Affiliation with overall ICAR performance (r = -.17) and all ICAR subscales (with *r's* ranging from -.17 to -.10).

Figure 4.7.



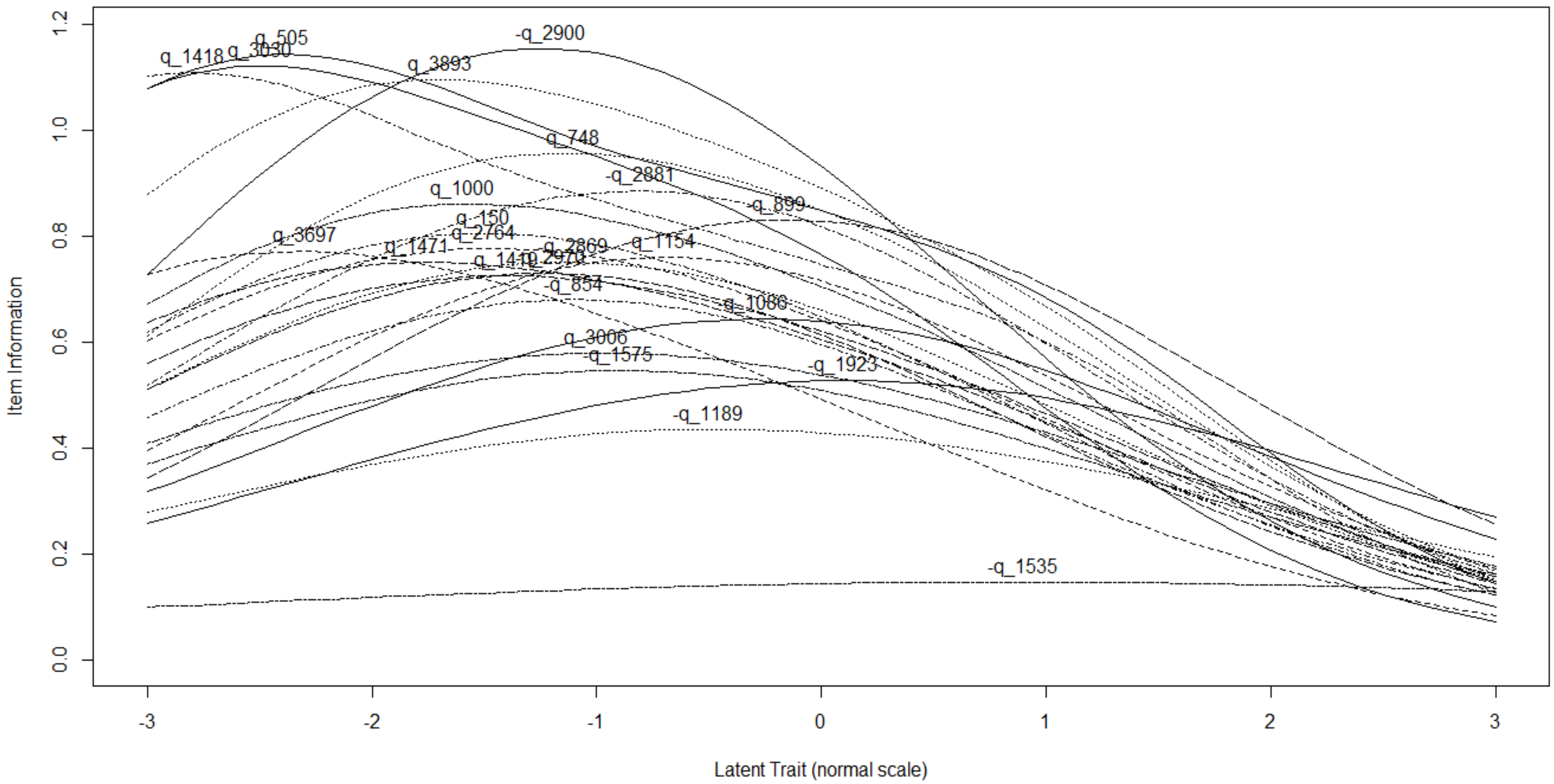
Associations of Affiliation and cognitive ability. VRiq = verbal reasoning, R3Diq = three-dimensional rotation ability, MRiq = matrix reasoning ability, LNiq = letter-number series ability, ICAR60 = general cognitive ability.

Scale Refinement

General factor saturation was reasonably high (ω_h = .59), providing some evidence for a strong general factor among the candidate items. Item information curves for the 24 Trait Affiliation Scale candidate items based are presented in Figure 4.8. Values for each item's discrimination parameter (a_i) and area under the information curve (AUC) are presented in Table 4.2.

Based on these additional IRT and factor analytic tests, I refined the selection of items, narrowing down the final version of the Trait Affiliation Scale to ten items. Items for the final version were selected to maximize item information and discrimination across levels of latent Trait Affiliation. I also removed any items that significantly reduced reliability and attempted to balance positively and negatively keyed items. The 10 items appearing in the final version of the scale are indicated in Table 4.2.

Figure 4.8.



Information curves for Trait Affiliation Scale candidate items.

Study 4c: Reliability and Validity of the Final Scale Form

Moving on from scale creation and refinement, Studies 4c through 4f focus on proving evidence of reliability and validity for the ten-item Trait Affiliation Scale. Two essential qualities for any well-performing personality instrument are internal consistency (at least when a measure is intended to assess an overarching dimension), as well as how strongly scores on a given scale are associated with scores on scales designed to measure similar constructs and, conversely, are not strongly associated with scores on measures of unrelated constructs (Litwin & Fink, 2003; Loevinger, 1957); Revelle & Condon, 2019).

In Study 4c, I examined internal consistency metrics and correlations of Trait Affiliation with various other personality constructs, across several independent datasets. To provide evidence of internal consistency reliability, I hoped to observe acceptable values of Cronbach's Alpha and Omega. To provide evidence of construct validity, I wanted to see evidence for both convergent validity—i.e., strong positive correlations with Compassion and Enthusiasm measures, moderate positive correlations with other Agreeableness and Extraversion measures, and moderate negative correlations with Antagonism and Detachment—and discriminant validity— i.e., minimal correlations with Big Five domains other than Agreeableness and Extraversion or with cognitive ability.

Method

Participants

Study 4c includes participants from Samples 1 (ESCS) and 2 (SAPA), which were described in the first two studies. In addition to Samples 1 and 2, Study 4c included data from four additional samples, described below:

Sample 3. This sample included 259 individuals (74 men and 185 women), recruited as part of a longitudinal study on personality, well-being, and personal projects. Participants' ages at Time 1 ranged from 19 to 25 years (*M* = 18.8, *SD* = 0.8). In terms of race, 159 participants identified as White (75.7%), 34 as Asian (13.1%), 5 as Latino (1.9%), 4 as Black (1.5%), and 18 as mixed race or other (6.9%). Of the 259 individuals tested at Time 1, 196 provided data at Time 2, 190 provided data at Time 3, and 151 provided data at Time 4.

Sample 4. This sample included 280 individuals (71 males, 203 females, and 6 non-binary or other), recruited as part of a study on personality and real-world social outcomes. Participants had a mean age of 20.5 (*SD* = 4.6). Recruitment took place at a liberal arts college in the Pacific Northwest, with a portion of subjects completing the research online and others in person. In terms of race, 204 participants identified as White (72.9%), 14 as Asian (5.0%), 24 as Hispanic/Latino (8.6%), 4 as Black (1.4%), 10 as Pacific Islander (3.6%), 3 as Native American (1.1%), and 21 as mixed race or other (7.5%).

Sample 5. This sample included 335 individuals, with ages ranging from 18 to 75 (*M* = 26.4, *SD* = 13.6). There were 267 females (79.7%), 67 males (20%), and 1 intersex individual (0.3%). In terms of race/ethnicity, 235 participants identified as White or Caucasian (70.1%), 59 as Asian or Pacific Islander (17.6%), 7 as Black or African American (2.1%), 4 as Latino or Hispanic (1.2%), and 30 as multiracial or other (9.0%). Participants were recruited via a combination of Qualtrics panels and from the campus of the University of Minnesota Twin Cities as part of a study on social cognition, personality, and psychopathology. Participants received an online informed consent document before beginning the study, then completed an online battery of questionnaires and behavioral tasks, including self-report measures of demographics, personality, psychopathology, and social functioning and several tests of social cognition. All protocols were approved by the University of Minnesota Institutional Review Board (ID# STUDY00003741).

Sample 6. This sample included 195 individuals (121 females) from ages 18 to 40 ($\bar{x}$ = 22.5, *SD* = 4.0). 91 participants were undergraduate students recruited on the campus of a small liberal arts college in the Pacific Northwest, and the remaining 104 individuals were recruited online using the Qualtrics Panel system. Participants received either course credit or monetary compensation for their participation. Participants completed modified online informed consent before beginning the study. All protocols were approved by the University of Minnesota Institutional Review Board (ID# STUDY00001844).

Procedure and Materials

Participants in all samples completed items from the Trait Affiliation Scale and BFAS, as well as providing demographic information. Participants in a subset of samples also completed measures of psychopathology (Samples 3 and 5), empathy (Samples 1, 3, and 5), and cognitive ability (Samples 2 and 3). To assess cognitive ability in Sample 3, participants completed a 16-item version of the ICAR (Condon & Revelle, 2014; Dworak et al., 2021). Measures not previously introduced are described below:

Personality Inventory for DSM-5 (PID-5; Krueger et al., 2012). To assess psychopathology in Sample 3, the PID-5 was administered at each of the four waves. The PID-5 is a questionnaire that includes 220 items rated on a four-point Likert scale. This inventory is used to assess symptoms of personality disorder in the alternative model from Section III of DSM-5. In the current study, I used participants' scores from the Antagonism and Detachment domain scales. Antagonism was computed as the average of facet-level scores for Manipulativeness, Deceitfulness, and Grandiosity. Detachment was computed as the average of facet-level scores for Anhedonia, Social Withdrawal, and Intimacy Avoidance.

Computer Adaptive Test of Personality Disorders: Static Form (CAT-PD; Simms et al., 2011; Wright & Simms, 2014). To assess psychopathology in Sample 5, the CAT-PD was administered. CAT-PD is a 216-item measure that assesses different dimensions of personality psychopathology. The form uses a 5-point Likert scale and allows for assessment of more than 20 facets of psychopathology, grouped into five broad categories similar to the Big Five (i.e., Negative Emotionality, Detachment, Antagonism, Disconstraint, and Psychoticism). The current study used the Antagonism and Detachment scales. Antagonism was computed as the average of facet-level scores for Callousness, Domineering, Grandiosity, Hostile Aggression, and Manipulativeness. Detachment was computed as the average of facet-level scores for Anhedonia, Emotional Detachment, Romantic Disinterest, and Social Withdrawal.

Interpersonal Reactivity Index (IRI; Davis, 1980). To assess empathy, participants in Samples 3 (at wave four only) and 5 were administered the IRI. The IRI is a multidimensional empathy questionnaire, including following subscales: Perspective Taking, Fantasy, Personal Distress, and Empathic Concern. The IRI consists of 28 items, to which participants respond with a five-point Likert scale. In the current study, I used only the Empathic Concern scale, which is the IRI subscale most related to Agreeableness and Compassion (Melchers et al., 2016).

Statistical Approach

Reliability and Structural Validity. To assess the internal consistency reliability of the final Trait Affiliation Scale items, I computed values of Cronbach's Alpha and Omega in six samples. Parallel analyses and Velicer's MAP test were also conducted in each sample, to examine the suggested factor structure of scale items.

Convergent and Discriminant Validity. To assess the construct validity of the scale, I computed correlations with various other personality variables and demographic constructs, including the Big Five, pathological personality traits, empathy, intelligence, age, and gender.

Results and Discussion

Reliability and Structural Validity

Internal consistency reliability and structural validity of the final scale form were investigated in a total of six samples. Values for Cronbach's Alpha, Omega total, and Omega hierarchical (using a two-group-factor bifactor model), in each sample are displayed in Table 4.4 (along with descriptive statistics). Across metrics and samples, observed alpha values suggested acceptable internal consistency. Values of omega total and omega hierarchical suggested adequate unidimensionality of the Trait Affiliation Scale, with at least one potential strong group factor.

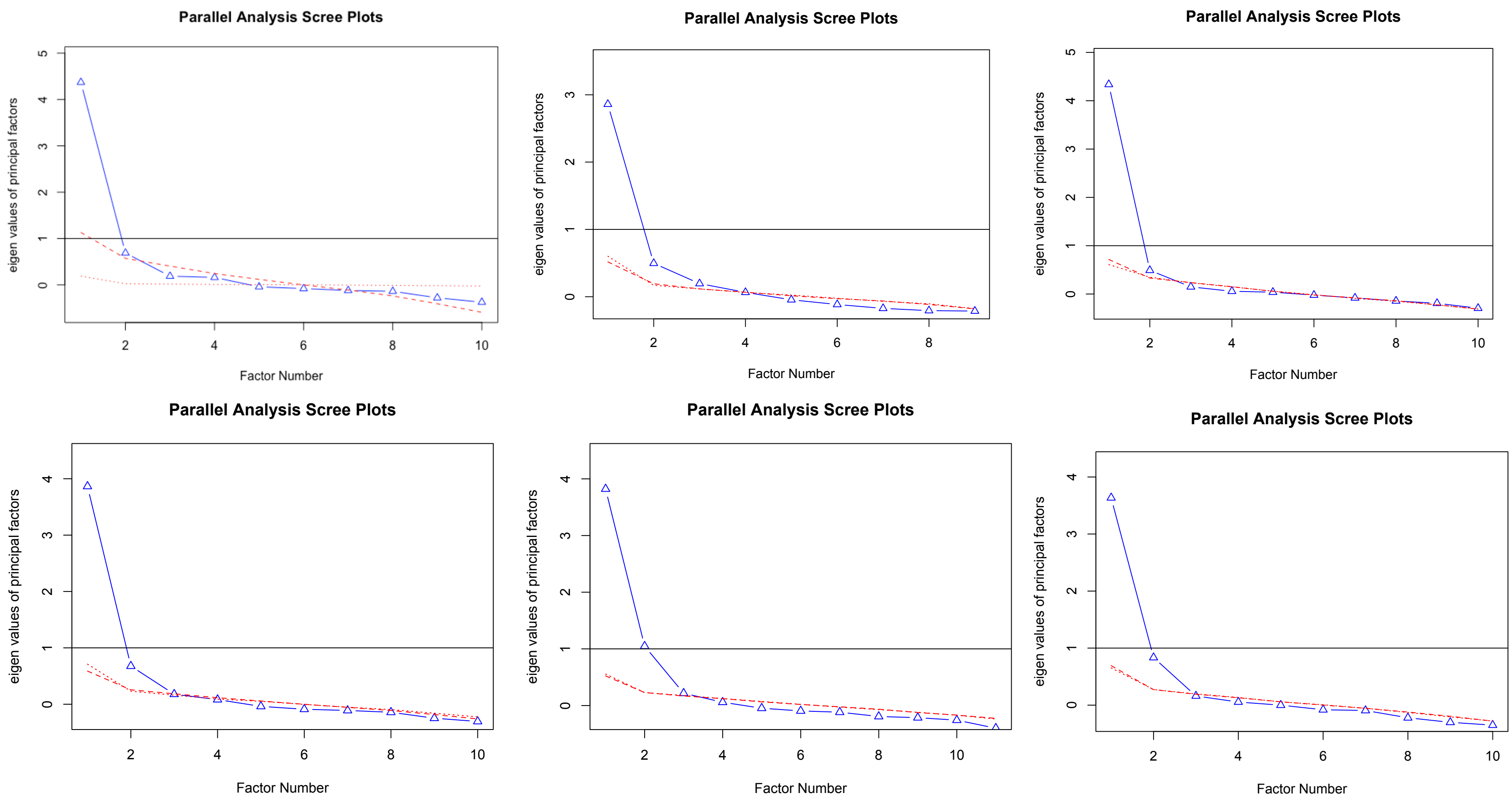
MAP tests yielded mixed results, with either one factor or two being suggested as optimal across the datasets. The results of parallel analyses and examination of scree plots in the datasets suggested a similar optimal number of factors, ranging from one to three (Figure 4.9). Closer examination of factor loadings suggested a first factor in all samples with relatively strong positive loadings and a second factor that was primarily related to item coding direction. Taken collectively, these findings suggest the Trait Affiliation Scale at least has a strong general factor but that continued investigation into the potential multidimensionality of Trait Affiliation may be warranted.

Table 4.4. *Descriptive statistics and internal consistency for the Trait Affiliation Scale across multiple samples*

Sample	Mean (SD)	Skew	[Min, Max]	Cronbach's α	ω	ω_h	MAP-1	MAP-2	MAP-3
1. ESCS (*N* = 409)	4.0 (0.6)	-0.6	[2.2, 5.0]	.80	.83	.58	.032	.044	.069
2. SAPA (*N* = 25732)	—	—	—	.88	.91	.62	.044	.043	.064
3. PPA Wave 1 (*N* = 259)	3.9 (0.5)	-0.4	[2.5, 4.9]	.82	.84	.42	.026	.060	.097
PPA Wave 2 (*N* = 196)	3.9 (0.4)	-0.2	[2.6, 4.9]	.81	.85	.65	.040	.055	.071
PPA Wave 3 (*N* = 190)	3.9 (0.4)	-0.2	[2.8, 5.0]	.80	.85	.49	.040	.044	.071
PPA Wave 4 (*N* = 151)	4.0 (0.5)	-0.5	[1.9, 5.0]	.87	.90	.69	.029	.038	.057
4. YW (*N* = 280)	4.0 (0.5)	-1.0	[1.2, 5.0]	.85	.88	.57	.037	.037	.055
5. SBSC (*N* = 335)	4.0 (0.6)	-0.5	[1.7, 5.0]	.85	.89	.54	.051	.030	.049
6. SBAV (*N* = 195)	3.9 (0.6)	-0.6	[1.6, 5.0]	.84	.88	.53	.045	.037	.060

Note. ESCS = Eugene Springfield Community Sample, SAPA = Synthetic Aperture Personality Assessment, PPA = Personality Projects Analysis Dataset, YW = Yanna Weisberg dataset on social goals and behaviors, SBSC = Scott Blain dataset on social cognition and personality, SBAV = Scott Blain dataset on induction of affiliative states, MAP-1, 2, 3 = Velicer's Minimum Average Partial for 1, 2, and 3 factors.

Figure 4.9.



Parallel analysis scree plots for the Trait Affiliation Scale final form. Top left = SAPA, Top middle = ESCS, Top right = PPA Wave 4, Bottom left = YW, Bottom middle = SBSC, Bottom right = SBAV

Convergent and Discriminant Validity

Validity of the final scale form was investigated in a total of six samples, by examining associations with various personality constructs. Correlations of the Trait Affiliation Scale with other personality variables are displayed in Table 4.5, organized by sample. Across all samples, the highest observed correlations with normal-range personality constructs were seen for Compassion and Enthusiasm, consistent with my research group's theory of Trait Affiliation and how the scale items were empirically derived; next highest correlations were seen for domain-level Extraversion and Agreeableness, with smaller correlations for Assertiveness, Politeness, and the other Big Five domains. There were also strong positive correlations with measures of empathy (i.e., the empathic concern scale from the IRI) and strong negative correlations with measures of Detachment (low Extraversion associated with psychopathology). General cognitive ability was not related to Trait Affiliation in Sample 3, but these variables showed a (relatively weak) negative correlation in Sample 2. Finally, in multiple samples, I observed associations with gender and age, with women and younger individuals tending to have higher Trait Affiliation.

Table 4.5. Construct (convergent and discriminant) validity for the Trait Affiliation Scale across multiple samples (Pearson Correlations)

Sample	A	E	Comp.	Enth.	Polite.	Assert.	Empath.	Antag.	Detach.	IQ	Sex/ Gender	Age	C	N	O/I
1. ESCS (*N* = 409)	.56	.59	.64	.66	.30	.32	.52	—	—	—	-.34	.01	.19	-.12	.24
2. SAPA (*N* = 25732)	.67	.70	.76	.71	.27	.40	—	—	—	-.16	-.29	-.03	.16	-.14	.16
3. PPA Wave 1 (*N* = 259)	.38	.63	.70	.69	.25	.36	—	-.08	-.63	.04	-.03	-.11	.03	-.19	.23
PPA Wave 2 (*N* = 196)	.57	.62	.74	.72	.21	.32	—	-.07	.61	—	—	—	.13	-.13	.24
PPA Wave 3 (*N* = 190)	.59	.61	.77	.73	.29	.30	—	-.16	-.58	—	—	—	.12	-.22	.27
PPA Wave 4 (*N* = 151)	.60	.63	.76	.78	.31	.27	.60	-.11	.66	—	—	—	.06	-.29	-.30
4. YW (*N* = 280)	.67	.56	.74	.66	.42	.31	—	—	—	—	-.20	-.14	.11	.04	.17
5. SBSC (*N* = 335)	.66	.65	.76	.72	.42	.43	.50	-.47	-.70	—	-.19	-.16	.28	-.30	.34
6. SBAV (*N* = 195)	.49	.51	.65	.64	.04	.24	—	—	—	—	-.09	-.26	.15	-.24	.34

Note. ESCS = Eugene Springfield Community Sample, SAPA = Synthetic Aperture Personality Assessment, PPA = Personality Projects Analysis Dataset, YW = Yanna Weisberg dataset on social goals and behaviors, SBSC = Scott Blain dataset on social cognition and personality, SBAV = Scott Blain dataset on induction of affiliative states, A = Agreeableness, E = Extraversion, Comp. = Compassion, Enth. = Enthusiasm, Polite. = Politeness, Assert. = Assertiveness, Empath. = Empathy, Antag. = Antagonism, Detach. = Detachment, C = Conscientiousness, N = Neuroticism, O/I = Openness/Intellect

The pattern of results seen across a range of samples and measures gives support for the reliability and validity of the new Trait Affiliation Scale. Trait Affiliation was consistently related to participants' scores on theoretically related measures—including Compassion, Enthusiasm, Agreeableness, Extraversion, Empathy, and Detachment—in the hypothesized directions, providing evidence for convergent validity. Correlations between Trait Affiliation and theoretically unrelated constructs—including other Big Five domains and cognitive ability—were generally low (in absolute value), providing evidence of discriminant validity.

Study 4d: Test-Retest Reliability

In addition to establishing the internal consistency and convergent validity of the new Trait Affiliation Scale, it is also vital to establish the test-retest reliability of this measure. As traits are defined by their relative stability over time (including across varying situations and throughout development), a trait questionnaire should provide similar scores for any given participant when taken at different timepoints. It is important to assess test-retest reliability apart from internal consistency, as having one form of reliability does not necessarily ensure that a given measure will display other forms of reliability (Leppink & Perez-Fuster, 2017). In Study 4d, I examined the test-retest reliability of the Trait Affiliation Scale, using a four-wave longitudinal dataset.

Method

Participants, Procedure, and Materials

This study used only participants from Sample 3 (described above). Participants provided data at four time points including 259 individuals at Time 1, 196 at Time 2, 190 at Time 3, and 151 at Time 4. At each of the time points, participants completed items from the Trait Affiliation Scale, as well as a variety of other questionnaires and a personal projects analysis task. Time points were spaced at approximately six-month intervals.

Statistical Approach

To quantify test-retest reliability, I first computed bivariate Pearson correlations among participants' Trait Affiliation Scale scores across each pair of time points, using pairwise deletion. Because the appropriateness of using Pearson correlations to assess test-retest reliability has been called into question (Aldridge et al., 2017; Koo & Li, 2016) and to facilitate simultaneous analysis of participants' scores across more than two timepoints, I also computed the intraclass correlation (ICC) among participants' Trait Affiliation Scale scores. Specifically, I computed a single-measurement, absolute-agreement, 2-way mixed-effects ICC (Aldridge et al., 2017; Koo & Li, 2016). For the sake of comprehensiveness, I computed and report ICCs using both pairwise and listwise deletion. 95% confidence intervals were computed for the ICCs.

Results and Discussion

Bivariate correlations of Trait Affiliation Scale scores across different time points are presented in Table 4.6. All correlations were strong, positive, and statistically significant, ranging from .70 to .84 (*p's* < .001).

Table 4.6. *Correlations among Trait Affiliation Scale scores at various time points*

	1	2	3	4
Time 1	–			
Time 2	.77**	–		
Time 3	.70**	.83**	–	
Time 4	.71**	.80**	.84**	–

Note. *p < .05, **p < 0.01

When computed using all participants in the dataset (*N* = 259), the intraclass correlation among Trait Affiliation Scale scores was .76 (95% CI: [.72, .79]). When computed using only participants with data points at all four times (*N* = 142), the intraclass correlation among Trait Affiliation Scale scores was .77 (95% CI: [.73, .81]).

Using established criteria, bivariate and intraclass correlation values suggest acceptable test-retest reliability for the Trait Affiliation Scale (Aldridge et al., 2017; Cicchetti, 1994; Koo & Li, 2016; Portney & Watkins, 2015; Shrout & Fleiss, 1979; Matheson, 2019). Given that test-retest reliability is often measured at shorter time intervals than those in the current dataset—i.e., the typical days or weeks vs. roughly six-month intervals in the current study (Marx et al., 2003; Park et al., 2018; Streiner et al., 2014)—and that college students in young adulthood are at a peak phase of life for personality change (Robins et al., 2001), results are particularly promising and suggest that Trait Affiliation as measured using the new scale is a relatively stable personality trait.

Study 4e: Criterion and Incremental Validity

In addition to providing evidence of convergent and discriminant validity by examining associations with existing personality questionnaires, another essential component of scale validation should be to provide evidence of criterion or predictive validity. Since personality traits represent stable cognitive, motivational, emotional, and behavioral patterns, scores on a given questionnaire should be able to predict real-world outcomes of interest relevant to the domain of functioning represented by a given trait (McAdams & Pals, 2006). Consequently, various personality traits have been studied in relation to social interaction, social cognition, and interpersonal outcomes (McCrae & John, 1992; Roberts et al., 2007; Dolan & Fullam, 2004). Specific interpersonal outcome variables that have been assessed include social network size and levels of social motivation and engagement, which can be investigated in relation to personality traits like Extraversion, Agreeableness, and Neuroticism (Pollet et al., 2011; Roberts et al., 2008).

Research supports a positive association between Extraversion and the general size of one's social network, operationalized as the number of people participants contacted in the last month or by asking participants to list people with whom they are close (Pollet et al., 2011; Roberts et al., 2008; Selfhout et al., 2010). Some evidence suggests this positive association with Extraversion is specific to individuals' support clique—the number of individuals in one's inner circle of close friends (Roberts et al., 2008). Others have also found positive associations between social network size and Agreeableness (Selfhout et al., 2010; Zhu et al., 2013).

Though a limited amount of research has examined social network size and social engagement in relation to levels below the Big Five in the personality trait hierarchy, I anticipated that outcomes such as social network size and other measures of social engagement would also be related to scores on the Trait Affiliation Scale, given its measurement content, theoretical importance for close relationships, and association with Agreeableness and Extraversion. In addition to social network size and social behavior, another potential correlate of Trait Affiliation is individual differences in social cognitive abilities such as emotion perception and theory of mind. Performance on tests of social cognition have been linked to individual differences in Agreeableness and its subfactors (Allen et al., 2017; Chapter 1 of this dissertation; Nettle & Liddle, 2008) and performance in such domains may contribute to individual differences in the interstitial trait of Affiliation as well.

In Study 4d, I examined the criterion and incremental validity of the Trait Affiliation Scale. Specifically, I sought to examine whether the scale could predict scores on real-world outcomes of interest, such as measures of social behavior, social cognitive ability, and interpersonal interactions. I also wanted to see evidence that the scale has incrementality validity—whether the scale can predict these interpersonal outcomes better than variance explained by existing measures of Compassion and Enthusiasm alone. Thus, in this study, I examined how Trait Affiliation was associated with participants' social goals, frequency of social events and behaviors, social network size, and social cognitive abilities, before and after controlling for other relevant variables (i.e., Compassion, Enthusiasm, gender, age, Politeness, and Assertiveness).

Method

Participants, Procedure, and Materials

This study included participants from Samples 4 and 5, which were all described in further detail in Study 4c. Participants in Sample 4 completed the Trait Affiliation Scale, BFAS, and a questionnaire assessing social goals, events, and behaviors. Sample items from the inventory of social goals, events, and behaviors are presented in Table 4.7.

Table 4.7. *Sample items for Study 4e social goals and behaviors*

Social Goals
I am trying to deepen my relationships with my friends.
I am trying to move toward growth and development in my friendships.
I am trying to enhance the bonding and intimacy in my close relationships.
I am trying to share many fun and meaningful experiences with my friends.
I am trying to avoid disagreements and conflicts with my friends.
I am trying to stay away from situations that could harm my friendships.
I am trying to avoid getting embarrassed, betrayed, or hurt by any of my friends.
I am trying to make sure that nothing bad happens to my close relationships.
Daily Events
Spent pleasant or relaxing time with friends/date/family.
Had a disagreement or conflict with a friend, boyfriend/girlfriend, or family member.
Received a compliment on my physical appearance.
Something happened that made me feel awkward or embarrassed in public.
Did not have enough privacy.
Had an unpleasant interaction with someone other than a friend, boyfriend/girlfriend, or family member.
Friends were not available when I wanted to socialize.
Had problems controlling negative feelings.

Others did not do something that I wanted them to do.
Others acted disinterested in something I said or did.
Had other type of unpleasant event (not listed above) with friends, family, or date.
Did something special for a friend/date that was appreciated.
Had especially good interactions with friend(s) or acquaintance(s).
Went out socializing with friends/date (e.g., party, club).
Went out to eat with a friend/date.
Provided support to someone I care for.
Had other type of pleasant event (not listed above) with friends, family, or date.
Social Events
I did something with my friends because I didn't want to be left out.
Others acted disinterested in something I said or did.
I went out with my friends or romantic partner even though I would have preferred to stay home.
Something happened that made me feel awkward or embarrassed in public.
I had a minor disagreement/conflict with a friend, romantic partner, or family member.
I went to an activity/place/event alone because none of my friends could go with me.
I had a major disagreement/conflict with a friend, romantic partner, or family member.
My friends were not available when I wanted to socialize.
I went out socializing with friends/date (e.g. party, dinner, club).
I had especially good interactions with friend(s) or acquaintance(s).
I did something special for a friend/steady date which was appreciated.
I started a friendly conversation with someone I did not know.
I sent an pleasant letter or e-mail to a friend, romantic partner, or family member.
I spent pleasant time with friends/romantic partner/family in a relaxed setting.
I asked someone out on a date.
I gave support to someone, a friend, family member, or romantic partner.

Key variables of interest were computed as the means of participants' ratings for approach friendship goals (e.g., "I am trying to deepen my relationships with my friends"), avoidance friendship goals (e.g., "I am trying to stay away from situations that could harm my friendships"), positive social events (e.g., "Spent pleasant or relaxing time with friends/date/family"), negative social events ("Others acted disinterested in something I said or did"), general positive events (e.g., "Had other type of pleasant event (not listed above) with friends, family, or date"), and general negative events (e.g., "Had problems controlling negative feelings").

Participants in Sample 5 completed the Trait Affiliation Scale, BFAS, and a variety of measures assessing interpersonal functioning, including a measure of social network size and tests of social cognitive ability. Sample 5 social outcome measures are further described below:

Lubben Social Network Scale – Revised (LSNS-R). The LSNS-R (Lubben, 1988) was used to measure social network size in the current study. The questionnaire has 12 items, half of which focus on family-related social engagement and the other half of which focus on friend or non-kin-related social engagement. All questions were rated on a Five-point Likert scale and final scores were computed as the total score of all questions, with higher scores representing higher social engagement and larger social network size. I also computed separate LSNS sub-scores for family and friends.

Theory of Mind Vignettes. The ToM vignette task (Stiller & Dunbar, 2007) uses a series of five short stories depicting social situations. Each story describes a social interaction involving multiple characters. Participants are asked to read each story to themselves twice, after which they answer five ToM questions and five memory questions pertaining to the story. All questions are in true-false format. Memory questions are designed to measure the participants' ability to retain the factual contents of the story, and the number of facts that the participant must retain varies between two and six in each question. Performance on memory questions within the task can be used as a covariate to ensure that any associations with variables of interest are due to participants' ToM ability rather than their memory for the details of the story. ToM questions require that the participant reason, or infer, a character's perspective in the story. To assess performance on the task, I adopted the procedure used by Nettle and Liddle (2008) and Allen et al. (2017) and computed simple sums of correct responses to memory questions and ToM questions for each participant.

Tricky Triangles Task. In the triangles task (Abell et al., 2000), participants are presented with a series of computerized animations of shapes interacting in a way that was random, physical, or social. In the random condition, the shapes did not interact with each other, but rather moved around purposelessly (e.g., bouncing or drifting). In the physical condition, the shapes moved in a way that could be identified as a particular goal-directed activity not involving ToM or mentalizing (e.g., fishing or swimming). The social condition included a variety of interaction types, each demonstrating a particular social sequence, such as coaxing, seducing, or mocking. Participants were tasked with indicating whether each animation was random, physical, or social in nature, then scored for their accuracy in correctly categorizing each animation in a series of 22 clips.

Reading the Mind in the Eyes Task. The eyes task (Baron-Cohen et al., 2001) consists of 36 grey-scale photos of people taken from magazines. These photos were cropped and rescaled so that only the area around the eyes could be seen. Each photo was accompanied by four mental state terms, from which the participant was instructed to choose the word that best described what the person in the photo was thinking or feeling. Only one of the four items was deemed correct (as judged by consensus from an independent panel of judges in the initial psychometric study). Participants were scored for their accuracy across all 36 stimuli.

Statistical Approach

First, I calculated descriptive statistics for task and personality measures. Then, bivariate Pearson correlations were computed to test for associations of Trait Affiliation with interpersonal variables of interest. Finally, I computed a series of multiple regression models to test whether Trait Affiliation was still associated with interpersonal outcome variables 1) after controlling for variance in Compassion and Enthusiasm, 2) after controlling for these variables as well as age and gender, and 3) after controlling for these variables as well as Politeness and Assertiveness.

Finally, I computed these models using a latent social cognition variable rather than just observed scores on the individual tasks. Accuracy scores from the three social cognition tasks were used as indicators of a single latent variable, predicted by 1) Affiliation controlling for variance in Compassion and Enthusiasm, 2) Affiliation controlling for these variables as well as age and gender, and 3) Affiliation controlling for these variables as well as Politeness and Assertiveness. Maximum likelihood estimation was used, and common fit indices were computed, including the root mean squared error of approximation (RMSEA), standardized root mean square resdiaul (SRMR), and Tucker Lewis index (TLI). The Latent Variable Analysis (LAVAAN) package for R was used for estimating this latent variable model (Rosseel, 2012). Performance on the memory condition of the vignette task was included as an additional predictor in models predicting the social cognition latent variable and those predicting performance on the mentalizing condition of the vignette task.

Results and Discussion

Bivariate correlations and regression coefficients (controlling for Enthusiasm and Compassion) of Trait Affiliation with social outcomes and accuracy on the various social cognition tasks are presented in Table 4.8. Trait Affiliation was positively correlated with frequency of approach friendship goals, avoidance friendship goals, positive social events, and general positive events, whether or not I controlled for other variables of interest. Trait Affiliation was not associated with frequency of negative events (general or social), at either the zero-order or partial-correlation level.

Trait Affiliation was positively correlated with social network size (for overall LSNS and the friends subscale), whether or not I controlled for other variables of interest. Correlations were stronger for the friends compared to the family subscale. There were significant positive bivariate correlations between Trait Affiliation and accuracy, for all of the social cognition tasks. After controlling for relevant BFAS aspects, however, Trait Affiliation was only significantly associated with performance on the eyes task and with the latent variable for social cognition. The latent variable model testing the bivariate association with Affiliation showed margin fit (RMSEA = .124, SRMR = .044, TLI = .883), whereas models showed acceptable fit across other versions including the model controlling for Compassion and Enthusiasm (RMSEA = .081, SRMR = .034, TLI = .953), age and gender (RMSEA = .066, SRMR = .030, TLI = .950), and Politeness and Assertiveness (RMSEA = .053, SRMR = .026, TLI = .961).

Table 4.8. *Associations of Trait Affiliation and Study 4e interpersonal variables*

Criterion Variable	r_{aff}	p	$\beta_{aff\text{-}1}$	p	$\beta_{aff\text{-}2}$	p	$\beta_{aff\text{-}3}$	p
Sample 4								
Approach Friendship Goals	.55	< .001	.35	.001	.34	.004	.32	.005
Avoidance Friendship Goals	.35	< .001	.38	.001	.35	.002	.42	.001
Positive Social Events	.47	< .001	.34	.001	.29	.009	.24	.023
Negative Social Events	.07	.302	.16	.161	.14	.247	.21	.086
Positive Event Frequency	.45	< .001	.42	< .001	.36	.001	.28	.009
Negative Event Frequency	.12	.057	.17	.153	.14	.258	.23	.059
Sample 5								
Social Network Size	.43	< .001	.32	.001	.23	.012	.22	.015
Social Network Size (Family)	.32	< .001	.20	.044	.16	.119	.14	.170
Social Network Size (Friends)	.43	< .001	.34	< .001	.23	.007	.24	.005
Eyes Task	.24	< .001	.27	.008	.20	.042	.20	.048
Triangles Task	.19	.001	.19	.059	.16	.122	.16	.121
Vignettes Task	.20	< .001	.13	.093	.08	.276	.10	.213
Latent Social Cognition	.22	< .001	.25	.004	.20	.022	.21	.015

Note. r_{aff} = Pearson correlation between Trait Affiliation and the given criterion variable, $\beta_{aff\text{-}1}$ = Standardized regression coefficient of Trait Affiliation predicting the given criterion variable controlling for Compassion and Enthusiasm, $\beta_{aff\text{-}2}$ = Standardized regression coefficient of Trait Affiliation predicting the given criterion variable controlling for Compassion, Enthusiasm, age, and gender, $\beta_{aff\text{-}3}$ = Standardized regression coefficient of Trait Affiliation predicting the given criterion variable controlling for Compassion, Enthusiasm, age, gender, Politeness, and Assertiveness.

Study 4f: Inducing Affiliative States

In addition to establishing associations of the Trait Affiliation Scale with existing personality questionnaires and more direct measures of social behavior and engagement, it is important to establish relations with individuals' state (i.e., moment-to-moment) experiences of warmth, affection, and a desire for Affiliation. As trait levels of a construct can be conceptualized as probability density distributions of states over time (Fleeson, 2001), one would expect scores on a valid trait measure of a given construct (i.e., Trait Affiliation) to correlate with state measures of the same construct (i.e., state Affiliation). In establishing the criterion and incremental validity of a given measure, levels of the given trait measure should predict state levels, to a degree beyond the utility of similar constructs existing in the literature. In this specific case, I sought to establish the strength of associations between state Affiliation and scores on the Trait Affiliation Scale, compared to existing measures of Extraversion, Agreeableness, and their component aspects. Further, I expected this pattern of associations to be present, both at rest and after the targeted induction of Affiliative states.

Emotion induction refers to the use of psychological techniques to elicit specific emotional or other affective states (Velten, 1968). The use of emotion induction to study psychological states has a long and rich history (Gross & Levenson, 1995; Ochsner et al., 2002; Polivy, 1981; Van Rooijen & Vlaander, 1984; Velten, 1968), with multiple stimulus modalities employed. These stimulus types have included viewing of images/imagery (e.g., Lang, 1979; Lang, et al., 1988; Wagner, 1990), interactions with trained confederates (e.g., Ax, 1953), hypnosis (e.g., Bower, 1983), repeating phrases (e.g., Velten, 1968), facial muscle movements (e.g., Ekman et al., 1983), and listening to music (e.g., Sutherland et al., 1982). However, perhaps the most consistent and effective method for inducing emotional/affective states has been the use of film clips (e.g., Gross & Levenson, 1995; Lazarus et al., 1962; McHugo et al., 1982; Philippot, 1993). Film clips are effective at inducing emotions because of their multimodal nature, utilizing moving images, dialogue, and music to elicit a specific emotional state. Though batteries of standardized emotional film clips exist for the induction of states such as amusement, anger, fear, and disgust (e.g., Gross & Levenson, 1995), similar standardized stimuli for more specific emotional states are harder to find.

As for states reflecting Affiliation specifically, I currently know of a single study establishing a stimulus specifically chosen and validated for the induction of affiliative states (Morrone-Strupinsky & Depue, 2004). Morrone-Strupinsky and Depue established a film stimulus that was specifically associated with induction of an affiliative state (feelings of warmth and affection), compared to a state of agentic Extraversion (feelings of enthusiasm, elation, or excitement). Increases in feelings of warmth and affection were positively associated with participants' scores on the MPQ Social Closeness scale, but not other components of Extraversion (Morrone-Strupinsky & Depue, 2004). These results suggest that changes in state Affiliation after watching targeted emotion induction stimuli may be predicted by trait levels of similar constructs (i.e., Social Closeness or Trait Affiliation). Although these results are promising for the current work, I wanted to examine associations of Trait Affiliation with induced state Affiliation across a range of affiliative film stimuli, rather than relying on a single film clip. Further, I wanted to establish specificity of the relations between state (and change in state) Affiliation with Trait Affiliation, compared changes in affiliative states predicted by measures of Compassion and Enthusiasm alone.

Given the importance of establishing associations between the Trait Affiliation Scale and affiliative states, Study 4f sought to examine associations between participants' Trait Affiliation and experience of warmth, affection, and desire to bond with others, before and after viewing film stimuli chosen to elicit these affiliative states. Film clips were chosen to display a variety of affiliative relationships, including parent-child, romantic, and friendship contexts. Affiliative cues in the film clips included physical touch/proximity, body language, narrative elements/dialogue, and verbal cues related to love/Affiliation.

I hypothesized that Trait Affiliation Scale scores would be strongly associated with levels of state Affiliation before and after emotion induction and would offer unique predictive utility above and beyond variance in state Affiliation explained by Extraversion, Agreeableness, and their corresponding aspects. Moreover, I believed change in state Affiliation after emotion induction, but not change in general positive emotionality, would be predicted by levels of Trait Affiliation; again, I anticipated the predictive ability of scores on the Trait Affiliation Scale to exceed that of relevant Big Five aspects.

Method

Participants and Procedure

Participants in Study 4f were those from Sample 6, introduced in Study 4c. First, participants completed a set of pre-emotion questions. These questions asked participants to describe how strongly they were currently experiencing each of a set of six statements, using a seven-point Likert scale. Four adjective-pair items included "Warm and Affectionate," "Compassionate and Kind," "Energetic and Happy," and "Calm and at Ease." Two additional items asked participants to rate their current desire to "spend time with loved ones" and "connect with others." These items were largely adapted from the PNAS-X (Watson & Clark, 1999) and research by Morrone-Strupinsky and Depue (2004).

After the pre-emotion questions, participants completed the BFAS and Trait Affiliation Scale, as discussed in previous studies. Participants then watched a set of three affiliative film stimuli, randomly selected for each participant from a pool of seven film clips. Videos were roughly two to three minutes in length, each. After each video, participants again answered the seven emotion questions noted above. The order of videos was consistent across participants, but selection of videos was randomized.

Affiliative Video Stimuli

Video stimuli were chosen to elicit an affiliative response, focused on specific aspects of Affiliation including friendship, romance, and familial bonds. Videos included segments from the following films/sources: *Juno, Forrest Gump, Blue Jay, Love Actually, The Spectacular Now,* a 2012 Olympics commercial, and an Extra Gum commercial. Clips included affiliative markers such as physical touch, presentation of parent-child relationships, initiation of new relationships, and verbal indicators of love/friendship.

Composite variables were created from the pre- and post- emotion questions. A composite Affiliation variable for pre- questions and for each video was calculated by averaging across the items “warm and affectionate,” “compassionate and kind,” “desire to spend time with loved ones,” and “desire to connect with others.” Composite positive emotion was averaged by averaging across the items “Energetic and Happy,” and “Calm and at Ease.” Finally, total composite post- variables were calculated by respectively averaging positive emotion and Affiliation variables corresponding to the three videos each participant watched.

Statistical Approach

First, descriptive statistics were calculated for trait and state items. Then, correlations were calculated to assess relations among pre-Affiliation, post-Affiliation, and scores on the Trait Affiliation Scale. The partial correlation of Trait Affiliation with post-Affiliation, controlling for pre-Affiliation was also computed. Subsequently, to determine the utility of the new Trait Affiliation Scale in predicting affiliative emotional response, a series of multiple regression models were computed, controlling for 1) Compassion and Enthusiasm, 2) these variables in addition to gender and age, and 3) these variables in addition to Politeness and Assertiveness. Finally, to determine specificity of these relations, similar analyses were done, using Trait Affiliation as a predictor of the positive emotion variables.

Results and Discussion

Descriptive statistics for self-report variables are presented in Table 4.9.

Table 4.9. *Descriptive statistics for Study 4f individual difference and task measures (across all videos)*

	Mean (SD)	Skew	[Minimum, Maximum]
Pre-Affiliation	5.0	-0.5	[1, 7]
Pre-Positive Emotion	4.3	-0.1	[1, 7]
Post-Affiliation	5.2	-0.8	[1, 7]
Post-Positive Emotion	4.7	-0.3	[1.2, 7]
Trait Affiliation	3.9	-0.6	[1.6, 5.0]
BFAS A	3.4	-0.1	[2.2, 4.5]
Compassion	3.2	-0.3	[1.3, 4.4]
Politeness	3.5	-0.3	[1.6, 4.7]
BFAS E	3.4	-0.3	[1.6, 5.0]
Enthusiasm	3.5	-0.5	[1.2, 5.0]
Assertiveness	3.3	-0.2	[1.5, 4.9]

Across videos, there was a significant increase in both Affiliation ($t_{(193)} = 3.1$, $p = .002$, $d = 0.45$) and positive emotion ($t_{(194)} = 3.8$, $p < .001$, $d = 0.55$) from pre- to post-video questions. Furthermore, repeated-measures ANCOVA models indicated significant Time (pre vs. post)-by-Trait-Affiliation interaction effects for affiliative ($F_{(1, 191)} = 4.9$, $p = .027$, *partial* $\eta^2 = .03$) and positive emotions ($F_{(1, 192)} = 4.8$, $p = .030$, *partial* $\eta^2 = .02$).

Zero-order correlations among personality and emotion variables are presented in Table 4.10. Trait Affiliation was correlated with pre- and post- levels for both affiliative emotions and positive emotions, but relations were significantly stronger for affiliative emotions ($p < .001$). Across variables, the correlation magnitudes were greater for scores on the Trait Affiliation Scale, compared to BFAS aspect and domain scores. Further, magnitudes of correlations were greater for Compassion and Enthusiasm on affiliative emotion endorsement, compared to Politeness or Assertiveness.

Table 4.10. *Correlations among Study 4f individual difference measures (personality and state items)*

	1	2	3	4	5	6	7	8	9	10	
1. Pre-Affiliation	–										
2. Pre-Positive	.59**	–									
3. Post-Affiliation	.66**	.34**	–								
4. Post-Positive	.50**	.58**	.75**	–							
5. Trait Affiliation	.56**	.24**	.63**	.40**	–						
6. BFAS A	.29**	.00	.36**	.13	.59**	–					
7. BFAS E	.46**	.40**	.35**	.33**	.59**	.11	–				
8. Compassion	.45**	.10	.50**	.26**	.72**	.87**	.33**	–			
9. Politeness	.03	-.11	.10	-.07	.25**	.82**	-.20**	.44**	–		
10. Enthusiasm	.51**	.34**	.43**	.30**	.74**	.37**	.82**	.49**	.10	–	
11. Assertiveness	.27**	.33**	.17*	.25**	.28**	-.15*	.86**	.09	-.39**	.40**	–

Note. *p < .05, **p < 0.01

Partial correlations were calculated to examine the relations of personality and scores on post-emotion variables, controlling for pre-values on the corresponding variables (Table 4.11). Across variables, associations were strongest for scores on the Trait Affiliation Scale, suggesting greater responses to affiliative stimuli (in terms of effects on both positive and affiliative emotions) for those with higher Trait Affiliation. Across personality variables, Enthusiasm and Compassion were stronger predictors of affiliative response compared to Politeness, Assertiveness, or Extraversion/Agreeableness domain scores. Finally, the partial correlation analyses controlling for post-levels on the opposing emotion category suggested specificity to the relation between Trait Affiliation and change in state Affiliation, rather than being a general effect on positive emotion.

Table 4.11. *Partial correlations among Study 4f personality and post-emotion controlling for pre-emotion*

	Affiliation	Positive Emotion	Affiliation (Controlling for Post Positive Emotion)	Positive Emotion (Controlling for Post Affiliation)
Trait Affiliation	.42**	.33**	.42**	-.16*
BFAS A	.24**	.16*	.33**	-.24**
BFAS E	.08	.13	-.01	.11
Compassion	.30**	.25**	.35**	-.19**
Politeness	.10	.00	.22**	-.22**
Enthusiasm	.16*	.14	.15*	-.04
Assertiveness	-.01	.07	-.13	.19**

Note. *p < .05, **p < 0.01

As a last step in my analyses, regression models were used to assess whether scores in the Trait Affiliation Scale had unique utility over Enthusiasm and Compassion. Results are presented in Table 4.12. Trait Affiliation predicted pre- and post- state Affiliation, even when controlling for relevant BFAS aspects, age, and gender. More rigorous models (controlling for pre-Affiliation and post-positive emotion) show the specificity of effects of Trait Affiliation on affiliative emotion response, compared to change in positive emotion (controlling for post- affiliative emotions).

Table 4.12. *Study 4f multiple regression analyses of personality and all emotion variables*

Criterion Variable	β_{aff-1}	p	β_{aff-2}	p	β_{aff-3}	p
Pre-Affiliation	.23	.011	.26	.005	.29	.002
Pre-Pos Emotion	-.02	.828	.06	.593	.09	.396
Post-Affiliation	.48	< .001	.48	< .001	.51	< .001
Post-Pos Emotion	.20	.046	.23	.029	.28	.009
Post-Affiliation (Controlling Pre)	.38	< .001	.35	< .001	.37	< .001
Post-Pos Emotion (Controlling Pre)	.22	.015	.20	.028	.23	.012
Post-Affiliation (Controlling Pre and Pos Emo)	.32	< .001	.28	< .001	.28	< .001
Post-Pos Emotion (Controlling Pre and Affil)	-.14	.048	-.13	.063	-.11	.117

Note. β_{aff-1} = Standardized regression coefficient of Trait Affiliation predicting the given criterion variable controlling for Compassion and Enthusiasm, β_{aff-2} = Standardized regression coefficient of Trait Affiliation predicting the given criterion variable controlling for Compassion, Enthusiasm, age, and gender, β_{aff-3} = Standardized regression coefficient of Trait Affiliation predicting the given criterion variable controlling for Compassion, Enthusiasm, age, gender, Politeness, and Assertiveness.

Overall Discussion

Across multiple studies and samples, I have shown that the new Trait Affiliation Scale is reliable, shows convergent and discriminant validity, and has unique utility for predicting important social outcome variables—above and beyond variance explained by

Enthusiasm and Compassion alone. The scale fits well within the broader factor space and nomological network of systems such as the Five Factor Model and IPC, while also offering unique measurement capabilities and incremental validity. The replication of results across multiple samples that differed substantially in their recruitment approaches and demographic makeup suggests findings are likely robust. This project replicates previous research advancing interpersonal theory through the integration of Big Five and IPC factor spaces (DeYoung et al., 2013), while also introducing a new Trait Affiliation Scale that can be usefully incorporated into future research.

Synthesizing Findings with Previous Frameworks

First, results using the items in the Trait Affiliation Scale replicated previous findings integrating measures of the Big Five, IPC, and Trait Affiliation into a single factor space. Specifically, Trait Affiliation (22.5°) fell equidistant between the Big Five aspects of Enthusiasm (45°) and Compassion (0°), which in turn corresponded to markers of the IPC's Love axis and the Gregarious-Extraverted octants, respectively. This means that Trait Affiliation also fell equidistant between Extraversion (67.5°) and Agreeableness (337.5°), as well as equidistant between Assertiveness or the Ambitious-Dominant octant (90°) and Politeness or the Unassuming-Ingenuous octant (315°). In past work, measures related to Trait Affiliation—Warmth facets from the NEO PI-R and AB5C-IPIP, the MPQ social closeness scale, and a single item measure of how "affectionate" participants were—also fell near the hypothesized 22.5° angle, with some of these measures falling nearer 0º (e.g., AB5C-IPIP Warmth), some nearer 45º (e.g., NEO PI-R Warmth and MPQ Social Closeness), and the single-item affectionate variable near 22.5º but closer to the origin, presumably due to comparatively lower reliability (DeYoung et al., 2013). I have now more precisely captured the 22.5° position of Trait Affiliation using a fuller measure and provided evidence of construct validity for the scale in multiple samples. The angular position of Trait Affiliation at 22.5° rather than immediately along the Love axis at (0°) is a modification of standard IPC conceptualizations of Affiliation but is consistent with my conceptualization of this trait as a blend of Enthusiasm and Compassion.

The blended, interstitial content of Trait Affiliation (which leads this trait to fall slightly above the IPC love axis) is of both theoretical and methodological consequence. For one, the fact that Trait Affiliation is not identical to markers of the IPC love axis supports the importance of reward processes (associated primarily with Extraversion rather than Agreeableness) in Trait Affiliation; individual differences in Trait Affiliation are likely related to underlying mechanisms and real-world outcomes other than those of Agreeableness, Compassion, or Empathy alone (Depue & Morrone- Strupinsky, 2005). Indeed, scores on the new Trait Affiliation Scale not only predicted metrics such as social goals and behaviors, social network size, and social cognitive ability, they also had incremental validity over variance explained by just Compassion and Enthusiasm.

The evidence for incremental validity over Compassion and Enthusiasm might lead readers to wonder what exact variance the scale is capturing, since Trait Affiliation is theoretically conceptualized as the intersection of Compassion and Enthusiasm rather than as a construct fully distinct from either aspect. Though one could speculate about underlying psychological or neurobiological mechanisms that are more specifically related to Affiliation and associated real-world behaviors than to either of its associated Big Five aspects—which could potentially drive the observed incremental associations with interpersonal outcomes—an equally likely possibility is that the new scale simply has greater measurement precision and ability to capture variance specific to Trait Affiliation, since Affiliation can only be roughly approximated by averaging scores on measures of Compassion and Enthusiasm (or modeling their shared variance). By using ten items that specifically target this intersection of Compassion and Enthusiasm rather than examining the average of (or shared variance in) items that measure the two aspects more broadly, I was able to pinpoint variance particularly relevant to interpersonal outcomes such as the size of one's social network or quantity of social goals and behaviors.

As a field, personality psychology has begun to move from a focus on descriptive models to explanatory theories. With this shift, it has become increasingly important to be able to reconcile and integrate various theoretical frameworks, which can sometimes appear to conflict with one another. To fully understand and explain complex traits, their interrelations, and mechanisms, relying exclusively on hierarchical models that depict simple structure (e.g., the Big Five) is sometimes inadequate. In the case of interpersonal traits, the circumplex nature of Extraversion, Agreeableness, their aspects, and their intersection (i.e., Trait Affiliation) must be considered. Creating a unified structural model of the IPC axes and octants, Extraversion, Agreeableness, and Trait Affiliation—a model developed in previous research and forwarded in the present work—allows us to integrate the most broadly used personality trait taxonomy with the most broadly used model for understanding individual differences in social behavior. I believe this integration, and in particular the development of the new Trait Affiliation Scale, can facilitate future development of integrative frameworks and theories to explain the functions and mechanisms of—as well as individual differences in—social behavior.

Facilitating Future Affiliation Research

While the current project introduces a new Trait Affiliation Scale and improves the psychometric characterization of this important trait, it also opens possibilities for future work on Trait Affiliation, which could be approached from a variety of methodological and theoretical perspectives. For instance, the Trait Affiliation Scale could be usefully incorporated into existing programs of research on topics such as psychopathology. As previously discussed, social dysfunction is seen across a broad array of mental disorders and symptom dimensions; incorporating the new scale into work on personality disorders, depression, and psychosis could help researchers better understand where exactly pathological low Affiliation falls in the broader factor space of traits such as Antagonism and Detachment, while potentially facilitating the development of screening and intervention programs that target Affiliation specifically. Future studies should also administer the Trait Affiliation Scale along with existing measures of social anhedonia and interpersonal pleasure to better clarify the relations among—and potential distinct components of—these constructs (Chapman et al., 1976; Gooding & Pflum, 2013). In both clinical and general population samples, the scale could also be usefully incorporated into studies using methods already prominent among interpersonal theory researchers, including dyadic designs and experience sampling methods (Beeney et al., 2019; Edershile & Wright, 2020; 2021; Elsaadawy et al., 2020; Human & Biesanz, 2011; 2012; Kerr et al., 2020; Ringwald & Wright, 2021; Ringwald et al., 2020; Vize et al., 2021). The scale could be particularly useful in helping elucidate the biological mechanisms of Trait Affiliation.

Previous theory and empirical work suggest that Agreeableness and Extraversion are related to individual differences in the function of distinct brain systems. For instance, Agreeableness has been linked to structure and function of the default and salience networks (Allen et al., 2017; Chapter 2 of this dissertation; Hou et al., 2017; Takeuchi et al., 2014), whereas Extraversion has been linked to the valuation network (including nucleus accumbens and ventromedial PFC) and the dopaminergic reward system (Civai et al., 2016; Depue & Fu, 2013; Kujawa et al., 2015; Smillie et al., 2011). Since Trait Affiliation also appears to have distinct neural substrates worth considering—in addition to the broader systems underlying Extraversion and Agreeableness—the new scale could be particularly useful for future studies that attempt to differentiate between the neural correlates of these three interrelated traits. Specific brain systems that warrant continued investigation in relation to individual differences in Affiliation include striatal reward circuits and associated endogenous psychoactive substances such as mu opioids, as well as hormones like oxytocin (Argiolas & Gessa, 1991; Depue & Morrone-Strupinsky, 2005; Marsh et al., 2010).

Differentiating the role of dopamine vs. endogenous opioids may be a particularly fruitful route for future work on Trait Affiliation. Whereas dopamine produces a desire to pursue rewards ("wanting"; Berridge & Robison, 2009), the pleasure one experiences from receiving rewards ("liking"), including social rewards, involves the endogenous opioid system (Depue & Morrone-Strupinsky, 2005). On this basis, Enthusiasm and Affiliation have been hypothesized to reflect sensitivity to hedonic reward and pleasure associated with the function of endogenous opioids, whereas Assertiveness appears to reflect the sensitivity to incentive reward and drive associated with dopaminergic function (Depue & Collins, 1999; DeYoung, 2013; DeYoung & Weisberg, 2018; Wacker et al., 2012). The endogenous opioid system appears to be a key mechanism mediating the positive emotions that follow attainment and consumption of rewards and is particularly important in social bonding (Depue & Morrone-Strupinsky, 2005). Specific evidence for the role of the opioid system in Affiliation is beginning to emerge; for example, sensitivity to psychopharmacological or tactile manipulation of endogenous opioid activity appears to be moderated by measures of social closeness, specific types of tactile stimulation and social touch lead to an increase of endogenous opioid activity, and opioid receptor availability is related to scores on Affiliation-related questionnaires (Depue & Morrone-Strupinsky, 2005; Karjalainen et al., 2016; Nummenmaa et al., 2015; 2016). Future research should work to further investigate the neural basis of individual differences in Trait Affiliation and the motivational aspects of social behavior. One worthwhile route might be to examine how Affiliation and its pathological counterparts (e.g., social anhedonia and intimacy avoidance) might be associated with connectivity and coordination between the neural systems associated with social cognitive ability (e.g., the default and salience networks) and those associated with reward (e.g., the ventromedial PFC and striatum).

Chapter 4 Conclusion

Trait Affiliation is a consequential, interstitial trait that represents individual differences in the desire to form and maintain relationships and falls at the intersection of the Compassion aspect of Agreeableness and the Enthusiasm aspect of Extraversion. The current research introduces a new Trait Affiliation Scale that shows high reliability, construct validity, and incremental validity in predicting relevant outcome variables such as social cognitive ability, social network size, and social behavior—above and beyond variance explained by measures of Compassion and Enthusiasm alone. The Trait Affiliation Scale shows promise for use in future research on individual differences in social behavior and may be particularly useful for forwarding research on the neurobiology of Affiliation and on transdiagnostic patterns of social dysfunction seen across a variety of mental disorders.

CONCLUSION

As social mammals, humans are predisposed toward a desire to relate with one another and with a curiosity to understand our often-complex relationships. The fields of social cognition, social neuroscience, and psychopathology research have produced hundreds of studies attempting to explain individual differences in social abilities and processes—endeavors that have been marked by plenty of breakthroughs but also persistent problems. Particularly, much of the research has been done in small samples that inhibit the ability to produce reliable and replicable results. Further, there is a dearth of approaches that emphasize broad explanatory frameworks to integrate findings across neurobiological systems and traits of interest. Finally, much of the current relevant work occurs in isolated subfields with little to no interdisciplinary crosstalk. Through this dissertation, I hope to have shown that personality psychology provides a promising scaffold for integrating past empirical research and creating new frameworks that allow us to better understand variation in social processes and corresponding outcomes.

In the preceding four chapters, I have presented novel empirical findings that improve the measurement and characterization of Agreeableness, Extraversion, and Affiliation—three personality traits key to individual differences in social cognition and behavior. Through a broad array of methodological approaches, including fMRI, psychometric methods, and behavioral tasks, I have further elucidated the underlying psychological and neural mechanisms of these traits. Collectively, this empirical work was drawn from eight independent samples and nearly 30,000 individual participants. In addition to offering methodological and theoretical contributions specific to interpersonal functioning, my research adds to a growing body of work using personality as a framework for understanding broader individual differences traditionally studied in subfields outside of the personality literature.

Clarifying the Mechanisms of Social Cognition and Behavior

Collectively, the research presented in my dissertation clarifies our understanding of several important psychological processes and neural systems that seem to underlie individual differences in social cognition and behavior. My first two chapters replicate and extend past work, providing further evidence that Agreeableness is associated with individual differences in default network function and social cognitive abilities such as theory of mind. Likewise, my third chapter provides further evidence of reward sensitivity as a likely mechanism for individual differences in Extraversion. Taken with the work of researchers such as Depue, Collins, Morrone-Strupinsky, and Nummenmaa, my fourth chapter suggests that Trait Affiliation may have both shared and unique measurement properties and underlying mechanisms, when compared to the individual Big Five domains of Agreeableness and Extraversion. Despite the contributions of my dissertation to the understanding of interpersonal functioning, there is still plenty of work to be done in this area.

For instance, future research could more specifically investigate connectivity and coordination among brain systems such as the default network, salience network, and reward circuits and the role their interaction might play in Agreeableness-Antagonism, Extraversion-Detachment, and Trait Affiliation. Moreover, compared to the aforementioned traits, the specific neural correlates of Honesty-Manipulativeness, Pacifism-Aggression, and Assertiveness remain comparatively understudied and elusive. There is a lack of high-powered research on the neurobiology of Honesty, but regions of both the salience and frontoparietal networks may be implicated (ten Brinke et al., 2005). Another potentially relevant brain system for social functioning is testosterone, which may help to explain the negative covariation between Politeness and Assertiveness (DeYoung & Blain, 2020; DeYoung & Weisberg, 2018). Several studies have connected individual differences in testosterone function to Extraversion, Assertiveness, and dominance as well as Aggression (Luxen & Buunk, 2005; Montoya et al., 2012; Netter, 2004; Nguyen et al., 2016; Smeets-Janssen et al., 2015; Turan et al., 2014). Further research is needed to better understand the shared and unique neurocognitive mechanisms of the various personality traits and psychopathology symptoms related to social cognition and behavior.

Contributions to Research Methodology and Reproducibility

In addition to replicating previous work and contributing new insights in personality psychology, social neuroscience, and interpersonal theory, this dissertation highlights several vital methodological considerations. First, Chapters 1 and 3 underscore the importance of high-powered replication studies. More specifically, Chapter 1 shows how multi-task designs and latent variable modeling can build on previous work using single-task designs (and in doing so potentially overcome some of the reliability issues often associated with behavioral tasks), whereas Chapter 3 serves as a cautionary tale against overinterpreting results obtained in small samples without the statistical power needed to achieve stable parameter estimates. Chapters 2 and 4—specifically through their use of the HCP, ESCS, and SAPA datasets—speak to the utility of utilizing large public-domain datasets in interpersonal and personality research. Use of such open datasets can increase research transparency and reproducibility, while also combatting many of the problems particular to small, under-powered research designs. These approaches may prove particularly useful in the fields of social, clinical, and personality neuroscience, as small samples are often the rule rather than the exception when it comes to fMRI.

Perhaps social and personality neuroscience's greatest problem when it comes to individual difference analyses has been and still is small sample sizes and resulting effects on statistical power (Barch & Yarkoni, 2013; Yarkoni, 2009; Cacioppo et al., 2014). Due largely to the high monetary cost of fMRI, many studies are conducted and published with sample sizes far too small for high-quality research on individual differences. For instance, analysis of more than 400 structural MRI studies published between 2006 and 2009 found the median statistical power was only 8% (Button et al., 2013; Ioannidis, 2011), and another study found that the median sample size for MRI studies was only 15 individuals (Carp, 2012). Because of low statistical power, by definition, only the largest of effect sizes are possible to detect. Vul et al.'s (2009) now infamous review of the social and personality neuroscience literature found an average correlation in the surveyed individual differences research of around $r = .6$. In contrast, the average effect size in social and personality research has been estimated near $r = .2$ (Gignac & Szodorai, 2016; Richard et al., 2003). Most likely, many findings from social neuroscience research on individual differences are Type I errors, or at the very least, are grossly overestimating effect sizes. It is important to note, however, that chronically low power in the field also undoubtedly contributes to magnified Type II error rates and failure to detect the kinds of effects one would expect in personality research; a sample size of roughly 200 individuals—obviously far greater than the median 15 individuals—would be needed to detect such an effect with 80% power, at a standard Type I error rate (Cohen, 1992).

It is worth noting that, despite these sample size recommendations for individual differences research, much of the social neuroscience research linking social cognitive phenomena to underlying systems—without an interest in interindividual variability—may be valid. The number of participants needed for adequate power to detect between-subjects effects is larger than what is needed for research using within-subjects designs (Thompson & Campbell, 2004; Yarkoni, 2009). The work presented in Chapter 2 of this dissertation is a step in the right direction for social neuroscience, as are initiatives such as the HCP, ABCD Study, and ENGIMA Consortia (Harms et al., 2018; Casey et al., 2018; Thompson et al., 2013; Van Essen et al., 2013). Future work in social neuroscience and related fields will need to overcome these power limitations to produce the consistent, robust findings necessary for developing comprehensive and accurate theories to explain the neurobiology of individual differences in social cognition and behavior.

The Importance of Hierarchy in Personality Research

Whereas a large portion of research attempting to characterize and explain variation in personality has focused on traits at the Big Five domain level, constructs at lower and higher levels of the personality trait hierarchy can be equally as important for understanding underlying sources of variation. Results from Chapters 1 and 4 of the current dissertation underscore the importance of investigating personality at resolutions other than the Big Five. More specifically, Chapter 1 shows that different facets of the Agreeableness-Antagonism dimension show divergent associations with social cognitive abilities, with Honesty-Manipulativeness predicting social cognition in the opposite direction from Compassion-Callousness and Pacifism-Aggression. These findings add to growing body of work documenting divergent associations at the aspect or facet level that would not be captured by examining associations at the domain (Big Five) level alone (e.g., Blain, Grazioplene et al., 2020; Blain, Longenecker et al., 2020; DeYoung et al., 2012; Grazioplene et al., 2016; Hirsh et al., 2010; Hou et al., 2017; Xu et al., 2021). Future research might usefully apply similar frameworks in understanding mechanisms and outcomes associated with traits such as Extraversion and its lower-order components.

Chapter 4 demonstrates the importance of another non-domain component of the personality trait hierarchy—interstitial traits. Although Trait Affiliation is theoretically conceptualized as the intersection of Enthusiasm and Compassion, findings show that a new scale specifically assessing this trait can predict relevant outcomes better than measures of related domains or aspects alone. Measuring interstitial traits with greater precision, as I do with the new Trait Affiliation Scale, may provide a promising avenue for identifying the underlying mechanisms of specific tendencies not captured by just the Big Five and for predicting real-world outcomes associated with these blended traits. One other trait that can be conceptualized as an interstitial blend of Big Five aspects is Enlightened Compassion, which blends high Openness and Compassion (Smillie et al., 2019). Enlightened Compassion might be a useful target for future studies of social cognition and behavior, given evidence that function of the default network is important for both of its constituent traits (Allen et al., 2017; Blain, Grazioplene et al., 2020; Chapter 2 of this dissertation; Takeuchi et al., 2014).

It is important to note that most of the research reviewed here (and that currently exists in the broader personality and social neuroscience literature) has focused on the trait level of analysis, but there are other important components to personality beyond traits. For instance, we should make greater efforts to understand the interaction between persons and situations, and the influence unique trait-context interactions might have on social cognition and behavior. Future work using experience sampling or longitudinal designs could be particularly useful for examining personality beyond the trait level. Finally, social and personality neuroscience would benefit from the further development of ecologically valid tasks and behavioral paradigms that better represent the abilities and skills humans are asked to draw upon during daily life for use in conjunction with laboratory tests of social cognition and related skills.

Concluding Remarks

My dissertation builds on the strengths of existing research on social cognition and social behavior from social, clinical, and neuroscientific perspectives, while leveraging insights from personality psychology to give us a better understanding of variation in social abilities throughout the population and across levels of functioning.

Other concepts highlighted throughout my work include the importance of high-quality replication-extension studies, how to leverage large public datasets, and why research on traits other than the Big Five is essential. In completing this dissertation, I hope to have made an argument for the utility of personality psychology and to have paved the way to a better understanding of how and why individuals vary in our social abilities, interpersonal interactions, and relationship success.

As mentioned at the outset of my dissertation, I believe the models and methods of personality psychology can be usefully adopted into existing programs of research on virtually any topic of psychological inquiry that involves individual differences. There is a rich history of this personality-focused approach in psychopathology research over the past few decades (Krueger et al., 1996; 2007; Markon et al., 2005; Ringwald et al., 2021; Thomas et al., 2013; Wright & Sims, 2014), and ongoing efforts such as HiTOP (Kotov et al., 2010) and RDoC (Insel et al., 2010) represent ever increasing interest in dimensional systems for understanding mental illness. Personality frameworks have also been usefully applied to a variety of other research topics, from psychosis proneness and pattern detection to emotional processing and self-regulation (Aaron et al., 2018; 2020; Blain, Longenecker, et al., 2020; Blain, Grazioplene, et al., 2020; DeYoung & Rueter, 2016; Rueter et al., 2018; Wang et al., 2020; 2021). Moving forward, researchers should more intentionally develop methods and models that connect the interrelated components of 1) brain function, 2) abilities, and 3) traits along the personality-psychopathology continuum, to more fully understand and explain individual differences at multiple levels of the trait hierarchy.

References

Aaron, R. V., Blain, S. D., Snodgress, M. A., & Park, S. (2020). Quadratic relationship between alexithymia and interoceptive accuracy, and results from a pilot mindfulness intervention. *Frontiers in psychiatry, 11,* 132.

Aaron, R. V., Snodgress, M. A., Blain, S. D., & Park, S. (2018). Affect labeling and other aspects of emotional experiences in relation to alexithymia following standardized emotion inductions. *Psychiatry research, 262,* 115-123.

Abell, F., Happe, F., & Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development, 15,* 1–16.

Abram, S. V., Wisner, K. M., Fox, J. M., Barch, D. M., Wang, L., Csernansky, J. G., MacDonald III, A. W., & Smith, M. J. (2016). Fronto-temporal connectivity predicts cognitive empathy deficits and experiential negative symptoms in schizophrenia. *Human Brain Mapping, 38*(3), 1111–1124. https://doi.org/10.1002/hbm.23439

Adolphs, R. (2009). The social brain: Neural basis of social knowledge. *Annual Review of Psychology, 60,* 693–716.

Aldridge, V. K., Dovey, T. M., & Wade, A. (2017). Assessing test-retest reliability of psychological measures. *European Psychologist, 22.* https://doi.org/10.1027/1016-9040/a000298

Ali, F., & Chamorro-Premuzic, T. (2010). The dark side of love and life satisfaction: Associations with intimate relationships, psychopathy and Machiavellianism. *Personality and Individual Differences, 48*(2), 228-233.

Allen, T. A., & DeYoung, C. G. (2017). Personality neuroscience and the Five-Factor Model. In T. A. Widiger (Ed.) *Oxford Handbook of the Five Factor Model of Personality* (p. 319–349). Oxford University Press.

Allen, T. A., Rueter, A. R., Abram, S. V., Brown, J. S., & DeYoung, C. G. (2017). Personality and neural correlates of mentalizing ability. *European Journal of Personality, 31,* 599–613.

Allison, P. D. (1990). Change scores as dependent variables in regression analysis. *Sociological Methodology, 20,* 93–114.

Altschuler, M., Sideridis, G., Kala, S., Warshawsky, M., Gilbert, R., Carroll, D., Burger-Caplan, R., & Faja, S. (2018). Measuring individual differences in cognitive, affective, and spontaneous theory of mind among school-aged children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 48,* 3945–3957.

Amelang, M., & Steinmayr, R. (2006). Is there a validity increment for tests of emotional intelligence in explaining the variance of performance criteria? *Intelligence, 34*(5), 459-468.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (5th ed.).* Washington, DC: American Psychiatric Association.

Anderson, K. M., Ge, T., Kong, R., Patrick, L. M., Spreng, R. N., Sabuncu, M. R., Yeo, B. T. T., & Holmes, A. J. (2021). Heritability of individualized cortical network topography. *PNAS*, *118*(9), e2016271118. https://doi.org/10.1073/pnas.2016271118

Anderson, J. L., Sellbom, M., & Salekin, R. T. (2018). Utility of the Personality Inventory for DSM-5-Brief Form (PID-5-BF) in the measurement of maladaptive personality and psychopathology. *Assessment, 25*(5)*,* 596–607.

Andreasen, N. C., Berrios, G. E., Bogerts, B., Brenner, H. D., Carpenter, W. T., Crow, T. J., ... & Lewine, R. R. J. (2012). *Negative Versus Positive Schizophrenia.* Springer Science & Business Media.

Andrews-Hanna, J. R., Smallwood, J., & Spreng, R. N. (2014). The default network and self-generated thought: Component processes, dynamic control, and clinical relevance. *Annals of the New York Academy of Sciences, 1316*(1)*,* 29–52.

Ang, Y. S., Kaiser, R., Deckersbach, T., Almeida, J., Phillips, M. L., Chase, H. W., ... & Weissman, M. (2020). Pretreatment reward sensitivity and frontostriatal resting-state functional connectivity are associated with response to bupropion after sertraline non-response. *Biological Psychiatry.* https://doi.org/10.1016/j.biopsych.2020.04.009

Anglim, J., & O'connor, P. (2019). Measurement and research using the Big Five, HEXACO, and narrow traits: A primer for researchers and practitioners. *Australian Journal of Psychology, 71*(1), 16-25.

Antonakis, J. (2004). On why" emotional intelligence" will not predict leadership effectiveness beyond IQ or the" big five": An extension and rejoinder. *Organizational Analysis, 12*(2).

Apperly, I. A. (2012). What is "theory of mind"? Concepts, cognitive processes and individual differences. *The Quarterly Journal of Experimental Psychology, 65,* 825–839.

Arbula, S., Pisanu, E., & Rumiati, R. I. (2021). Representation of social content in dorsomedial prefrontal cortex underlies individual differences in agreeableness trait. *NeuroImage*, *235*, 118049. https://doi.org/10.1016/j.neuroimage.2021.118049

Ascher, J. A., Cole, J. O., Colin, J. N., Feighner, J. P., Ferris, R. M., Fibiger, H. C., ... & Richelson, E. (1995). Bupropion: a review of its mechanism of antidepressant activity. *The Journal of Clinical Psychiatry*, *56*(9), 395–401.

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and social psychology review, 11*(2), 150-166.

Ashton, M. C., Lee, K., & Goldberg, L. R. (2007). The IPIP–HEXACO scales: An alternative, public-domain measure of the personality constructs in the HEXACO model. *Personality and Individual Differences, 42*(8), 1515-1526.

Ashton, M. C., Lee, K., & De Vries, R. E. (2014). The HEXACO honesty-humility, agreeableness, and emotionality factors: A review of research and theory. *Personality and Social Psychology Review, 18*(2), 139-152.

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural equation modeling: a multidisciplinary journal, 16*(3), 397-438.

Ávila, C., & Parcet, M. A. (2002). Individual differences in reward sensitivity and attentional focus. *Personality and Individual Differences, 33, 979*-996. https://doi.org/10.1016/S0191-8869(01)00207-0

Ax, A. F. (1953). The physiological differentiation between fear and anger in humans. *Psychosomatic medicine, 15*(5), 433-442.

Bainbridge, T. F., Quinlan, J. A., Mar, R. A., Smillie, L. D., & Fajkowska, M. (2019). Openness/intellect and susceptibility to pseudo–profound bullshit: A replication and extension. *European Journal of Personality, 33*(1), 72–88.

Baker, J. (2003). *Social Skills Training: For Children and Adolescents with Asperger Syndrome and Social-Communication Problems*. Mission, KS: Autism Asperger Publishing Company.

Barch, D. M., Burgess, G. C., Harms, M. P., Petersen, S. E., Schlaggar, B. L., Corbetta, M., Glasser, M. F., Curtiss, S., Dixit, S., Feldt, C., Nolan, D., Bryant, E., Hartley, T., Footer, O., Bjork, J. M., Poldrack, R., Smith, S., Johansen-Berg, H., Snyder, A. Z., & Van Essen, D. C. (2013). Function in the human connectome: task-fMRI and individual differences in behavior. *NeuroImage, 80,* 169–189.

Barch, D. M., Carter, C. S., Gold, J. M., Johnson, S. L., Kring, A. M., MacDonald III, A. W., ... & Strauss, M. E. (2017). Explicit and implicit reinforcement learning across the psychosis spectrum. *Journal of Abnormal Psychology, 126*(5), 694.

Barch, D. M., Oquendo, M. A., Pacheco, J., & Morris, S. (2016). *Behavioral assessment methods for RDoC constructs: A report by the National Advisory Mental Health Council Workgroup on tasks and measures for RDoC.* Washington, DC: National Institutes of Mental Health.

Barch, D. M., & Yarkoni, T. (2013). Introduction to the special issue on reliability and replication in cognitive and affective neuroscience research. *Cognitive, Affective, and Behavior Neuroscience, 13,* 687-689.

Barends, A. J., de Vries, R. E., & van Vugt, M. (2019). Power influences the expression of Honesty-Humility: The power-exploitation affordances hypothesis. *Journal of Research in Personality, 82,* 103856.

Barford, K. A., Zhao, K., & Smillie, L. D. (2015). Mapping the interpersonal domain: Translating between the Big Five, HEXACO, and Interpersonal Circumplex. *Personality and Individual Differences, 86,* 232-237.

Bar-On, R. (2001). Emotional intelligence and self-actualization. *Emotional Intelligence in Everyday Life: A Scientific Inquiry,* 82-97.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21,* 37–46.

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.

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science, 20*(5)*,* 286–290.

Beaty, R. E., Chen, Q., Christensen, A. P., Qiu, J., Silvia, P. J., & Schacter, D. L. (2018). Brain networks of the imaginative mind: Dynamic functional connectivity of default and cognitive control networks relates to Openness to Experience. *Human Brain Mapping, 39*(2), 811–821.

Beaty, R. E., Kaufman, S. B., Benedek, M., Jung, R. E., Kenett, Y. N., Jauk, E., Neubauer, A. C., & Silvia, P. J. (2016). Personality and complex brain networks: The role of Openness to experience in default network efficiency. *Human Brain Mapping, 37*, 773–779.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *BDI-II: Beck Depression Inventory Manual*. (2nd ed.) TX, San Antonio: The Psychological Corporation.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives General Psychiatry, 4,* 561-571. http://dx.doi.org/10.1001/archpsyc.1961.01710120031004

Beeney, J. E., Stepp, S. D., Hallquist, M. N., Ringwald, W. R., Wright, A. G., Lazarus, S. A., ... & Pilkonis, P. A. (2019). Attachment styles, social behavior, and personality functioning in romantic relationships. *Personality Disorders: Theory, Research, and Treatment, 10*(3), 275.

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

Best, C. S., Moffat, V. J., Power, M. J., Owens, D. G. C., & Johnstone, E. C. (2008). The boundaries of the cognitive phenotype of autism: Theory of mind, central coherence and ambiguous figure perception in young people with autistic traits. *Journal of Autism and Developmental Disorders*, *38*(5), 840–847. https://doi.org/10.1007/s10803-007-0451-8

Blain, S. D., Grazioplene, R. G., Ma, Y., & DeYoung, C. G. (2020). Toward a neural model of the Openness-Psychoticism dimension: Functional connectivity in the default and frontoparietal control networks. *Schizophrenia Bulletin, 46*(3), 540–551.

Blain, S. D., Longenecker, J. M., Grazioplene, R. G., Klimes-Dougan, B., & DeYoung, C. G. (2020). Apophenia as the disposition to false positives: A unifying framework for openness and psychoticism. *Journal of Abnormal Psychology, 129*(3), 279.

Blain, S. D., Peterman, J. S., & Park, S. (2017). Subtle cues missed: Impaired perception of emotion from gait in relation to schizotypy and autism spectrum traits. *Schizophrenia Research, 183,* 157–160.

Blain, S. D., Sassenberg, T. A., Xi, M., Zhao, D., & DeYoung, C. G. (2020). Extraversion but not depression predicts reward sensitivity: Revisiting the measurement of anhedonic phenotypes. *Journal of Personality and Social Psychology.* https://doi.org/10.1037/pspp0000371

Bogdan, R., & Pizzagalli, D. A. (2006). Acute stress reduces hedonic capacity: Implications for depression. *Biological Psychiatry, 60,* 1147-1154.

Bogdan, R., Santesso, D. L., Fagerness, J., Perlis, R. H., & Pizzagalli, D. A. (2011). Corticotropin-releasing hormone receptor type 1 (CRHR1) genetic variation and stress interact to influence reward learning. *Journal of Neuroscience, 31*(37), 13246-13254.

Bora, E. (2020). Theory of mind and schizotypy: A meta-analysis. *Schizophrenia Research*, *222*, 97–103. https://doi.org/10.1016/j.schres.2020.04.024

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

Bower, G. H. (1983). Affect and cognition. *Philosophical Transactions of the Royal Society of London. B, 302*(1110), 387-402.

Bowman, L. C., Thorpe, S. G., Cannon, E. N., & Fox, N. A. (2017). Action mechanisms for social cognition: Behavioral and neural correlates of developing theory of mind. *Developmental Science, 20*(5)*,* e12447.

Bress, J. N., & Hajcak, G. (2013). Self-report and behavioral measures of reward sensitivity predict the feedback negativity. *Psychophysiology, 50*(7), 610-616.

Brothers, L. (1990). The social brain: A project for integrating primate behavior and neurophysiology in a new domain. *Concepts in Neuroscience, 1,* 27–51.

Brown, G. S., & White, K. G. (2005). The optimal correction for estimating extreme discriminability. *Behavior Research Methods, 37*(3), 436-449.

Brown, G. S., & White, K. G. (2009). Measuring discriminability when there are multiple sources of bias. *Behavior Research Methods, 41*(1), 75-84.

Brune, M., & Brune-Cohrs, U. (2006). Theory of mind—evolution, ontogeny, brain mechanisms and psychopathology. *Neuroscience & Biobehavioral Reviews, 30*(4)*,* 437–455.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences, 1124,* 1–38.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafo, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews | Neuroscience, 14,* 1–14.

Byrne, R. W. (1996). Machiavellian intelligence. *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews, 5*(5), 172-180.

Byrne, R., & Whiten, A. (1994). *Machiavellian Intelligence.* Oxford: Oxford University Press.

Cacioppo, J. T., Cacioppo, S., Dulawa, S., & Palmer, A. A. (2014). Social neuroscience and its potential contribution to psychiatry. *World Psychiatry, 13*(2), 131-139.

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

Caprara, G. V., Alessandri, G., Giunta, L. D., Panerai, L., & Eisenberg, N. (2010). The contribution of agreeableness and self-efficacy beliefs to prosociality. *European Journal of Personality*, *24*(1), 36–55. https://doi.org/10.1002/per.739

Carp, J. (2012). The secret lives of experiments: Methods reporting in the fMRI literature. *NeuroImage, 63,* 289–300.

Carrington, S. J., & Bailey, A. J. (2009). Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Human Brain Mapping*, *30*(8), 2313-2335.

Casey, B. J., Cannonier, T., Conley, M. I., Cohen, A. O., Barch, D. M., Heitzeg, M. M., ... & Dale, A. M. (2018). The adolescent brain cognitive development (ABCD) study: imaging acquisition across 21 sites. *Developmental cognitive neuroscience, 32,* 43-54.

Cassidy, K. W., Werner, R. S., Rourke, M., Zubernis, L. S., & Balaraman, G. (2003). The relationship between psychological understanding and positive social behaviors. *Social Development, 12*(2), 198–221.

Castelli, F., Happe, F., Frith, U., & Frith, C. (2000). Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage, 12,* 314–325.

Chapman, L. J., Chapman, J. P., & Raulin, M. L. (1976). Scales for physical and social anhedonia. *Journal of abnormal psychology, 85*(4), 374.

Chong, M., Bhushan, C., Joshi, A. A., Choi, S., Haldar, J. P., Shattuck, D. W., Spreng, R. N., & Leahy, R. M. (2017). Individual parcellation of resting fMRI with a group functional connectivity prior. *NeuroImage, 156,* 87–100. https://doi.org/10.1016/j.neuroimage.2017.04.054

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment, 6*(4), 284.

Cicero, D. C., Jonas, K. G., Li, K., Perlman, G., & Kotov, R. (2019). Common taxonomy of traits and symptoms: linking schizophrenia symptoms, schizotypy, and normal personality. *Schizophrenia Bulletin, 45*(6), 1336-1348.

Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology, 100*(3), 316.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7,* 249-253. http://dx.doi.org/10.1177/014662168300700301

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science, 1*(3), 98-101.

Collison, K. L., Vize, C. E., Miller, J. D., & Lynam, D. R. (2018). Development and preliminary validation of a five factor model measure of Machiavellianism. *Psychological Assessment, 30*(10), 1401.

Condon, D. M., & Revelle, W. (2014). The International Cognitive Ability Resource: Development and initial validation of a public-domain measure. *Intelligence, 43,* 52–64.

Condon, D., & Revelle, W. (2015). Selected personality data from the SAPA-Project: On the structure of phrased self-report items. *Journal of Open Psychology Data, 3*(1).

Condon, D. M., Roney, E., & Revelle, W. (2017). A SAPA project update: On the structure of phrased self-report personality items. *Journal of Open Psychology Data, 5*(1).

Corr, P. J. (2008). *The Reinforcement Sensitivity Theory of Personality*. Cambridge University Press.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI) Professional Manual.* Odessa, FL: Psychological Assessment Resources.

Costa Jr, P. T., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: robust and surprising findings. *Journal of personality and social psychology, 81*(2), 322.

Costello, C. G. (1972). Depression: Loss of reinforcers or loss of reinforcer effectiveness? *Behavior Therapy, 3,* 240-247. https://doi.org/10.1016/S0005-7894(72)80084-4

Côté, S., & Moskowitz, D. S. (1998). On the dynamic covariation between interpersonal behavior and affect: Prediction from Neuroticism, Extraversion, and Agreeableness. *Journal of Personality and Social Psychology*, *75*(4), 1032–1046.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297–334.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American psychologist, 12*(11), 671.

Curtis, R., Kenworthy, D., Bevan, T., Fellner, E., Rickman, A., Nighy, B., Firth, C., ... (2004). *Love Actually.* London, UK, DNA Films.

Dahlgren, S., Dahlgren Sandberg, A., & Hjelmquist, E. (2010). The non-specificity of theory of mind deficits: Evidence from children with communicative disabilities. *European Journal of Cognitive Psychology, 15*(1), 129–155.

Dale, A. M., Fischl, B., Sereno, M. I., (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage, 9*(2), 179–194. https://doi.org/10.1006/nimg.1998.0395

Dang, J., King, K. M., & Inzlicht, M. (2020). Why Are Self-Report and Behavioral Measures Weakly Correlated? *Trends in Cognitive Sciences, 24*(4), 267-269.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology, 10,* 85.

De Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory.* Guilford Press.

De Vane, C. L., Liston, H. L., & Markowitz, J. S. (2002). Clinical pharmacokinetics of sertraline. *Clinical Pharmacokinetics, 41*(15), 1247-1266.

del Barrio, V., Aluja, A., & García, L. F. (2004). Relationship between empathy and the Big Five personality traits in a sample of Spanish adolescents. *Social Behavior and Personality: An International Journal, 32*(7), 677–681.

Delgado, M. R., Glimcher, P. W., Hajcak, G., Pizzagalli, D. A., Treadway, M. T., Yerys, B. E. (2016). Positive Valence Systems Final Report. In Barch, D. M., Oquendo, M. A., Pacheco, J., & Morris, S. *Behavioral Assessment Methods for RDoC Constructs: A Report by the National Advisory Mental Health Council Workgroup on Tasks and Measures for RDoC.* Washington, DC: National Institutes of Mental Health.

Depue, R. A., & Collins, P. F. (1999). Neurobiology of the structure of personality: Dopamine, facilitation of incentive motivation, and Extraversion. *Behavioral and Brain Sciences, 22,* 491-517. http://dx.doi.org/10.1017/S0140525X99002046

Depue, R. A., & Morrone-Strupinsky, J. V. (2005). A neurobehavioral model of affiliative bonding: Implications for conceptualizing a human trait of affiliation. *Behavioral and Brain Sciences, 28*(3), 313-349.

Devine, R. T., & Hughes, C. (2013). Silent films and strange stories: Theory of mind, gender, and social experiences in middle childhood. *Child Development, 84*(3), 989–1003.

Devine, R. T., White, N., Ensor, R., & Hughes, C. (2016). Theory of mind in middle childhood: Longitudinal associations with executive function and social competence. *Developmental Psychology, 52*(5), 758–771.

DeYoung, C. G. (2013). The neuromodulator of exploration: A unifying theory of the role of dopamine in personality. *Frontiers in Human Neuroscience, 7,* article 762. http://dx.doi.org/10.3389/fnhum.2013.00762

DeYoung, C. G. (2015). Cybernetic big five theory. *Journal of Research in Personality, 56,* 33–58.

DeYoung, C. G. & Blain, S. D. (2020). Personality neuroscience: Explaining individual differences in motivation, emotion, cognition, and behavior. In Corr, P. J., & Matthews, G. (Eds.) *The Cambridge Handbook of Personality Psychology* (2nd ed.). New York: Cambridge University Press.

DeYoung, C. G., Carey, B. E., Krueger, R. F., & Ross, S. R. (2016). Ten aspects of the Big Five in the Personality Inventory for DSM–5. *Personality Disorders: Theory, Research, and Treatment, 7*(2), 113–123.

DeYoung, C. G., Grazioplene, R. G., & Peterson, J. B. (2012). From madness to genius: The Openness/Intellect trait domain as a paradoxical simplex. *Journal of Research in Personality, 46*, 63–78.

DeYoung, C. G., & Krueger, R. F. (2018). A cybernetic theory of psychopathology. *Psychological Inquiry, 29*(3), 117-138. https://doi.org/10.1080/1047840X.2018.1513680

DeYoung, C. G., Weisberg, Y. J., Quilty, L. C., & Peterson, J. B. (2013). Unifying the aspects of the Big Five, the interpersonal circumplex, and trait affiliation. *Journal of Personality, 81*(5), 465–475.

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology, 93*(5), 880.

DeYoung, C. G., & Weisberg, Y. J. (2018). Cybernetic approaches to personality and social behavior. In M. Snyder & K. Deaux (Eds). *Oxford Handbook of*

Personality and Social Psychology, Second edition (p. 387–414). New York: Oxford University Press.

Ding, X. P., Wellman, H. M., Wang, Y., Fu, G., & Lee, K. (2015). Theory-of-mind training causes honest young children to lie. *Psychological Science, 26*(11), 1812-1821.

Di Nicola, M., De Risio, L., Battaglia, C., Camardese, G., Tedeschi, D., Mazza, M., ... & Siracusano, A. (2013). Reduced hedonic capacity in euthymic bipolar subjects: a trait-like feature? *Journal of Affective Disorders, 147*(1-3), 446-450.

Di Tella, M., Miti, F., Ardito, R. B., & Adenzato, M. (2020). Social cognition and sex: Are men and women really different? *Personality and Individual Differences*, *162*, 110045. https://doi.org/10.1016/j.paid.2020.110045

Dolan, M. & Fullam, R. (2004). Theory of mind and mentalizing ability in antisocial personality disorders with and without psychopathy. *Psychological Medicine, 34,* 1093–1102.

Dworak, E. M., Revelle, W., Doebler, P., & Condon, D. M. (2021). Using the International Cognitive Ability Resource as an open source tool to explore individual differences in cognitive ability. *Personality and Individual Differences, 169,* 109906.

Edershile, E. A., & Wright, A. G. (2020). Fluctuations in grandiose and vulnerable narcissistic states: A momentary perspective. *Journal of Personality and Social Psychology. 120(5),* 1386–1414. https://doi.org/10.1037/pspp0000370

Edershile, E. A., & Wright, A. G. (2021). Grandiose and vulnerable narcissistic states in interpersonal situations. *Self and Identity, 20*(2), 165-181.

Elsaadawy, N., Carlson, E. N., & Human, L. J. (2020). Who influences meta-accuracy? It takes two to know the impressions we make. *Journal of Personality and Social Psychology.* Advance online publication. https://doi.org/10.1037/pspp0000376

Edwards, J. R. (1994). Regression analysis as an alternative to difference scores. *Journal of Management, 20*(3)*,* 683–689.

Edwards, J. R. (1996). An examination of competing versions of the person-environment fit approach to stress. *Academy of Management Journal, 39*(2), 292–339.

Eisenberg, I. W., Bissett, P. G., Enkavi, A. Z., Li, J., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Uncovering the structure of self-regulation through data-driven ontology discovery. *Nature Communications, 10*(1), 1-13.

Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions*. Science, 221*(4616), 1208-1210.

Emanuel, E. J., & Gudbranson, E. (2018). Does medicine overemphasize IQ? *JAMA, 319*(7), 651-652.

Enkavi, A. Z., Eisenberg, I. W., Bissett, P. G., Mazza, G. L., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019a). Large-scale analysis of test–retest reliabilities of self-regulation measures. *Proceedings of the National Academy of Sciences, 116*(12), 5472-5477.

Enkavi, A. Z., Eisenberg, I. W., Bissett, P. G., Mazza, G. L., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019a). Large-scale analysis of test–retest reliabilities of self-regulation measures. *Proceedings of the National Academy of Sciences, 116*(12), 5472-5477.

Enkavi, A. Z., Eisenberg, I. W., Bissett, P. G., Mazza, G. L., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019b). Reply to Friedman and Banich: Right measures for the research question. *Proceedings of the National Academy of Sciences, 116*(49), 24398-24399.

Extra Gum: The Story of Sarah & Juan [Advertisement]. (2015, October 7).

Fan, X. (1998). Item response theory and classical test theory: An empirical comparison of their item/person statistics. *Educational and psychological measurement, 58*(3), 357-381.

Feingold, A. (1994). Gender differences in personality: a meta-analysis. *Psychological Bulletin, 116*(3), 429.

Fleeson, W. (2001). Toward a structure-and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology, 80*(6), 1011.

Fleeson, W. (2007). Situation-based contingencies underlying trait-content manifestation in behavior. *Journal of personality, 75*(4), 825-862.

Fleeson, W., & Gallagher, P. (2009). The implications of Big Five standing for the distribution of trait manifestation in behavior: fifteen experience-sampling studies and a meta-analysis. *Journal of personality and social psychology, 97*(6), 1097.

Fox, S., & Spector, P. E. (2000). Relations of emotional intelligence, practical intelligence, general intelligence, and trait affectivity with interview outcomes: It's not all just ‘G’. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior, 21*(2), 203-220.

Fried, E. I. (2017). The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *Journal of Affective Disorders, 208,* 191-197.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron, 50*(4)*,* 531–534.

Frith, U., & Frith, C. D. (2010). The social brain: Allowing humans to boldly go where no other species has been. *Philosophical Transactions of the Royal Society B, 365,* 165–175.

Furnham, A., Richards, S. C., & Paulhus, D. L. (2013). The Dark Triad of personality: A 10 year review. *Social and personality psychology compass, 7*(3), 199-216.

Furnham, A., & Treglown, L. (2021). The dark side of high-fliers: The Dark Triad, high-flier traits, engagement, and subjective success. *Frontiers in Psychology, 12.* https://doi.org/%E2%80%8B10.3389/fpsyg.2021.647676

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends in Cognitive Sciences, 7(*2), 77-83.

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences, 8*, 396–403.

Gardner, H. (2011). *Frames of Mind: The Theory of Multiple Intelligences.* Hachette UK

Garrett, N., Lazzaro, S. C., Ariely, D., & Sharot, T. (2016). The brain adapts to dishonesty. *Nature neuroscience, 19*(12), 1727-1732.

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences, 102,* 74-78. http://dx.doi.org/10.1016/j.paid.2016.06.069

Gneezy, U., Kajackaite, A., & Sobel, J. (2018). Lying aversion and the size of the lie. *American Economic Review, 108*(2), 419-53.

Gold, J. M., Barch, D. M., Carter, C. S., Dakin, S., Luck, S. J., MacDonald III, A. W., ... & Strauss, M. (2012). Clinical, functional, and intertask correlations of measures developed by the Cognitive Neuroscience Test Reliability and Clinical Applications for Schizophrenia Consortium. *Schizophrenia Bulletin, 38*(1), 144-152.

Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower levels of several five-factor models. In F. Ostendorf (Ed.), *Personality psychology in Europe* (Volume 7, pp. 7–28). Tilburg: Tilburg University Press.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in personality, 40*(1), 84-96.

Goleman, D. (1996). *Emotional Intelligence: Why it can Matter More than IQ.* Bloomsbury Publishing.

Gooding, D. C., & Pflum, M. J. (2014). The assessment of interpersonal pleasure: introduction of the Anticipatory and Consummatory Interpersonal Pleasure Scale (ACIPS) and preliminary findings. *Psychiatry research, 215*(1), 237-243.

Gore, W. L., & Widiger, T. A. (2013). The DSM-5 dimensional trait model and five-factor models of general personality. *Journal of Abnormal Psychology, 122(*3), 816.

Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence, 24*(1), 79-132.

Graziano, W. G., & Eisenberg, N. (1997). Agreeableness: A dimension of personality. In R. Hogan, J. A. Johnson, & S. R. Briggs (Eds.), *Handbook of Personality Psychology* (pp. 795–824). Academic Press.

Graziano, W. G., Habashi, M. M., Sheese, B. E., & Tobin, R. M. (2007). Agreeableness, empathy, and helping: A person × situation perspective. *Journal of Personality and Social Psychology, 93*(4), 583–599.

Graziano, W. G., & Tobin, R. M. (2013). The cognitive and motivational foundations underlying agreeableness. *Handbook of cognition and emotion,* 347-364.

Grazioplene, R. G., Chavez, R. S., Rustichini, A., & DeYoung, C. G. (2016). White matter correlates of psychosis-linked traits support continuity between personality and psychopathology. *Journal of abnormal psychology, 125*(8), 1135.

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics (Vol. 1).* New York: Wiley.

Gross, J. J., & Levenson, R. W. (1995). Emotion elicitation using films. *Cognition & Emotion, 9*(1), 87-108.

Grover, S., & Furnham, A. (2021). The Dark Triad, emotional intelligence, self-monitoring and executive coach effectiveness and satisfaction. *Coaching: An International Journal of Theory, Research and Practice,* 1–21.

Gur, R. C., Ragland, J. D., Moberg, P. J., Turner, T. H., Bilker, W. B., Kohler, C., Siegel, S. J., & Gur, R. E (2001). Computerized neurocognitive scanning: I. Methodology and validation in healthy people. *Neuropsychopharmacology, 25*(5)*,* 766–776.

Gurtman, M. B. (2009). Exploring personality with the interpersonal circumplex. *Social and personality psychology compass, 3*(4), 601-619.

Habashi, M. M., Graziano, W. G., & Hoover, A. E. (2016). Searching for the prosocial personality: A big five approach to linking personality and prosocial behavior. *Personality and Social Psychology Bulletin, 42*(9)*,* 1177–1192.

Harms, M. P., Somerville, L. H., Ances, B. M., Andersson, J., Barch, D. M., Bastiani, M., ... & Yacoub, E. (2018). Extending the Human Connectome Project across ages: Imaging protocols for the Lifespan Development and Aging projects. *Neuroimage, 183*, 972-984.

Harrington, L., Siegert, R., & Mcclure, J. (2005). Theory of mind in schizophrenia: A critical review. *Cognitive Neuropsychiatry*, *10*(4), 249–286. https://doi.org/10.1080/13546800444000056

Hasler, G., Drevets, W. C., Manji, H. K., & Charney, D. S. (2004). Discovering endophenotypes for major depression. *Neuropsychopharmacology, 29*, 1765-1781. http://dx.doi.org/ 10.1038/sj.nnp.1300506

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50*(3), 1166-1186.

Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients. *American Psychologist, 58,* 78-80.

Hirsh, J. B., DeYoung, C. G., Xu, X., & Peterson, J. B. (2010). Compassionate liberals and polite conservatives: Associations of agreeableness with political ideology and moral values. *Personality and Social Psychology Bulletin, 36*(5), 655-664.

Hoffman, M. L. (1977). Sex differences in empathy and related behaviors. *Psychological Bulletin*, *84*(4), 712–722. https://doi.org/10.1037/0033-2909.84.4.712

Hofstee, W. K., De Raad, B., & Goldberg, L. R. (1992). Integration of the big five and circumplex approaches to trait structure. *Journal of Personality and Social Psychology, 63*(1), 146.

Holmes, A., Vuilleumier, P., & Eimer, M. (2003). The processing of emotional facial expression is gated by spatial attention: Evidence from event-related brain potentials. *Cognitive Brain Research*, *16*(2), 174–184. https://doi.org/10.1016/S0926-6410(02)00268-9

Hou, X., Allen, T. A., Wei, D., Huang, H., Wang, K., DeYoung, C. G., & Qiu, J. (2017). Trait compassion is associated with the neural substrate of empathy. *Cognitive, Affective, & Behavioral Neuroscience, 17*(5), 1018-1027.

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.

Huber, P. J. (1967). UNDER NONSTANDARD CONDITIONS. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability: Weather modification* (Vol. 5, p. 221). Univ of California Press.

Hughes, C., Cassidy, B. S., Faskowitz, J., Avena-Koenigsberger, A., Sporns, O., & Krendl, A. C. (2019). Age differences in specific neural connections within the Default Mode Network underlie theory of mind. *NeuroImage*, *191*, 269–277. https://doi.org/10.1016/j.neuroimage.2019.02.024

Human, L. J., & Biesanz, J. C. (2011). Through the looking glass clearly: accuracy and assumed similarity in well-adjusted individuals' first impressions. *Journal of Personality and Social Psychology, 100*(2), 349.

Human, L. J., & Biesanz, J. C. (2012). Accuracy and assumed similarity in first impressions of personality: Differing associations at different levels of analysis. *Journal of Research in Personality, 46*(1), 106-110.

Human, L. J., Carlson, E. N., Geukes, K., Nestler, S., & Back, M. D. (2020). Do accurate personality impressions benefit early relationship development? The bidirectional associations between accuracy and liking. *Journal of Personality and Social Psychology, 118*(1), 199.

Hyatt, C. J., Calhoun, V. D., Pearlson, G. D., & Assaf, M. (2015). Specific default mode subnetworks support mentalizing as revealed through opposing network recruitment by social and semantic fMRI tasks. *Human Brain Mapping, 36,* 3047–3063.

Hyman, S. (2002). A new beginning for research on borderline personality disorder. *Biological Psychiatry, 51*, 933-935. http://dx.doi.org/10.1016/S006-3223(02)01408-7

Inagaki, T. K., & Meyer, M. L. (2020). Individual differences in resting-state connectivity and giving social support: implications for health. *Social Cognitive and Affective Neuroscience, 15*(10), 1076-1085.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., ... & Wang, P. (2010). Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *American Journal of Psychiatry, 167*(7), 748-751.

Ioannidis, J. P. A. (2011). Excess significance bias in the literature on brain volume abnormalities. *Archives of General Psychiatry, 68,* 773–780.

Jackson, P. L., Rainville, P., & Decety, J. (2006). To what extent do we share the pain of others? Insight from the neural bases of pain empathy. *Pain, 125*(1), 5–9.

Jang, K. L., Livesley, W. J., Angleitner, A., Riemann, R., & Vernon, P. A. (2002). Genetic and environmental influences on the covariance of facets defining the domains of the five-factor model of personality. *Personality and individual Differences, 33*(1), 83-101.

Jensen-Campbell, L., Adams, R. E., Perry, D. G., Workman, K. A., Furdella, J. Q., & Egan, S. K. (2002). Agreeableness, extraversion, and peer relations in early adolescence: Winning friends and deflecting aggression. *Journal of Research in Personality, 36,* 224–251.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.). *Handbook of Personality: Theory and Research* (p. 114–158). New York: Guilford.

Johnson, M. D., Lim, H. H., Netoff, T. I., Connolly, A. T., Johnson, N., Roy, A., Holt, A., Lim, K. O., Carey, J. R., Vitek, J. L., & He, B. (2013). Neuromodulation for brain disorders: challenges and opportunities. *IEEE Transactions on Biomedical Engineering*, *60*(3), 610–624.

Joiner, T. E., Brown, J. S., & Metalsky, G. I. (2003). A test of the tripartite model's prediction of anhedonia's specificity to depression: Patients with major depression versus patients with schizophrenia. *Psychiatry Research, 119,* 243-250. http://dx.doi.org/10.1016/s0165-1781(03)00131-8

Jonason, P. K., & Webster, G. D. (2010). The dirty dozen: a concise measure of the dark triad. *Psychological assessment, 22*(2), 420.

Jonason, P. K., & Krause, L. (2013). The emotional deficits associated with the Dark Triad traits: Cognitive empathy, affective empathy, and alexithymia. *Personality and Individual Differences, 55*(5), 532-537.

Jones, D. N., & Paulhus, D. L. (2014). Introducing the short dark triad (SD3) a brief measure of dark personality traits. *Assessment, 21*(1), 28-41.

Jones, A. P., Happé, F. G., Gilbert, F., Burnett, S., & Viding, E. (2010). Feeling, caring, knowing: different types of empathy deficit in boys with psychopathic tendencies and autism spectrum disorder. *Journal of Child Psychology and Psychiatry, 51*(11), 1188-1197.

Kaiser, R. H., Treadway, M. T., Wooten, D. W., Kumar, P., Goer, F., Murray, L., ... & Alpert, N. M. (2018). Frontostriatal and dopamine markers of individual differences in reinforcement learning: a multi-modal investigation. *Cerebral Cortex, 28*(12), 4281-4290.

Kajonius, P. J., & Björkman, T. (2020). Individuals with dark traits have the ability but not the disposition to empathize. *Personality and Individual Differences, 155,* 109716.

Kamarck, T. W., Haskett, R. F., Muldoon, M., Flory, J. D., Anderson, B., Bies, R., ... & Manuck, S. B. (2009). Citalopram intervention for hostility: Results of a randomized clinical trial. *Journal of Consulting and Clinical Psychology, 77*(1), 174-188.

Kaplan, J. T., & Iacoboni, M. (2006). Getting a grip on other minds: Mirror neurons, intention understanding, and cognitive empathy. *Social Neuroscience, 1*(3-4), 175-183.

Karjalainen, T., Tuominen, L., Manninen, S., Kalliokoski, K. K., Nuutila, P., Jääskeläinen, I. P., ... & Nummenmaa, L. (2016). Behavioural activation system sensitivity is associated with cerebral μ-opioid receptor availability. *Social Cognitive and Affective Neuroscience, 11*(8), 1310-1316.

Keedwell, P. A., Andrew, C., Williams, S. C., Brammer, M. J., & Phillips, M. L. (2005). The neural correlates of anhedonia in major depressive disorder. *Biological Psychiatry, 58*(11), 843-853. http://dx.doi.org/10.1016/j.biopsych.2005.05.019

Keith, T. Z. (2006). *Multiple Regression and Beyond.* Boston: Allyn & Bacon.

Kerr, L. G., Tissera, H., McClure, M. J., Lydon, J. E., Back, M. D., & Human, L. J. (2020). Blind at first sight: The role of distinctively accurate and positive first impressions in romantic interest. *Psychological Science, 31*(6), 715-728.

Kerr, N., Dunbar, R. I. M., & Bentall, R. P. (2003). Theory of mind deficits in bipolar affective disorder. *Journal of Affective Disorders*, *73*(3), 253–259. https://doi.org/10.1016/S0165-0327(02)00008-3

Kong, R., Li, J., Orban, C., Sabuncu, M. R., Liu, H., Schaefer, A., Sun, N., Zuo, X., Holmes, A. J., Eickhoff, S. B., & Yeo, B. T. T. (2019). Spatial topography of individual-specific cortical networks predicts human cognition, personality, and emotion. *Cerebral Cortex, 29,* 2533–2551. https://doi.org/10.1093/cercor/bhy123

Kong, R., Yang, Q., Gordon, E., Xue, A., Yan, X., Orban, C., ... & Yeo, B. T. (2021). Individual-specific areal-level parcellations improve functional connectivity prediction of behavior. *bioRxiv*. https://doi.org/10.1101/2021.01.16.426943

Koole, S. L., Jager, W., van den Berg, A. E., Vlek, C. A. J., & Hofstee, W. K. B. (2001). On the social nature of personality: Effects of Extraversion, Agreeableness, and feedback about collective resource use on cooperation in a resource dilemma. *Personality and Social Psychology Bulletin*, *27*(3), 289–301. https://doi.org/10.1177/0146167201273003

Kotov, R., Gamez, W., Schmidt, F., & Watson, D. (2010). Linking “big” personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. *Psychological Bulletin, 136,* 768.

Kotov, R., Foti, D., Li, K., Bromet, E. J., Hajcak, G., & Ruggero, C. J. (2016). Validating dimensions of psychosis symptomatology: Neural correlates and 20-year outcomes. *Journal of Abnormal Psychology, 125*(8), 1103.

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., … Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454-477. http://dx.doi.org/10.1037/abn0000258

Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2012). Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychological Medicine, 42,* 1879-1890.

Krueger, R. F., & Markon, K. E. (2014). The role of the DSM-5 personality trait model in moving toward a quantitative and empirically based approach to classifying personality and psychopathology. *Annual Review of Clinical Psychology, 10*, 477-501.

Kwapil, T. R., & Barrantes-Vidal, N. (2015). Schizotypy: looking back and moving forward. *Schizophrenia Bulletin, 41*(suppl_2), S366-S373.

Lammers, J., & Burgmer, P. (2019). Power increases the self-serving bias in the attribution of collective successes and failures. *European Journal of Social Psychology, 49*(5), 1087-1095.

Landy, F. J. (2005). Some historical and scientific issues related to research on emotional intelligence. *Journal of Organizational Behavior, 26*(4), 411-424.

Laske, K., Saccardo, S., & Gneezy, U. (2018). Do fines deter unethical behavior? The effect of systematically varying the size and probability of punishment. *The Effect of Systematically Varying the Size and Probability of Punishment* (April 5, 2018).

Laursen, B., Pulkkinen, L., & Adams, R. (2002). The antecedents and correlates of agreeableness in adulthood. *Developmental Psychology*, *38*(4), 591–603. https://doi.org/10.1037/0012-1649.38.4.591

Lawlor, V. M., Webb, C. A., Wiecki, T. V., Frank, M. J., Trivedi, M., Pizzagalli, D. A., & Dillon, D. G. (2019). Dissecting the impact of depression on decision-making. *Psychological Medicine.* https://doi.org/10.1017/S0033291719001570

Lazarus, R. S., Speisman, J. C., Mordkoff, A. M., & Davison, L. A. (1962). A laboratory study of psychological stress produced by a motion picture film. *Psychological Monographs: General and Applied, 76*(34), 1-35.

Leary, T. (2004). *Interpersonal diagnosis of personality: A functional theory and methodology for personality evaluation.* Wipf and Stock Publishers.

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate behavioral research, 39*(2), 329-358.

Leech, R., & Sharp, D. J. (2014). The role of the posterior cingulate cortex in cognition and disease. *Brain, 137*(1)*,* 12–32.

Lehmann, A. (2016). *Blue Jay.* Los Angeles, CA, Duplass Brothers Productions.

Leppink, J., & Pérez-Fuster, P. (2017). We need more replication research–A case for test-retest reliability. *Perspectives on Medical Education, 6*(3), 158-164.

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in 'theory of mind.' *Trends in Cognitive Sciences*, *8*(12), 528–533. https://doi.org/10.1016/j.tics.2004.10.001

Liddle, B., & Nettle, D. (2006). Higher-order theory of mind and social competence in school-age children. *Journal of Cultural and Evolutionary Psychology, 4,* 231–246.

Litwin, M. S., & Fink, A. (2003). *How to assess and interpret survey psychometrics (Vol. 8).* Sage.

Liu, W. H., Roiser, J. P., Wang, L. Z., Zhu, Y. H., Huang, J., Neumann, D. L., ... & Chan, R. C. (2016). Anhedonia is associated with blunted reward sensitivity in first-degree relatives of patients with major depression. *Journal of Affective Disorders, 190,* 640-648.

Loas, G. (1996). Vulnerability to depression: A model centered on anhedonia. *Journal of Affective Disorders, 41,* 39-53. https://doi.org/10.1016/0165-0327(96)00065-1

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological reports, 3*(3), 635-694.

Lonigro, A., Laghi, F., Baiocco, R., & Baumgartner, E. (2014). Mind reading skills and empathy: Evidence for nice and nasty ToM behaviours in school-aged children. *Journal of Child and Family Studies, 23*(3), 581–590.

Lorenzo-Seva, U., & Ten Berge, J. M. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology, 2*(2), 57-64.

Lozier, L.M., Vanmeter, J.W., Marsh, A.A. (2014). Impairment in facial affect recognition associated with autism spectrum disorders: a meta-analysis. *Developmental Psychopathology, 26*, 933-45.

Lubben, J. (1988). Assessing social networks among elderly populations. *Family & Community Health: The Journal of Health Promotion & Maintenance, 11,* 42-52.

Lucas, R. E., Diener, E., Grob, A., Suh, E. M., & Shao, L. (2000). Cross-cultural evidence for the fundamental features of extraversion. *Journal of Personality and Social Psychology*, *79*(3), 452–468. https://doi.org/10.1037/0022-3514.79.3.452

Ludeke, S. G., Bainbridge, T. F., Liu, J., Zhao, K., Smillie, L. D., & Zettler, I. (2019). Using the Big Five Aspect Scales to translate between the HEXACO and Big Five personality models. *Journal of personality, 87*(5), 1025-1038.

Luxen, M. F., & Buunk, B. P. (2005). Second-to-fourth digit ratio related to verbal and numerical intelligence and the Big Five. *Personality and Individual Differences, 39*(5), 959-966.

Lynam, D. R., & Miller, J. D. (2019). The basic trait of Antagonism: An unfortunately underappreciated construct. *Journal of Research in Personality, 81,* 118-126.

Lyons, M., Caldwell, T., & Shultz, S. (2010). Mind-reading and manipulation—Is Machiavellianism related to theory of mind? *Journal of Evolutionary Psychology, 8,* 261–274.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7,* 19-40. http://dx.doi.org/10.1037/1082-989X.7.1.19

Malouff, J. M., Thorsteinsson, E. B., Schutte, N. S., Bhullar, N., & Rooke, S. E. (2010). The five-factor model of personality and relationship satisfaction of intimate partners: A meta-analysis. *Journal of Research in Personality, 44,* 124–127. https://doi. org/10.1016/j.jrp.2009.09.004.

Manuck, S. B., Flory, J. D., McCaffery, J. M., Matthews, K. A., Mann, J. J., & Muldoon, M. F. (1998). Aggression, impulsivity, and central nervous system serotonergic responsivity in a nonpatient sample. *Neuropsychopharmacology, 19,* 287–299.

Mar, R. A., Spreng, R. N., & DeYoung, C. G. (2013). How to produce personality neuroscience research with high statistical power and low additional cost. *Cognitive, Affective, & Behavioral Neuroscience, 13*(3), 674-685.

Markon, K. E., Chmielewski, M., & Miller, C. J. (2011). The reliability and validity of discrete and continuous measures of psychopathology: A quantitative review. *Psychological Bulletin, 137*(5), 856-879. http://dx.doi.org/10.1037/a0023678

Markon, K. E., Krueger, R. F., & Watson, D. (2005). Delineating the structure of normal and abnormal personality: an integrative hierarchical approach. *Journal of Personality and Social Psychology, 88*(1), 139.

Mars, R. B., Neubert, F. X., Noonan, M. P., Sallet, J., Toni, I., & Rushworth, M. F. (2012). On the relationship between the "default mode network" and the "social brain". *Frontiers in Human Neuroscience, 6,* 189.

Marsh, A. A., Henry, H. Y., Pine, D. S., & Blair, R. J. R. (2010). Oxytocin improves specific recognition of positive facial expressions. *Psychopharmacology, 209,* 225–232.

Martinez Gutierrez, N., & Cribbie, R. (2021). Incidence and interpretation of statistical suppression in psychological research. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement.*

Marx, R. G., Menezes, A., Horovitz, L., Jones, E. C., & Warren, R. F. (2003). A comparison of two time intervals for test-retest reliability of health status instruments. *Journal of clinical epidemiology, 56*(8), 730-735.

Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., & Macrae, C. N. (2007). Wandering minds: The default network and stimulus-independent thought. *Science, 315*, 393–395.

Matheson, G. J. (2019). We need to talk about reliability: making better use of test-retest studies for study design and interpretation. *PeerJ, 7,* e6918.

Mayer, J. D., Roberts, R. D., & Barsade, S. G. (2008). Human abilities: Emotional intelligence. *Annu. Rev. Psychol., 59,* 507-536.

Mazar, N., Amir, O., & Ariely, D. (2008). More ways to cheat-expanding the scope of dishonesty. *Journal of Marketing Research, 45*(6), 651-653.

McAdams, D. P., & Pals, J. L. (2006). A new Big Five: Fundamental principles for an integrative science of personality. *American Psychologist, 61*, 204–217. https://doi.org/10.1037/0003-066X.61.3.204

McCarthy, D., & Davison, M. (1979). Signal probability, reinforcement and signal detection. *Journal of Experimental Analysis of Behavior, 32,* 373-386. http://dx.doi.org/10.1901/jeab.1979.32-373

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality, 60*(2), 175–215. https://doi.org/10.1111/j.1467-6494.1992.tb00970.x

McDonald, R. P. (1999). *Test theory: A unified treatment.* Lawrence Earlbaum Associates. Inc., Mahwah, NJ, 142–145.

McEnrue, M. P., & Groves, K. (2006). Choosing among tests of emotional intelligence: what is the evidence? *Human Resource Development Quarterly, 17*(1), 9-42.

McHugo, G. J., Smith, C. A., & Lanzetta, J. T. (1982). The structure of self-reports of emotional responses to film segments. *Motivation and Emotion, 6*(4), 365-385.

McNulty, J. K., & Russell, V. M. (2016). Forgive and forget, or forgive and regret? Whether forgiveness leads to less or more offending depends on offender agreeableness. *Personality and Social Psychology Bulletin, 42*(5), 616-631.

Meehl, P. E. (1975). Hedonic capacity: Some conjectures. *Bulletin of the Menninger Clinic, 39,* 295-307.

Meier, B. P., Robinson, M. D., & Wilkowski, B. M. (2006). Turning the other cheek: Agreeableness and the regulation of aggression-related primes. *Psychological Science, 17,* 136–142.

Mestre, M. V., Samper, P., Frías, M. D., & Tur, A. M. (2009). Are Women More Empathetic than Men? A Longitudinal Study in Adolescence. *The Spanish Journal of Psychology*, *12*(1), 76–83. https://doi.org/10.1017/S1138741600001499

Meyer, M. L. (2019). Social by default: characterizing the social functions of the resting brain. *Current Directions in Psychological Science, 28*(4), 380-386.

Moeller, S. K., & Robinson, M. D. (2010). Cognitive sources of evidence for neuroticism's link to punishment-reactivity processes. *Cognition and Emotion, 24*(5), 741–759.

Mohr, P., Howells, K., Gerace, A., Day, A., & Wharton, M. (2007). The role of perspective taking in anger arousal. *Personality and Individual Differences, 43,* 507–517.

Montagne, B., Kessels, R. P. C., Frigerio, E., de Haan, E. H. F., & Perrett, D. I. (2005). Sex differences in the perception of affective facial expressions: Do men really lack emotional sensitivity? *Cognitive Processing*, *6*(2), 136–141. https://doi.org/10.1007/s10339-005-0050-6

Montoya, E. R., Terburg, D., Bos, P. A., & Van Honk, J. (2012). Testosterone, cortisol, and serotonin as key regulators of social aggression: A review and theoretical perspective. *Motivation and Emotion, 36*(1), 65-73.

Mooradian, T. A., Davis, M., & Matzler, K. (2011). Dispositional empathy and the hierarchical structure of personality. *The American Journal of Psychology, 124*(1), 99–109.

Mooradian, T. A., Matzler, K., & Szykman, L. (2008). Empathetic responses to advertising: Testing a network of antecedents and consequences. *Marketing Letters, 19*(2), 79–92.

Moore, S. R., Fu, Y., & Depue, R. A. (2014). Social traits modulate attention to affiliative cues. *Frontiers in psychology, 5,* 649.

Moran, J. M., Kelley, W. M., & Heatherton, T. F. (2013). What can the organization of the brain's default mode network tell us about self-knowledge? *Frontiers in Human Neuroscience, 7,* 391.

Morrison, K. E., Pinkham, A. E., Kelsven, S., Ludwig, K., Penn, D. L., & Sasson, N. J. (2019). Psychometric evaluation of social cognitive measures for adults with autism. *Autism Research, 12*(5), 766-778.

Morrone-Strupinsky, J. V., & Depue, R. A. (2004). Differential relation of two distinct, film-induced positive emotional states to affiliative and agentic extraversion. *Personality and Individual Differences, 36*(5), 1109-1126.

Morrone, J. V., Depue, R. A., Scherer, A. J., & White, T. L. (2000). Film-induced incentive motivation and positive activation in relation to agentic and affiliative components of extraversion. *Personality and Individual Differences, 29*(2), 199-216.

Muthén, L.K. and Muthén, B.O. (2017). *Mplus User's Guide. Eighth Edition.* Los Angeles, CA: Muthén & Muthén.

Mwilambwe-Tshilobo, L., Ge, T., Chong, M., Ferguson, M. A., Misic, B., Burrow, A. L., Leahy, R. M., & Spreng, R. N. (2019). Loneliness and meaning in life are reflected in the intrinsic network architecture of the brain. *Social Cognitive and Affective Neuroscience, 14*(4), 423–433. https://doi.org/10.1093/scan/nsz021

Navarra-Ventura, G., Fernandez-Gonzalo, S., Turon, M., Pousa, E., Palao, D., Cardoner, N., & Jodar, M. (2018). Gender differences in social cognition: A cross-sectional pilot study of recently diagnosed patients with schizophrenia and healthy subjects. *Canadian Journal of Psychiatry. Revue Canadienne de Psychiatrie*, *63*(8), 538–546. https://doi.org/10.1177/0706743717746661

Netter, P. (2004). Personality and hormones. In R. M. Stelmack (Ed.), *On the psychobiology of personality: Essays in honor of Marvin Zuckerman* (pp. 353–377). New York: Elsevier.

Nettle, D. (2007). Empathizing and systemizing: What are they, and what do they contribute to our understanding of psychological sex differences? *British Journal of Psychology, 98*(2), 237-255.

Nettle, D., & Liddle, B. (2008). Agreeableness is related to social-cognitive, but not social-perceptual, theory of mind. *European Journal of Personality, 22,* 323–335.

Nguyen, T. V., McCracken, J. T., Albaugh, M. D., Botteron, K. N., Hudziak, J. J., & Ducharme, S. (2016). A testosterone-related structural brain phenotype predicts aggressive behavior from childhood to adulthood. *Psychoneuroendocrinology, 63,* 109–118.

Nosek, B. A., & Smyth, F. L. (2007). A multitrait-multimethod validation of the implicit association test. *Experimental Psychology, 54*(1), 14-29.

Nummenmaa, L., Manninen, S., Tuominen, L., Hirvonen, J., Kalliokoski, K. K., Nuutila, P., ... & Sams, M. (2015). Adult attachment style is associated with cerebral μ-opioid receptor availability in humans. *Human brain mapping, 36*(9), 3621-3628.

Nummenmaa, L., Tuominen, L., Dunbar, R., Hirvonen, J., Manninen, S., Arponen, E., ... & Sams, M. (2016). Social touch modulates endogenous μ-opioid system activity in humans. *NeuroImage, 138,* 242-247.

Ochsner, K. N., Bunge, S. A., Gross, J. J., & Gabrieli, J. D. (2002). Rethinking feelings: an FMRI study of the cognitive regulation of emotion. *Journal of cognitive neuroscience, 14*(8), 1215-1229.

Ones, D. S., Dilchert, S., & Viswesvaran, C. (2012). 10 Cognitive Abilities. *The Oxford Handbook of Personnel Assessment and Selection,* 179.

Østby, Y., Walhovd, K. B., Tamnes, C. K., Grydeland, H., Westlye, L. T., & Fjell, A. M. (2012). Mental time travel and default-mode network functional connectivity in the developing brain. *Proceedings of the National Academy of Sciences*, *109*(42), 16800–16804. https://doi.org/10.1073/pnas.1210627109

Ozer, D. J., & Benet-Martinez, V. (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology, 57,* 401–421.

Paal, T., & Bereczkei, T. (2007). Adult theory of mind, cooperation, Machiavellianism: The effect of mindreading on social relations. *Personality and Individual Differences, 43,* 541–551.

Pardini, D. A., Lochman, J. E., & Frick, P. J. (2003). Callous/unemotional traits and social-cognitive processes in adjudicated youths. *Journal of the American Academy of Child & Adolescent Psychiatry, 42*(3), 364-371.

Park, M. S., Kang, K. J., Jang, S. J., Lee, J. Y., & Chang, S. J. (2018). Evaluating test-retest reliability in patient-reported outcome measures for older people: A systematic review. *International Journal of Nursing Studies, 79,* 58-69.

Patrick, C. J., Kramer, M. D., Krueger, R. F., & Markon, K. E. (2013). Optimizing efficiency of psychopathology assessment through quantitative modeling: Development of a brief form of the Externalizing Spectrum Inventory. *Psychological Assessment, 25,* 1332–1348.

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.

Pedersen, A., Koelkebeck, K., Brandt, M., Wee, M., Kueppers, K. A., Kugul, H., Kohl, W., Bauer, J., & Ohrmann, P. (2012). Theory of mind in patients with schizophrenia: Is mentalizing delayed? *Schizophrenia Research, 137*(1–3)*,* 224–229.

Penner, L. A., Fritzsche, B. A., Craiger, J. P., Freifeld, T. R., Butcher, J. N., & Spielberger, C. D. (1995). Measuring the prosocial personality. *Advances in Personality Assessment, 10,* 147–163.

Philippot, P. (1993). Inducing and assessing differentiated emotion-feeling states in the laboratory. *Cognition and emotion, 7*(2), 171-193.

Phillips, L. H., Channon, S., Tunstall, M., Hedenstrom, A., & Lyons, K. (2008). The role of working memory in decoding emotions. *Emotion*, *8*(2), 184–191. https://doi.org/10.1037/1528-3542.8.2.184

Pinkham, A. E., Harvey, P. D., & Penn, D. L. (2018). Social cognition psychometric evaluation: results of the final validation study. *Schizophrenia Bulletin, 44*(4), 737-748.

Pinkham, A. E., Hopfinger, J. B., Pelphrey, K. A., Piven, J., & Penn, D. L. (2008). Neural bases for impaired social cognition in schizophrenia and autism spectrum disorders. *Schizophrenia Research, 99*(1–3), 164–175.

Pizzagalli, D. A., Bogdan, R., Ratner, K. G., Jahn, A. L. (2007). Increased perceived stress is associated with blunted hedonic capacity: Potential implications for depression research. *Behaviour Research and Therapy, 45,* 2742-2753. http://dx.doi.org/10.1016/j.brat.2007.07.013

Pizzagalli, D. A., Goetz, E., Ostacher, M., Iosifescu, D. V., & Perlis, R. H. (2008a). Euthymic patients with bipolar disorder show decreased reward learning in a probabilistic reward task. *Biological Psychiatry, 64*(2), 162-168.

Pizzagalli, D. A., Iosifescu, D., Hallett, L. A., Ratner, K. G., & Fava, M. (2008b). Reduced hedonic capacity in major depressive disorder: Evidence from a probabilistic reward task. *Journal of Psychiatric Research, 43,* 76-87. http://dx.doi.org/10.1016/j.jpsychires.2008.03.001

Pizzagalli, D. A., Jahn, A. L., & O'Shea, J. P. (2005). Toward an objective characterization of an anhedonic phenotype: A signal-detection approach. *Biological Psychiatry, 57,* 319-327. http://dx.doi.org/10.1016/j.biopsych.2004.11.026

Pizzagalli, D. A., Oakes, T. R., Fox, A. S., Chung, M. K., Larson, C. L., Abercrombie, H. C., Schaefer, S. M., Benca, R. M., Davidson, R. J. (2004). Functional but not structural subgenual prefrontal cortex abnormalities in melancholia. *Molecular Psychiatry, 9,* 393-405. http://dx.doi.org/10.1038/sj.mp.4001469

Poldrack, R. A. (2010). Mapping mental function to brain structure: How can cognitive neuroimaging succeed? *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *5*(5), 753–761. https://doi.org/10.1177/1745691610388777

Polivy, J. (1981). On the induction of emotion in the laboratory: Discrete moods or multiple affect states? *Journal of Personality and Social Psychology, 41*(4), 803.

Pollet, T. V, Roberts, S. G. B., & Dunbar, R. I. M. (2011). Extraverts have larger social network layers: But do not feel emotionally closer to individuals at any layer. *Journal of Individual Differences, 32(3), 161–169.* https://doi.org/10.1027/1614-0001/a000048

Ponsoldt, J., McNulty, T., Levy, S., Lauren, A., Krumm, M., Neustadter, S., Weber, M. H., ... (2014). *The Spectacular Now.* Santa Monica, CA, Lions Gate Films.

Portney, L. G., & Watkins, M. P. (2009). *Foundations of clinical research: applications to practice* (Vol. 892). Upper Saddle River, NJ: Pearson/Prentice Hall.

Preißler, S., Dziobek, I., Ritter, K., Heekeren, H. R., & Roepke, S. (2010). Social cognition in borderline personality disorder: evidence for disturbed recognition of the emotions, thoughts, and intentions of others. *Frontiers in Behavioral Neuroscience, 4,* 182.

Reitman, J., et al. (2008). *Juno*. Beverly Hills, CA, 20th Century Fox Home Entertainment.

Revelle W (2021). *psych: Procedures for Psychological, Psychometric, and Personality Research.* Northwestern University, Evanston, Illinois. R package version 2.1.6, https://CRAN.R-project.org/package=psych.

Revelle, W., & Condon, D. M. (2019). Reliability from α to ω: A tutorial. *Psychological Assessment, 31*(12), 1395–1411.

Revelle, W., Condon, D. M., Wilt, J., French, J. A., Brown, A., & Elleman, L. G. (2016). Web and phone based data collection using planned missing designs. *Sage handbook of online research methods* (2nd ed., p. 578-595). Sage Publications, Inc.

Revelle, W., Dworak, E. M., & Condon, D. M. (2021). Exploring the persome: The power of the item in understanding personality structure. *Personality and Individual Differences, 169,* 109905.

Richard, F. D., Bond Jr, C. F., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology, 7*(4), 331-363.

Ringwald, W. R., Hopwood, C. J., Pilkonis, P. A., & Wright, A. G. (2020). Dynamic features of affect and interpersonal behavior in relation to general and specific personality pathology. *Personality Disorders: Theory, Research, and Treatment.* https://doi.org/10.1037/per0000469

Ringwald, W. R., & Wright, A. G. (2021). The affiliative role of empathy in everyday interpersonal interactions. *European Journal of Personality, 35*(2), 197-211.

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The Power of Personality: The Comparative Validity of Personality Traits, Socioeconomic Status, and Cognitive Ability for Predicting Important Life Outcomes. *Perspectives on psychological science: A journal of the Association for Psychological Science, 2*(4), 313–345. https://doi.org/10.1111/j.1745-6916.2007.00047.x

Roberts, S. G., Wilson, R., Fedurek, P., & Dunbar, R. I. (2008). Individual differences and personal social network size and structure. *Personality and individual differences, 44*(4), 954-964.

Robins, R. W., Fraley, R. C., Roberts, B. W., & Trzesniewski, K. H. (2001). A longitudinal study of personality change in young adulthood. *Journal of personality, 69*(4), 617-640.

Robinson, M. D., Moeller, S. K., & Ode, S. (2010). Extraversion and reward-related processing: Probing incentive motivation in affective priming tasks. *Emotion, 10*(5), 615-626. http://dx.doi.org/10.1037/a0019173

Rosseel, Y. (2012). LAVAAN: An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48*, 1–36. URL http://www.jstatsoft.org/v48/i02/

Rueter, A. R., Abram, S. V., MacDonald III, A. W., Rustichini, A., & DeYoung, C. G. (2018). The goal priority network as a neural substrate of Conscientiousness. *Human Brain Mapping, 39*, 3574–3585.

Rush, A. J., & Weissenburger, J. E. (1994). Melancholic symptom features and DSM-IV. *The American Journal of Psychiatry, 151,* 489-498. http://dx.doi.org/10.1176/ajp.151.4.489

Sabbagh, M. A., Moulson, M. C., & Harkness, K. L. (2004). Neural correlates of mental state decoding in human adults: An event-related potential study. *Journal of Cognitive Neuroscience, 16*(3)*,* 415–426.

Samejima, F. (1997). Graded response model. In *Handbook of modern item response theory* (pp. 85-100). Springer, New York, NY.

Santarnecchi, E., Emmendorfer, A., & Pascual-Leone, A. (2017). Dissecting the parieto-frontal correlates of fluid intelligence: A comprehensive ALE meta-analysis study. *Intelligence, 63,* 9–28.

Santesso, D. L., Dillon, D. G., Birk, J. L., Holmes, A. J., Goetz, E., Bogdan, R., & Pizzagalli, D. A. (2008). Individual differences in reinforcement learning: behavioral, electrophysiological, and neuroimaging correlates. *Neuroimage, 42*(2), 807-816.

Santesso, D. L., Evins, A. E., Frank, M. J., Schetter, E. C., Bogdan, R., & Pizzagalli, D. A. (2009). Single dose of a dopamine agonist impairs reinforcement learning in humans: Evidence from event-related potentials and computational modeling of striatal-cortical function. *Human Brain Mapping, 30*(7), 1963-1976.

Sasamoto, A., Miyata, J., Hirao, K., Fujiwara, H., Kawada, R., Fujimoto, S., Tanaka, Y., Kubota, M., Sawamoto, N., Fukuyama, H., Takahashi, H., & Murai, T. (2011). Social impairment in schizophrenia revealed by Autism-Spectrum Quotient correlated with gray matter reduction. *Social Neuroscience of Psychiatric Disorders, 6,* 548–558.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*(4), 507-514.

Saucier, G. (1992). Openness versus intellect: Much ado about nothing? *European Journal of Personality*, *6*(5), 381–386. https://doi.org/10.1002/per.2410060506

Saxe, R., & Powell, L. J. (2006). It's the thought that counts: Specific brain regions for one component of theory of mind. *Psychological Science, 17,* 692–699. https://doi.org/10.1111/j.1467- 9280.2006.01768.x.

Saxe, R., & Wexler, A. (2005). Making sense of another mind: The role of the right temporo- parietal junction. *Neuropsychologia, 43,* 1391–1399. https://doi.org/10.1016/j.neuropsychologia.2005.02.013.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind". *Neuroimage, 19*(4), 1835-1842.

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X. N., Holmes, A. J., Eickhoff, S. B., & Yeo, B. T. T. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cerebral Cortex, 28*(9), 3095–3114.

Schilbach, L., Bzdok, D., Timmermans, B., Fox, P. T., Laird, A. R., Vogeley, K., & Eickhoff, S. B. (2012). Introspective minds: using ALE meta-analyses to study commonalities in the neural correlates of emotional processing, social & unconstrained cognition. *PloS one, 7*(2), e30920.

Schilbach, L., Eickhoff, S. B., Rotarska-Jagiela, A., Fink, G. R., & Vogeley, K. (2008). Minds at rest? Social cognition as the default mode of cognizing and its putative relationship to the "default system" of the brain. *Consciousness and cognition, 17*(2), 457–467.

Schnabel, K., Asendorpf, J. B., & Greenwald, A. G. (2008). Assessment of individual differences in implicit cognition: A review of IAT measures. *European Journal of Psychological Assessment, 24*(4), 210.

Schimmenti, A., Jonason, P. K., Passanisi, A., La Marca, L., Di Dio, N., & Gervasi, A. M. (2019). Exploring the dark side of personality: Emotional awareness, empathy, and the Dark Triad traits in an Italian Sample. *Current Psychology, 38*(1), 100–109.

Schmidt, F. L., Shaffer, J. A., & Oh, I. S. (2008). Increased accuracy for range restriction corrections: Implications for the role of personality and general mental ability in job and training performance. *Personnel Psychology, 61*(4), 827-868.

Schnabel, K., Asendorpf, J. B., & Greenwald, A. G. (2008). Assessment of individual differences in implicit cognition: A review of IAT measures. *European Journal of Psychological Assessment, 24*(4), 210–217.

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality, 74*(5), 609-612. https://doi.org/10.1016/j.jrp.2013.05.009

Schönemann, P. H. (1966). Varisim: A new machine method for orthogonal rotation. *Psychometrika, 31*(2), 235-248.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin, 86*(2), 420.

Schultebraucks, K., Deuter, C. E., Duesenberg, M., Schulze, L., Hellmann-Regen, J., Domke, A., Lockenvitz, L., Kuehl, L. K., Otte, C., & Wingenfeld, K. (2016). Selective attention to emotional cues and emotion recognition in healthy subjects: The role of mineralocorticoid receptor stimulation. *Psychopharmacology*, *233*(18), 3405–3415. https://doi.org/10.1007/s00213-016-4380-0

Schumacker, R. E., & Lomax, R. G. (2004). *A Beginner's Guide to Structural Equation Modeling.* Psychology Press.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews*, *42*, 9–34.

Sebastian, C. L., McCrory, E. J., Cecil, C. A., Lockwood, P. L., De Brito, S. A., Fontaine, N. M., & Viding, E. (2012). Neural responses to affective and cognitive theory of mind in children with conduct problems and varying levels of callous-unemotional traits. *Archives of General Psychiatry, 69*(8), 814-822.

Seitz, R. J., Nickel, J., & Azari, N. P. (2006). Functional modularity of the medial prefrontal cortex: involvement in human empathy. *Neuropsychology, 20*(6), 743.

Selfhout, M., Burk, W., Branje, S., Denissen, J., van Aken, M., & Meeus, W. (2010). Emerging Late Adolescent Friendship Networks and Big Five Personality Traits:

A Social Network Approach. *Journal of Personality.* https://doi.org/10.1111/j.1467-6494.2010.00625.x

Shamay-Tsoory, S. G., Harari, H., Aharon-Peretz, J., & Levkovitz, Y. (2010). The role of the orbitofrontal cortex in affective theory of mind deficits in criminal offenders with psychopathic tendencies. *Cortex, 46*(5), 668-677.

Shin, D. J., Lee, T. Y., Jung, W. H., Kim, S. N., Jang, J. H., & Kwon, J. S. (2015). Away from home: the brain of the wandering mind as a model for schizophrenia. *Schizophrenia Research, 165*(1), 83–89.

Sibley, C. G., & Duckitt, J. (2008). Personality and prejudice: A meta-analysis and theoretical review. *Personality and Social Psychology Review, 12*(3), 248–279. https://doi.org/10.1177/1088868 308319226.

Silvia, P. J., & Christensen, A. P. (2020). Looking up at the curious personality: individual differences in curiosity and openness to experience. *Current Opinion in Behavioral Sciences, 35*, 1–6.

Simms, L. J., Goldberg, L. R., Roberts, J. E., Watson, D., Welte, J., & Rotterman, J. H. (2011). Computerized adaptive assessment of personality disorder: Introducing the CAT-PD Project. *Journal of Personality Assessment, 93,* 380–389.

Singer, T., & Klimecki, O. M. (2014). Empathy and compassion. *Current Biology, 24,* R875–878.

Slaughter, V., Peterson, C. C., & Moore, C. (2013). I can talk you into it: Theory of mind and persuasion behavior in young children. *Developmental Psychology, 49*(2), 227–231.

Smeets-Janssen, M. M., Roelofs, K., Van Pelt, J., Spinhoven, P., Zitman, F. G., Penninx, B. W., & Giltay, E. J. (2015). Salivary testosterone is consistently and positively associated with extraversion: results from the Netherlands study of depression and anxiety. *Neuropsychobiology, 71,* 76–84.

Smillie, L. D. (2013). Extraversion and reward processing. C*urrent Directions in Psychological Science, 22*(3), 167-172.

Smillie, L. D., Cooper, A. J., & Pickering, A. D. (2011a). Individual differences in reward–prediction–error: extraversion and feedback-related negativity. *Social Cognitive and Affective Neuroscience, 6*(5), 646-652.

Smillie, L. D., Cooper, A. J., Wilt, J., & Revelle, W. (2012). Do extraverts get more bang for the buck? Refining the affective-reactivity hypothesis of Extraversion. *Journal of Personality and Social Psychology, 103*(2), 306–326.

Smillie, L. D., Dalgleish, L. I., & Jackson, C. J. (2007). Distinguishing between learning and motivation in behavioral tests of the reinforcement sensitivity theory of personality. *Personality and Social Psychology Bulletin, 33*(4), 476-489.

Smillie, L. D., Jach, H. K., Hughes, D. M., Wacker, J., Cooper, A. J., & Pickering, A. D. (2019). Extraversion and reward-processing: Consolidating evidence from an electroencephalographic index of reward-prediction-error. *Biological psychology, 146,* 107735.

Smillie, L. D., Lawn, E. C., Zhao, K., Perry, R., & Laham, S. M. (2019). Prosociality and morality through the lens of personality psychology. *Australian Journal of Psychology, 71*(1), 50-58.

Smillie, L. D., Loxton, N. J., & Avery, R. E. (2011b). Reinforcement Sensitivity Theory, Research, Applications and Future. In Chamorro-Premuzic, T., Furnham, A. F., & von Stumm, S. (eds.) *Handbook of Individual Differences.* (pp. 101–131). Wiley-Blackwell.

Smillie, L. D., & Wacker, J. (2014). Dopaminergic foundations of personality and individual differences. *Frontiers in Human Neuroscience, 8*, article 874, https://doi.org/10.3389/fn-hum.2014.00874

Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., ... & Beckmann, C. F. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the National Academy of Sciences, 106*(31), 13040-13045.

Snaith, P. (1993). Anhedonia: A neglected symptom of psychopathology. *Psychological Medicine, 23,* 957-966. http://dx.doi.org/10.1017/S0033291700026428

Song, M., Liu, Y., Yuan, Z., Wang, K., Yu, C., & Jiang, T. (2009). Default network and intelligence difference. *Autonomous Mental Development, IEEE Transactions On*, *2009*, 101–109. https://doi.org/10.1109/TAMD.2009.2029312

Spreng, R. N. (2013). Examining the role of memory in social cognition. *Frontiers in Psychology*, *4,* 1–2. https://doi.org/10.3389/fpsyg.2013.00437

Spreng, R. N., & Andrews-Hanna, J. R. (2015). The default network and social cognition. *Brain Mapping: An Encyclopedic Reference, 3,* 165–169.

Spreng, R. N., Mar, R. A., & Kim, A. S. N. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: A quantitative meta-analysis. *Journal of Cognitive Neuroscience, 21*(3), 489–510.

Spunt, R. P., & Lieberman, M. D. (2012). An integrative model of the neural systems supporting the comprehension of observed emotional behavior. *NeuroImage, 59*(3), 3050–3059. https://doi.org/10.1016/j.neuroimage.2011.10.005

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137-149.

Stein, S. J., & Book, H. E. (2011). *The EQ Edge: Emotional Intelligence and Your Success.* John Wiley & Sons.

Steinberg, L., & Thissen, D. (1995). Item response theory in personality research. In P. E. Shrout & S. T. Fiske (Eds.), *Personality research, methods, and theory: A festschrift honoring Donald W. Fiske* (pp. 161–181). Lawrence Erlbaum Associates, Inc.

Stellwagen, K. K., & Kerig, P. K. (2013). Dark triad personality traits and theory of mind among school-age children. *Personality and Individual Differences*, *54*(1), 123–127.

Stiller, J., & Dunbar, R. I. M. (2007). Perspective-taking and memory capacity predict social network size. *Social Networks, 29,* 93–104.

Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales: a practical guide to their development and use*. Oxford University Press, USA.

Sullivan, H. S. (1953). *The interpersonal theory of psychiatry.* New York: Norton.

Sullivan, H. S. (1964). A Theory of Interpersonal Relations-The Illusion of Personal Individuality. *Varieties of personality theory,* 143.

Sun, J., Kaufman, S. B., & Smillie, L. D. (2018). Unique associations between Big Five personality aspects and multiple dimensions of well-being. *Journal of Personality, 86*(2), 158–172.

Sutherland, G., Newman, B., & Rachman, S. (1982). Experimental investigations of the relations between mood and intrusive unwanted cognitions. *British Journal of Medical Psychology, 55*(2), 127-138.

Suzuki, T., Samuel, D. B., Pahlen, S., & Krueger, R. F. (2015). DSM-5 alternative personality disorder model traits as maladaptive extreme variants of the five-factor model: An item-response theory analysis. *Journal of Abnormal Psychology, 124*(2), 343–354.

Swanner, J. K., & Beike, D. (2015). Throwing you under the bus: High power people knowingly harm others when offered small incentives. *Basic and Applied Social Psychology, 37*(5), 294-302.

Tackett, J. L., Lahey, B. B., Van Hulle, C., Waldman, I., Krueger, R. F., & Rathouz, P. J. (2013). Common genetic influences on negative emotionality and a general psychopathology factor in childhood and adolescence. *Journal of Abnormal Psychology, 122*(4), 1142.

Tager-Flusberg, H. (2011). *False-belief tasks are distinct from theory of mind*. Retrieved from https://www.spectrumnews.org/opinion/viewpoint/false-belief-tasks-are-distinct-from-theory-of-mind/

Tager-Flusberg, H., & Sullivan, K. (2000). A componential view of theory of mind: Evidence from Williams syndrome. *Cognition, 76,* 59–90.

Takeuchi, H., Taki, Y., Hashizume, H., Sassa, Y., Nagase, T., Nouchi, R., & Kawashima, R. (2011). Failing to deactivate: The association between brain activity during a working memory task and creativity. *Neuroimage*, *55*(2), 681–687.

Takeuchi, H., Taki, Y., Nouchi, R., Sekiguchi, A., Hashizume, H., Sassa, Y., Kotozaki, Y., Miyauchi, C. M., Yokoyama, R., Iizuka, K., Nakagawa, S., Nagase, T., Kunitoki, K., & Kawashima, R. (2014). Association between resting-state functional connectivity and empathizing/systemizing. *NeuroImage, 99,* 312–322.

Tamir, D. I., Bricker, A. B., Dodell-Feder, D., & Mitchell, J. P. (2016). Reading fiction and reading minds: The role of simulation in the default network. *Social Cognitive and Affective Neuroscience*, *11*(2), 215–224. https://doi.org/10.1093/scan/nsv114

Talwar, V., Gordon, H. M., & Lee, K. (2007). Lying in the elementary school years: verbal deception and its relation to second-order belief understanding. *Developmental psychology, 43*(3), 804.

Terraciano, A., McCrae, R. R., & Costa Jr, P. T. (2003). Factorial and construct validity of the Italian Positive and Negative Affect Schedule (PANAS). *European journal of psychological assessment, 19*(2), 131.

Terracciano, A., McCrae, R. R., Hagemann, D., & Costa Jr, P. T. (2003). Individual difference variables, affective differentiation, and the structures of affect. *Journal of personality, 71*(5), 669-704.

Thank you, Mom (P&G) - Olympic Games 2012 London [Advertisement]. (2012, July 9).

Thorndike, R. L., & Stein, S. (1937). An evaluation of the attempts to measure social intelligence. *Psychological Bulletin, 34*(5), 275.

Thornton, M. A., & Conway, A. R. A. (2013). Working memory for social information: Chunking or domain-specific buffer? *NeuroImage*, *70*, 233–239. https://doi.org/10.1016/j.neuroimage.2012.12.063

Tompson, S. H., Falk, E. B., Vettel, J. M., & Bassett, D. S. (2018). Network approaches to understand individual differences in brain connectivity: Opportunities for personality neuroscience. *Personality Neuroscience, 1*(e5)*,* 1–12.

Thompson, V. A., & Campbell, J. I. (2004). A power struggle: Between-vs. within-subjects designs in deductive reasoning research. *Psychologia, 47*(4), 277-296.

Thompson, P. M., Stein, J. L., Medland, S. E., Hibar, D. P., Vasquez, A. A., Renteria, M. E., ... & Le Hellard, S. (2014). The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain imaging and behavior, 8*(2), 153-182.

Tripp, G., & Alsop, B. (1999). Sensitivity to reward frequency in boys with attention deficit hyperactivity disorder. *Journal of Clinical and Child Psychology, 28,* 366-375. http://dx.doi.org/10.1207.S15374424jccp280309

Trivedi, M. H., McGrath, P. J., Fava, M., Parsey, R. V., Kurian, B. T., Phillips, M. L., ... & Cooper, C. (2016). Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC): rationale and design. *Journal of Psychiatric Research, 78,* 11-23.

Turan, B., Guo, J., Boggiano, M. M., & Bedgood, D. (2014). Dominant, cold, avoidant, and lonely: Basal testosterone as a biological marker for an interpersonal style. *Journal of Research in Personality, 50,* 84–89.

Tzelgov, J., & Henik, A. (1991). Suppression situations in psychological research: Definitions, implications, and applications. *Psychological bulletin, 109*(3), 524.

Uğurbil, K., Xu, J., Auerbach, E. J., Moeller, S., Vu, A. T., Duarte-Carvajalino, J. M., Lenglet, C., Wu, X., Schmitter, S., Van de Moortele, P. F., Strupp, J., Sapiro, G., De Martino, F., Wang, D., Harel, N., Garwood, M., Chen, L., Feinberg, D. A., Smith, S. M., … Yacoub, E. (2013). Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. *NeuroImage, 80,* 80–104.

van der Linden, D., Pekaar, K. A., Bakker, A. B., Schermer, J. A., Vernon, P. A., Dunkel, C. S., & Petrides, K. V. (2017). Overlap between the general factor of personality and emotional intelligence: A meta-analysis. *Psychological Bulletin, 143*(1), 36.

van der Linden, D., te Nijenhuis, J., & Bakker, A. B. (2010). The general factor of personality: A meta-analysis of Big Five intercorrelations and a criterion-related validity study. *Journal of Research in Personality, 44*(3), 315-327.

van der Linden, D., Tsaousis, I., & Petrides, K. V. (2012). Overlap between General Factors of Personality in the Big Five, Giant Three, and trait emotional intelligence. *Personality and Individual Differences, 53*(3), 175-179.

Van Egeren, L. F. (2009). A cybernetic model of global personality traits. *Personality and Social Psychology Review, 13*(2), 92-108.

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., Ugurbil, K., & WU-Minn HCP Consortium. (2013). *NeuroImage, 80,* 62–79.

Van Rooijen, L., & Vlaander, G. P. (1984). Dramatic induction of depressive mood. *Journal of clinical psychology, 40*(6), 1318-1322.

Van Rooy, D. L., & Viswesvaran, C. (2004). Emotional intelligence: A meta-analytic investigation of predictive validity and nomological net. *Journal of Vocational Behavior, 65*(1), 71-95.

Vellante, M., Baron-Cohen, S., Melis, M., Marrone, M., Petretto, D. R., Masala, C., & Preti, A. (2013). The “Reading the Mind in the Eyes” test: systematic review of psychometric properties and a validation study in Italy. *Cognitive Neuropsychiatry, 18*(4), 326-354.

Velten, E. (1968). A laboratory task for the induction of mood states. *Behaviour Research and Therapy, 6,* 473-482.

Vize, C. E., Ringwald, W. R., Edershile, E. A., & Wright, A. G. (2021). Antagonism in daily life: An exploratory ecological momentary assessment study. *Clinical Psychological Science,* 21677026211013507.

Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happé, F., Falkai, P., ... & Zilles, K. (2001). Mind reading: neural mechanisms of theory of mind and self-perspective. *Neuroimage*, *14*(1), 170–181.

Vonk, J., Zeigler-Hill, V., Ewing, D., Mercer, S., & Noser, A. E. (2015). Mindreading in the dark: Dark personality features and theory of mind. *Personality and Individual Differences, 87,* 50-54.

Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science, 4*(3), 274–290.

Vrieze, E., Pizzagalli, D. A., Demyttenaere, K., Hompes, T., Sienaert, P., de Boer, P., ... & Claes, S. (2013). Reduced reward learning predicts outcome in major depressive disorder. *Biological Psychiatry, 73*(7), 639-645.

Wacker, J., Mueller, E. M., Hennig, J., & Stemmler, G. (2012). How to consistently link extraversion and intelligence to the catechol-O-methyltransferase (COMT) gene: On defining and measuring psychological phenotypes in neurogenetic research. *Journal of Personality and Social Psychology, 102*(2), 427.

Wagner, D. D., Kelley, W. M., & Heatherton, T. F. (2011). Individual differences in the spontaneous recruitment of brain regions supporting mental state understanding when viewing natural social scenes. *Cerebral Cortex, 21*(12), 2788–2796. https://doi.org/10.1093/cercor/bhr074

Wai, M., & Tiliopoulos, N. (2012). The affective and cognitive empathic nature of the dark triad of personality. *Personality and Individual Differences, 52*(7), 794–799.

Wang, X., Blain, S. D., Meng, J., Liu, Y., & Qiu, J. (2021). Variability in emotion regulation strategy use is negatively associated with depressive symptoms. *Cognition and emotion, 35*(2), 324-340.

Wang, X., Blain, S. D., Wei, D., Yang, W., Yang, J., Zhuang, K., ... & Qiu, J. (2020). The role of frontal-subcortical connectivity in the relation between coping styles and reactivity and downregulation of negative emotion. *Brain and Cognition, 146,* 105631.

Wang, Z., Devine, R. T., Wong, K. K., & Hughes, C. (2016). Theory of mind and executive function during middle childhood across cultures. *Journal of Experimental Child Psychology, 149,* 6–22.

Watkin, C. (2000). Developing emotional intelligence. *International Journal of Selection and Assessment, 8*(2), 89-92.

Watson, D., & Clark, L. A. (1999). *The PANAS-X: Manual for the positive and negative affect schedule-expanded form.*

Waytz, A., Zaki, J., & Mitchell, J. P. (2012). Response of dorsomedial prefrontal cortex predicts altruistic behavior. *Journal of Neuroscience, 32*(22), 7646-7650.

Webb, C. A., Dillon, D. G., Pechtel, P., Goer, F. K., Murray, L., Huys, Q. J., ... & Kurian, B. T. (2016). Neural correlates of three promising endophenotypes of depression: evidence from the EMBARC study. *Neuropsychopharmacology, 41*(2), 454-463.

Webb, C. A., Trivedi, M. H., Cohen, Z. D., Dillon, D. G., Fournier, J. C., Goer, F., ... & Adams, P. (2020). Personalized prediction of antidepressant v. placebo response: evidence from the EMBARC study. *Psychological Medicine, 49*(7), 1118-1127.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale--Fourth Edition (WAIS-IV).* [Database record]. APA PsycTests.

Weidmann, R., Ledermann, T., & Grob, A. (2017). Big Five traits and relationship satisfaction: The mediating role of self-esteem. *Journal of Research in Personality, 69,* 102–109. https://doi.org/10.1016/j.jrp.2016.06.001

Weisberg, Y. J., DeYoung, C. G., & Hirsh, J. B. (2011). Gender differences in personality across the ten aspects of the Big Five. *Frontiers in psychology, 2,* 178.

Weiss, B., Jahn, A., Hyatt, C. S., Owens, M. M., Carter, N. T., Sweet, L. H., ... & Haas, B. W. (2021). Investigating the neural substrates of Antagonistic Externalizing and social-cognitive Theory of Mind: An fMRI examination of functional activity and synchrony. *Personality Neuroscience,* 4.

Wheatley, T., Milleville, S. C., & Martin, A. (2007). Understanding animate agents: Distinct role for the social network and mirror system. *Psychological Science, 18*(6)*,* 469–474.

Whitton, A. E., Kakani, P., Foti, D., Van't Veer, A., Haile, A., Crowley, D. J., & Pizzagalli, D. A. (2016). Blunted neural responses to reward in remitted major depression: a high-density event-related potential study. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 1*(1), 87-95.

Widiger, T. A. (2011). Personality and psychopathology. *World Psychiatry, 10*, 103-106. http://dx.doi.org/10.1002/j.2051-5545.2011.tb00024.x

Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of personality and social psychology, 37*(3), 395.

Wilkowski, B. M., Robinson, M. D., & Meier, B. P. (2006). Agreeableness and the prolonged spatial processing of antisocial and prosocial information. *Journal of Research in Personality, 4*0(6), 1152-1168.

Willoughby, E., & Boutwell, B. B. (2018). Importance of Intelligence and Emotional Intelligence for Physicians. *JAMA, 320*(2), 205-205.

Wingenbach, T. S. H., Ashwin, C., & Brosnan, M. (2018). Sex differences in facial emotion recognition across varying expression intensity levels from videos. *PLOS ONE*, *13*(1), e0190634. https://doi.org/10.1371/journal.pone.0190634

Wittenborn, J. R. (1951). An evaluation of the use of difference scores in prediction. *Journal of Clinical Psychology, 7,* 108–111.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: journal of the Econometric Society,* 817-838.

Wright, A. G., Creswell, K. G., Flory, J. D., Muldoon, M. F., & Manuck, S. B. (2019). Neurobiological Functioning and the Personality-Trait Hierarchy: Central Serotonergic Responsivity and the Stability Metatrait. *Psychological Science, 30*(10), 1413-1423.

Wright, A. G., Hopwood, C. J., Skodol, A. E., & Morey, L. C. (2016). Longitudinal validation of general and specific structural features of personality pathology. *Journal of Abnormal Psychology, 125*(8), 1120.

Wright, A. G. C., Krueger, R. F., Hobbs, M. J., Markon, K. E., Eaton, N. R., & Slade, T. (2013). The structure of psychopathology: Toward an expanded quantitative empirical model. *Journal of Abnormal Psychology, 122*(1), 281-294. http://dx.doi.org/10.1037/a0030133

Wright, A. G. C., & Simms, L. J. (2014). On the structure of personality disorder traits: Conjoint analysis of the CT-PD, PID-5, and NEO PI-3 trait models. *Personality Disorders: Theory, Research, and Treatment, 5,* 43–54.

Xiao, M., Zhu, W., Wei, J., Lei, X., & Xia, L. (2019). The relationship among resting-state brain activity and connectivity, agreeableness and displaced aggression: Two possible mediation models. *Journal of Affective Disorders*, *256*, 641–649. https://doi.org/10.1016/j.jad.2019.06.046

Xu, X., Soto, C. J., & Plaks, J. E. (2021). Beyond openness to experience and conscientiousness: Testing links between lower-level personality traits and american political orientation. *Journal of Personality.* https://doi.org/10.1111/jopy.12613

Yarkoni, T. (2009). Big correlations in little studies: Inflated fMRI correlations reflect low statistical power—Commentary on Vul et al. (2009). *Perspectives on Psychological Science, 4*(3), 294–298.

Yarkoni, T., Poldrack, R. A., Van Essen, D. C., & Wager, T. D. (2010). Cognitive neuroscience 2.0: Building a cumulative science of human brain function. *Trends in Cognitive Sciences, 14*(11), 489–496.

Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zollei, L., Polimeni, J. R., Fischl, B., Liu, H., & Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology, 106,* 1125–1165.

Young, L., Dodell-Feder, D., & Saxe, R. (2010). What gets the attention of the temporo-parietal junction? An fMRI investigation of attention and theory of mind. *Neuropsychologia, 48,* 2658–2664. https://doi.org/10.1016/j.neuropsychologia.2010. 05.012.

Zaki, J., Bolger, N., & Ochsner, K. (2008). It takes two: The interpersonal nature of empathic accuracy. *Psychological Science, 19*(4), 399-404.

Zemeckis, R., Tisch, S., Finerman, W., Starkey, S., Roth, E., Burgess, D., In Schmidt, A., ...Paramount Pictures Corporation. (2001). *Forrest Gump.* Hollywood, CA, Paramount Pictures.

Zhang, M., Savill, N., Margulies, D. S., Smallwood, J., & Jefferies, E. (2019). Distinct individual differences in default mode network connectivity relate to off-task thought and text memory during reading. *Scientific Reports*, *9*(1), 16220. https://doi.org/10.1038/s41598-019-52674-9

Zhu, X., Woo, S. E., Porter, C., & Brzezinski, M. (2013). Pathways to happiness: From personality to social networks and perceived support. *Social Networks.* https://doi.org/10.1016/j.socnet.2013.04.005

Zillig, L. M. P., Hemenover, S. H., & Dienstbier, R. A. (2002). What do we assess when we assess a Big 5 trait? A content analysis of the affective, behavioral, and cognitive processes represented in Big 5 personality inventories. *Personality and Social Psychology Bulletin, 28*(6), 847-858.