Eavesdropping on Interpersonal Behavior in Everyday Life

By

JESSIE SUN
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Simine Vazire, Chair

Richard W. Robins

Geoffrey P. Goodwin

Committee in Charge

2020

Acknowledgments

I dedicate this dissertation to my parents, Sophie Zheng and Abel Sun. You worked so hard, made many sacrifices, and immigrated to New Zealand to create a better life for me and Zheng. Thank you for instilling the value of education in me from a young age, for trusting me to find my own path, for supporting my educational aspirations and career goals even when that means being 10,000 (and soon, 14,000) km apart, and for always being proud of me. I hope that a physical Commencement ceremony will eventually happen so that you can fully experience the satisfaction of producing a “Dr. Sun”. In the meanwhile, know that I love you and cannot wait to come home and see you once it is safe to do so.

When I was deciding on who to work with during graduate school, a little bird told me that Simine Vazire was a “good scientist and a good person”. Nearly five years later, I can say that this is both completely true and also a complete understatement. Simine, I already had a deep admiration for you as a scientist before I even arrived in Davis. I admired you for your leadership in the open science community, the transparency with which you reported your articles, and your vision and determination in undertaking a ground-breaking study like the Personality and Interpersonal Roles Study (PAIRS). Since then, you have taught me many things about being a good scientist. You have shown me what it looks like to stand up for scientific ideals—and that it’s worth doing, even when things get ugly. You have encouraged me to better calibrate my claims...I *might* even be ready to replace “people” with “university students” in the title of our next paper. Working with you on PAIRS gave me a newfound appreciation of what constitutes “hard-to-collect data”, taught me that adding any dimension beyond self-reports (e.g., informant reports, longitudinal data, experience sampling) can exponentially increase the value

of a research program, and demonstrated the abundance and costlessness of sharing these data with others without asking for anything in return.

Simine, thank you for also being such an extraordinarily good person. You have been a true role model for what a good mentor looks like. When I needed advice on various professional crises, you immediately jumped on the phone to strategize with me. When I wanted to work with Geoff for a semester, you not only encouraged me to do so, but paid for the visit and for a study that you weren't even involved in. When we were in Croatia, you saw an opportunity for us to hang out and “work” over a biking and donkey-watching expedition. When I had something worth celebrating, you added extra exclamation marks and enthusiastically overcame your natural introverted tendencies to CELEBRATE. When I was on the job market, you seemed even more invested and excited than I was! These are just a few ways in which you've constantly shown me that you have my best interests in mind, that you have my back, and that you care about both my success and my well-being. I want you to know that I have never for a single moment taken your kindness for granted, and I look forward to paying it forward.

I would also like to express my heartfelt thanks to the many other faculty members at Davis and beyond who have contributed to my intellectual development. Mijke Rhemtulla taught me almost everything I know about structural equation modeling, psychometrics, and simulation, and has been a patient mentor, helpful collaborator, and inspiring instructor. Rick Robins has been a wise presence on my advising and dissertation committees, providing valuable feedback on this dissertation as well as my research and professional development more generally. Wiebke Bleidorn and Chris Hopwood pulled no punches with their critiques of my work. Their tough love challenged me to think deeper about the definition of morality and what it means to have a

moral personality. I can't claim to have a convincing answer just yet, but I'm working on it. Thanks also to Chris for sending me several key references on the interpersonal circumplex.

Geoff Goodwin became an important mentor and trusted collaborator to me in 2018. Geoff, I learned so much from visiting your lab for just four months, and our paper together is the one I'm proudest of so far. Thank you for believing in my ideas, for being such a reliable and thorough collaborator, for your support on the job market, and for your thoughtful feedback on this dissertation. I cannot wait to continue our collaborations during my postdoc years and beyond.

Matthias Mehl, whose innovations in developing the Electronically Activated Recorder (EAR) made this research possible, has been incredibly generous with his mentorship and encouragement over the past few years. Matthias, I so appreciate the way you have taken an interest in my career and kept up with my EAR papers. You always have a kind word for the directions I have taken with the EAR. I also wanted to thank you for generously taking the time to coach me on how to respond to challenging questions about the limitations of the EAR.

Luke Smillie, my undergraduate advisor, put me on the right trajectory when he encouraged me to join the Personality Processes Lab. Luke, your mentorship and encouragement gave me the confidence to pursue a research career and helped me to develop my passion for personality science. You built a positive, fun, and supportive lab culture that I hope to recreate one day soon. The lifelong friendships I formed within the PPL are a big part of the reason why Melbourne very much feels like home. I am so glad that we have continued to collaborate throughout my graduate school years, and hope that we will continue to do so for a long time.

I would also like to thank Peggy Kern and Lisa Williams, who were an important part of my early research training and helped me to reach milestones during my undergraduate years. I

started my research journey at Penn as a research assistant to Peggy in 2014, and it feels like I am coming full circle to be starting a postdoc there next year. Lisa trusted me enough to give me the opportunity to co-author my first empirical publication in psychology.

My friends have supported me, inspired me, and added a sense of fun and adventure to my life. Hayley, it is so wonderful that although you are still in Melbourne, our friendship feels stronger than ever. I am grateful that our professional paths have converged such that we can spend quality time together at conferences all over the world. We have shared so many peak moments axe-throwing in Montréal, hiking in Yosemite, eating gelato in Prague, deep-water soloing in Croatia, and climbing on Cát Bá Island. Thank you for being such a constant, reliable, and compassionate presence in my life for so many years.

Michelle, I am so grateful that we immediately became vegan besties in grad school. Your optimism has inspired me to believe that good things are just around the corner even when things look bleak. You have helped me see the world through your eyes, and it is a wonderful world filled with simple pleasures that are already here. I cherish the many deep and meaningful conversations that we shared, and very much hope that we will be friends for a long time.

Lucy, I am glad that we became closer friends over the last year. I always looked forward to catching up with you because we always had so much to talk about. You are so full of surprises and hidden talents, and I admire you for your courage and independence.

Sometimes grad school life can feel little bit too serious. I thank Ted, Elizabeth, and Glenn for balancing that out with well-measured doses of rowdiness. Ted and PsychHaus hosted way more than their fair share of parties that brought all the grad students together. Elizabeth and Glenn brought out my goofy side, inspired me to see and seize opportunities all around, and opened my heart and mind in many other ways.

Kumara also deserves recognition for tolerating excessive amounts of cuddling and for filling my heart with joy. He has been the cutest, fluffiest sidekick, and an extremely comforting presence during the second half of grad school.

As my time in Davis comes to a close, I have come to more fully appreciate—and continue to discover—its hidden gems. Some of my happiest, most productive hours in Davis were spent reading, writing, co-working and holding journal club meetings at Cloud Forest Cafe. I eagerly looked forward to the camaraderie and community of working alongside Olivia, Kailey, Rick, and Aurélien during Cloud Forest Fridays. TH, thank you for creating such a cozy atmosphere, for generously connecting me to your private WiFi network, and for letting me stay and work for unreasonably long sessions. I can only hope to find another space as welcoming as this in my next cities.

Another gem I only discovered this past year was the Latin dance community at Ketmoree. I suspected that I would like dancing, but I didn't know that it would immediately become a life-changing obsesión. I left every Bachata Tuesday feeling exhilarated, alive, connected, and inspired by the beauty and skill I had witnessed on the dancefloor. Bachata Tuesday provided an important weekly reminder that there is much joy to be experienced in life outside of work. I thank all the leads who patiently showed me the ropes, and Ketmoree for creating such a safe space. It was also through Ketmoree that I met Jonathan, who quickly became a loyal friend, dance partner, and occasional gym-buddy. Jonathan, thank you for all the practice sessions, for the rides to the BachaKiz socials, and for teaching me to love Kizomba. I look forward to the day when we can all get back on the dance floor.

Sometimes, hidden gems are revealed in the most unexpected ways. Aurélien, you and I hit it off as friends and intellectual buddies from the moment you arrived in Davis, but it took a

pandemic and a hammock to open my eyes to the sweetness beneath your unassuming facade. You showed up and wished me a calm happiness during the final stretch of this dissertation. The intellectual excitement we shared while nerding out over moral psychology reminded me of all that I love about this job. We experienced the glory of Davis in the springtime, made only more beautiful by our togetherness. Most profoundly—and healingly—you showed me that I am both deserving and capable of tenderness. Although we will part ways too soon, it gives me solace to believe that we will both be better off having had the opportunity to fully know, trust, and accept one another. You are, and will always be special to me.

Finally, I feel extremely lucky to have been blessed with the PAIRS dataset, which formed the basis of this dissertation and my graduate school career. This research would not have been possible without the contributions of past and present members of the Personality and Self-Knowledge Lab—Brittany Solomon, Kathryn Bollich, Robert Wilson, Kelci Harris, Katie Finnigan, and Julia Bottesini—who each played substantial roles in designing PAIRS, collecting and cleaning the PAIRS data, or supervising EAR coding. This research would especially not have been possible without the research assistants at Wash U who collected the data, the nearly 300 research assistants at UC Davis who coded the EAR files over the past six years, or the reliable lab managers who coordinated these efforts. Filled-to-the-brim sticker charts and hundreds of thousands of Qualtrics entries cannot do justice to the brute human effort involved in listening to file after file—many of which contained nothing but silence or otherwise mundane, quotidian sounds of day-to-day existence. I am in awe of your collective efforts and feel humbled to have been a part of this Herculean effort to study what people actually do in everyday life.

Abstract

People are defined in large part by how they relate to and interact with others. Yet, relatively little research has examined what people actually do to each other during their everyday social interactions. Across seven chapters, this dissertation explores the connections between personality traits, interpersonal behaviors, and social consequences. In Chapter 1, I argue for the centrality of behavior in the conceptualization of interpersonal traits. In Chapter 2, I review the tradeoffs between different methods of capturing interpersonal behavior in everyday life and introduce the dataset that I use in the following four empirical chapters. This dataset ($N = 287$) includes personality traits, 18 interpersonal behaviors coded from audio recordings of everyday life, and social consequences reported by friends. The first two empirical chapters (Chapters 3–4) examine the psychometric properties of the 18 interpersonal behaviors. Chapter 3 explores their factor structure and suggests that 14 of the behaviors may reflect more general tendencies towards supportiveness, negativity, mockery, and gossip. Chapter 4 addresses one of the most fundamental issues in personality psychology—behavioral consistency—by showing that these interpersonal behaviors show modest amounts of rank-order stability across one year. Chapter 5 demonstrates that personality traits are manifested in concrete interpersonal behaviors and reveals several gender differences in interpersonal behaviors, but limited implications for friend-reported social consequences. In the last empirical chapter (Chapter 6), I conduct metric calibration to make the interpersonal behavioral manifestations of personality traits (based on the correlates from Chapter 5) more concrete and meaningful by mapping frequencies of behaviors to otherwise arbitrary scores on personality trait measures. Finally, I conclude by discussing the complexities of establishing connections between traits and behaviors and the unique value and challenges of measuring actual behavior (Chapter 7).

Table of Contents

Acknowledgments	ii
Abstract	viii
Chapter 1: Introduction	1
The Role of Actual Behavior in Personality Psychology.....	2
Taxonomies of Interpersonal Traits	5
Behavioral Implications of Interpersonal Traits	9
Overview of the Dissertation	13
Chapter 2: Measuring Interpersonal Behavior in Everyday Life	16
Methods for Measuring Everyday Interpersonal Behavior	16
Interpersonal Behaviors Included in this Dissertation	22
Method Overview	28
Chapter 3: Factor Structure of Everyday Interpersonal Behaviors.....	38
Introduction	38
Method and Results.....	41
Discussion.....	52
Chapter 4: Rank-Order Stability Across One Year	57
Introduction	57
Method and Results.....	60
Discussion.....	65
Chapter 5: Correlates and Consequences of Everyday Interpersonal Behavior.....	68
Introduction	68
Method	82
Results	88
Personality Trait Correlates	88
Gender Differences	91
Social Consequences	94
Discussion.....	95
Chapter 6: Calibration of Arbitrary Metrics.....	108
Introduction	108
Method and Results.....	110
Discussion.....	114

Chapter 7: General Discussion	116
Overview of the Findings	116
Contributions of Observing Actual Interpersonal Behavior	119
Complexities of Establishing Connections Between Traits and Behaviors	120
What Have We Learned About Interpersonal Behaviors?	122
Challenges of Measuring Actual Behavior	124
Limitations	125
Conclusion	127
References	128
Supplemental Tables	155

Chapter 1: Introduction

[W]e are incapable of giving a complete popular description of personality without indicating the manner in which the personality in question stimulates or influences other human beings and the manner in which the behavior of other human beings produces adjustments or responses in the personality in question. (Allport & Allport, 1921, p. 7)

Historically, interpersonal processes have been the domain of social psychology, whereas personality psychology has focused on the individual (Tracy, Robins, & Sherman, 2009).

However, as the opening quote by Allport and Allport (1921) illustrates, personality theorists have always been aware of the potential for intrapsychic variables to influence and be influenced by other people. Moreover, the development of modern trait taxonomies such as the Big Five and HEXACO was based on the assumption that socially relevant personality characteristics should be reflected in language (i.e., the lexical hypothesis; Allport, 1937; Norman, 1967). Far from existing in vacuums, individuals are inextricably embedded in social contexts, and patterns of how we relate to and interact with one another help to define who we are as individuals (Gurtman, 2009; Wiggins, 1979). Cultivating warm and secure relationships, gaining status, treating others well, and being liked and respected are all outcomes that human beings care deeply about. Personality psychologists are well-placed to examine individual differences in the social interaction processes that give rise to these kinds of consequential social outcomes (Back, in press; Back & Vazire, 2015). This dissertation continues in this spirit of bridging the traditional divide between social and personality psychology.

This dissertation explores the connections between personality traits, interpersonal behaviors coded from audio recordings of everyday conversations, and social consequences. Specifically, I examine a) the interrelations among different behaviors (e.g., are sympathetic people less likely to blame others?), b) the extent to which people show stable individual

differences in interpersonal behaviors (e.g., are people who apologize more than others across one week also relatively more apologetic, one year later?), c) the behavioral implications of personality traits (e.g., do narcissists brag more?), and d) the social consequences of such behaviors (e.g., are grateful people liked more?). Ultimately, the goal of this dissertation is to explore the correlates and consequences of individual differences in how people treat each other. In the present chapter, I argue for the unique role of behavior for understanding the social consequences of personality, review the role of behavior in the conceptualization of interpersonal traits, and provide an overview of the dissertation.

The Role of Actual Behavior in Personality Psychology

A central mission of psychology is to describe, explain, and understand the consequences of what people actually do (Baumeister, Vohs, & Funder, 2007; Furr, 2009). Yet, social–personality psychologists have lamented that the field has become “a science of self-reports and finger movements” (Baumeister et al., 2007, p. 396; Furr, 2009). Instead of directly observing *actual behavior*—the ordinary activities and social interactions that people naturally engage in as they go about their daily lives—social and personality psychologists disproportionately rely on people’s self-reports of how they typically act, recollections of past behavior, hypothetical predictions about future behavior, and single behaviors that have tenuous connections with what people normally do in everyday life (e.g., dictator game allocations, Zhao & Smillie, 2015; how intensely a participant ostensibly shocked someone, Milgram, 1963; how much hot sauce a participant gave to someone as a measure of aggression, Ritter & Eslea, 2005).

The study of behavior has special significance for personality psychology specifically (Furr, 2009). At the most basic level, if personality traits comprise relatively stable, enduring patterns of affect, behavior, and cognition (i.e., the “ABCs”; DeYoung, 2015; McAdams & Pals,

2006), a comprehensive science of personality needs to include all three components of the ABCs. Here, I argue that behavior is perhaps the most important component of personality and that this is especially true for interpersonal processes.

Historically, behavioral assessment has played an important role in refuting situationist critiques of personality. Situationists (e.g., Mischel, 1968) argued that the primary determinant of behavior was situations and that because people act differently from one situation to the next, the concept of a trait is not useful. A particularly devastating critique was that the relationships between traits and behavior are too small to be important (Mischel, 1968). Behavioral assessment refuted these critiques by demonstrating that people demonstrate substantial behavioral consistency across situations (Funder & Colvin, 1991; Furr & Funder, 2004). Moreover, although correlations between traits and single behavioral observations tend to be low, correlations with behaviors averaged over a sample of situations can be quite high (Epstein, 1979). This moved the field beyond the person–situation debate—which had intensified the antagonism between personality and social psychology (Kenrick & Funder, 1988)—to a more productive integration of how behaviors are jointly determined by people, situations, and interactions between the two (Donnellan, Lucas, & Fleeson, 1989; Fleeson, 2004).

Behavioral assessment also plays an important role in clarifying the concrete, observable implications of personality trait measures. Although traits cannot be reduced to their behavioral manifestations, behavior is where other psychological ingredients (e.g., thoughts, feelings, desires) become concrete and tangibly consequential for oneself and others (Furr, 2009). Thus, a personality construct is unlikely to be widely interesting or important if it has no behavioral implications (Cronbach & Meehl, 1955; Furr, 2009). A further issue is that personality traits, like many other psychological constructs, are typically measured using arbitrary metrics such as a 5-

point scale (Blanton & Jaccard, 2006). Arbitrary metrics lack clear behavioral referents; for example, it is unclear how a person who scores a “5” on extraversion acts differently compared to someone who scores a “1”. Measuring actual behavior can help reveal the concrete implications of scoring at particular levels of personality trait measures, an issue I will address in Chapter 6.

Finally, in the absence of telepathic communication, behavior is the only component of the ABCs that can be directly observed by others. Thus, traits can only be judged and have an impact on other people via their behavioral manifestations. This idea is reflected in models of interpersonal perception such as the lens model (Brunswik, 1956; see Figure 1) and its extension, the Realistic Accuracy Model (Funder, 1995). Such models assume that people form interpersonal impressions of targets by observing their behaviors, and that accurate personality judgment is achieved to the extent that personality traits are manifested in trait-relevant behaviors that are detected by perceivers. If, on the other hand, people never do anything relevant to their traits, perceivers will be unable to form accurate judgments of their personality traits. Thus, even the most well-intentioned person may be judged as cold-hearted and indifferent if their good intentions never make it out of their heads (e.g., as acts of sympathy, helpfulness, and gratitude). The lens model can easily be extended beyond personality judgment to other social consequences (e.g., being liked, having a positive or negative impact on others). In each case, personality traits only have social consequences via their behavioral manifestations. For this reason, I argue that behavior plays a particularly central role in interpersonal processes, and more generally, in bridging the gap between personality and social psychology.

Figure 1. The lens model (Brunswik, 1956) illustrates the indispensable role of actual behavior in mediating between personality traits and social consequences.

Taxonomies of Interpersonal Traits

Interpersonal traits can be broadly understood as a subset of personality traits that primarily describe individual differences in what people do to each other (Wiggins, 1979). The English language contains at least 800 adjectives (e.g., boastful, manipulative, disrespectful, warm, sympathetic, shy, modest, kind, outgoing, friendly) that capture such traits (Wiggins, 1979). Here, I first review how three major structural models—the Big Five, HEXACO, and Interpersonal Circumplex—describe common dimensions that underlie the vast landscape of interpersonal behaviors. This lays the conceptual foundation for contextualizing the interpersonal behaviors of interest in this dissertation (to be introduced in Chapter 2). In the next section, I argue that interpersonal behavior is central to the definition of major interpersonal traits.

Trait models.

The Big Five. The Big Five model is an integrative, descriptive model of the major dimensions of personality variation (John, Naumann, & Soto, 2008). *Extraversion* describes tendencies towards being outgoing, sociable, assertive, and talkative. *Agreeableness* describes

tendencies towards being helpful, sympathetic, respectful, modest, and trusting.

Conscientiousness describes tendencies towards productiveness, orderliness, and responsibility.

Neuroticism describes the tendency to experience negative emotions such as anxiety, depression, and anger. Finally, *openness* describes tendencies towards intellectual curiosity, creativity, imagination, and aesthetic appreciation. Thus, within the Big Five model, extraversion and agreeableness are the two traits that are most obviously related to interpersonal behavior (but see Chapter 5 for a consideration of the interpersonal implications of conscientiousness, neuroticism, and openness). DeYoung (2015) posits that interpersonal behaviors associated with extraversion are driven by general reward processes, whereas interpersonal behaviors associated with agreeableness reflect affiliative bonding and social regulation systems.

HEXACO. The HEXACO model is a six-factor alternative to the Big Five model (Ashton & Lee, 2007; Ashton et al., 2004). Its eXtraversion, conscientiousness, and openness factors are parallel to those in the Big Five model. However, the agreeableness, emotionality, and honesty-humility factors represent a reorganization of the variance associated with the Big Five agreeableness and neuroticism domains, as well as additional variance not captured within the classic Big Five (Ashton, Lee, & de Vries, 2014). Specifically, tendencies towards anger are captured by the neuroticism domain in the Big Five model but are captured by low agreeableness in the HEXACO model. The HEXACO emotionality domain also captures tendencies towards sentimentality and empathy that associated with Big Five agreeableness rather than neuroticism.

The sixth factor, honesty-humility (H), describes tendencies towards interpersonal sincerity (i.e., being unwilling to manipulate others), fairness, greed avoidance, and modesty. Proponents of HEXACO argue that the H and agreeableness (A) factors are related to different aspects of individual differences in cooperation or reciprocal altruism (Ashton et al., 2014).

Specifically, within the HEXACO model, H represents tendencies towards *active cooperation* (treating others fairly even when you could successfully exploit them), whereas A represents tendencies towards *reactive cooperation* (i.e., tolerating and forgiving others' exploitativeness; being forgiving, lenient, accommodating, and patient). H is thought to capture moral aspects of prosocial behavior that are beyond the scope of the Big Five (Ashton & Lee, 2007); however, there is an ongoing debate about the extent to which H is encompassed by the Big Five. For example, some researchers argue that H shows substantial conceptual and empirical overlap with a narrower aspect of Big Five agreeableness, *politeness*, which captures a consideration of and respect for others' needs and desires (Barford, Zhao, & Smillie, 2015; DeYoung, Quilty, & Peterson, 2007).

Interpersonal circumplex. Finally, the interpersonal circumplex model (IPC; Wiggins, 1979) is a powerful and parsimonious framework for organizing interpersonal phenomena. The IPC proposes that interpersonal variables (e.g., behaviors, needs, motives, problems, traits) can be organized along two orthogonal dimensions: agency (dominance vs. submissiveness) and communion (warmth vs. coldness), depicted in Figure 2. Agency captures phenomena relating to “getting ahead” (e.g., dominance, power, status), whereas communion captures phenomena relating to “getting along” (e.g., love, affiliation, friendliness). It is possible to derive broader (e.g., agreeableness) or narrower (e.g., modesty) interpersonal variables depending on how finely and at which angles we wish to “slice the circumplex pie” (e.g., into 4ths, 8ths, or 16ths; Wiggins, 1979, p. 400).

Figure 2. The interpersonal circumplex and approximate angular positions (bold dashed lines) of extraversion (E), agreeableness (A), and honesty-humility (H) (Barford et al., 2016; DeYoung et al., 2013).

The IPC model assumes that every interpersonal variable involves a blend of agency and communion and that differences between interpersonal variables are largely reducible to differences along these two fundamental dimensions (Gurtman, 2009; cf. Goodwin, Piazza, & Rozin, 2014). Thus, each of the interpersonal traits described above can be mapped onto the IPC. Extraversion, which captures a blend of high warmth and high dominance, typically falls near 60° (Barford et al., 2015; Costa & McCrae, 1989; DeYoung, Weisberg, Quilty, & Peterson, 2013; Pincus, 2002). Big Five and HEXACO agreeableness, which both capture a blend of high warmth and low dominance, typically fall near 330° (Barford et al., 2015; Costa & McCrae, 1989; DeYoung et al., 2013; Pincus, 2002). Finally, honesty-humility also captures a blend of

high warmth and low dominance, falling along the axis running from unassuming–ingenuous to arrogant–calculating (i.e., around 315°).

The geometry of the circumplex also provides a direct measure of the conceptual similarities between different interpersonal behaviors (Gurtman, 1992). The circular ordering implies that a) behaviors that are closer together (e.g., in the same octant) should be more positively related than behaviors that are further apart (e.g., in neighboring octants), b) behaviors that are at right angles to each other (e.g., cold-hearted vs. unassured–submissive) should be uncorrelated, and c) behaviors that are in opposite octants (e.g., gregarious–extraverted vs. aloof–introverted) should be negatively correlated (Markey & Markey, 2009).

Behavioral Implications of Interpersonal Traits

Are interpersonal traits primarily defined by overt behavior, or do thoughts and feelings matter more? For example, can a person truly be agreeable if they often have kind thoughts and feelings, but never do anything kind or helpful? Thoughts and feelings likely have behavioral implications (as I discuss below), can change the interpretation of a given behavior (an issue I consider in Chapters 5 and 7), and are important to understand in their own right. Moreover, in practice, the ABCs of personality tend to cohere, such that people who experience sympathetic affect tend to also have trusting cognitions and express considerate behavior (Wilt & Revelle, 2015). Ultimately, however, I argue that interpersonal traits are primarily defined by what people *do* to each other.

In Table 1, I provide example items from measures of Big Five extraversion, Big Five and HEXACO agreeableness, and honesty-humility to illustrate the role of behavior in the conceptualization and measurement of major interpersonal traits. Behavior tends to be emphasized more than affect and cognition in the measurement of Big Five extraversion,

whereas measures of Big Five agreeableness tend to have relatively balanced amounts of ABC content (Wilt & Revelle, 2015; Zillig, Hemenover, & Dienstbier, 2002). To my knowledge, a formal analysis of the relative balance of ABC content has not been conducted on measures of HEXACO agreeableness and honesty-humility, but as can be seen in Table 1, behavior is well-represented in existing measures.

Beyond the behavioral items in Table 1, thoughts and feelings likely also have behavioral implications. For example, the affective tendency to “Sympathize with others’ feelings”, may manifest behaviorally as kind, caring actions such as reassuring a friend or giving to charity, and anticipatory guilt (“Would feel very badly for a long time if I were to steal from someone”) predisposes people to act in ethical ways (Cohen, Panter, & Turan, 2012). Similarly, a cognitive belief that “I think that I am entitled to more respect than the average person is” would likely be manifested as boastful and condescending actions. Thus, interpersonal behaviors are central to each of these traits and are likely facilitated by accompanying affective and cognitive tendencies.

What might these trait conceptualizations imply about the everyday interpersonal behaviors that should be associated with each of these traits? To provide some examples, a) extraverts should act in more talkative, sociable, gregarious, dominant, attention-seeking, and self-disclosing ways, b) people who are high on Big Five agreeableness should be more helpful, express more sympathy and concern, express more gratitude, and be less antagonistic (e.g., less likely to argue with, criticize, and insult others), c) people who are high on HEXACO agreeableness should express less interpersonal anger, be more accommodating, and be less likely to criticize, complain about, and argue with others, and d) people who are high on honesty-humility should be less likely to act in boastful and condescending ways, and less likely to manipulate, deceive, cheat, and take advantage of others.

Table 1

Example Affect, Behavior, and Cognition Items from Measures of Extraversion, Agreeableness, and Honesty-Humility

	Big Five Extraversion	Big Five Agreeableness	HEXACO Agreeableness	HEXACO Honesty-Humility
Affect	Love excitement	Sympathize with others' feelings	Love my enemies	Would feel very badly for a long time if I were to steal from someone
	Is full of energy	Feel others' emotions	Become frustrated and angry with people when they don't live up to my expectations (r)	I would get a lot of pleasure from owning expensive luxury goods (r)
	Rarely feels excited or eager (r)	Is compassionate, has a soft heart	I rarely feel anger, even when people treat me quite badly	Admire a really clever scam (r)
Behavior	Is talkative	Is helpful and unselfish with others	Rarely complain	If I want something from a person I dislike, I will act very nicely toward that person in order to get it (r)
	Is outgoing, sociable	Show my gratitude	Speak ill of others (r)	Pretend to be concerned for others (r)
	Is dominant, acts as a leader	Is polite, courteous to others	Am nice to people I should be angry at	Put on a show to impress people (r)
Cognition	Reveal little about myself (r)	Starts arguments with others (r)	When people tell me that I'm wrong, my first reaction is to argue with them (r)	Cheat to get ahead (r)
	Know how to captivate people	Believe that I am better than others (r)	Am quick to judge others (r)	I am an ordinary person who is no better than others
	React quickly	Has a forgiving nature	Am inclined to forgive others	I think that I am entitled to more respect than the average person is (r)
	Come up with a solution right away	Assumes the best about people	Find fault with everything (r)	Having a lot of money is not especially important to me

Note. Items are drawn from Wilt and Revelle (2015), DeYoung et al. (2007), Soto and John (2017), Lee and Ashton (2018), and Ashton, Lee, and Goldberg (2007). (r) = reverse-scored.

A final issue is whether interpersonal behaviors should be related to traits in a decontextualized or context-specific way. Some theorists have conceptualized traits as being “nonconditional” and “decontextualized” (e.g., McAdams & Pals, 2006), implying that the expression of traits should not depend on situational features. Other trait theorists argue that traits are contextualized (e.g., Allport, 1937); for example, Tellegen (1981) defined traits as “the disposition to exhibit reaction R under condition S” (p. 219). This perspective implies that traits require appropriate eliciting stimuli in order to be manifested in behavior (DeYoung, 2015). The relevant eliciting stimuli may be more frequently present for some behaviors than for others. For example, individual differences in talkative, outgoing, boastful, argumentative, and helpful behaviors have the opportunity to manifest whenever other people are around. Other behaviors may have more specific elicitors. For example, expressions of gratitude are only relevant when a person perceives that others have done something to benefit them (Algoe, 2012), individual differences in forgiving and patient (vs. retaliatory and angry) behaviors only have the opportunity to manifest in response to the transgressions of others, and it makes no sense to express sympathy in the absence of a suffering other.

These potential context-dependencies imply that interpersonal traits may more strongly predict interpersonal behaviors when relevant eliciting stimuli are present and that the association between interpersonal traits and behaviors may be weaker when generalizing across all situations. If interpersonal traits are still moderately to strongly correlated with interpersonal behavior without considering context, this would imply either that interpersonal traits have decontextualized behavioral implications, or that the relevant eliciting stimuli are often present. If, on the other hand, interpersonal traits are only weakly or are not associated with theoretically

related behaviors, this would be consistent with conditional, contextualized conceptualizations of traits.

Overview of the Dissertation

In this chapter, I have made the case for studying actual behavior. Behavior is usually considered the ultimate outcome of interest—something that has a greater impact on the world and other people than mere thoughts and feelings. Behavior is the only medium by which thoughts and feelings can have an impact on others, and is the only information that others can use to judge a person. For this reason, behavior is especially central to the conceptualization of major interpersonal traits. Finally, studying behavior provides the opportunity to test whether self-reports of interpersonal traits are manifested in interpersonal behaviors across an average of all situations.

Researchers have made substantial progress towards understanding interpersonal behaviors in laboratory settings (for a review, see Back, in press). Studies in the lens model tradition have uncovered behavioral cues that underlie the accuracy of personality judgments (e.g., Borkenau & Liebler, 1992; Cuperman & Ickes, 2009; Funder & Sneed, 1993). For example, Funder and Sneed (1993) found that extraverts tended to speak in a loud voice, were more expressive in their face, voice, or gestures, initiated humor, and were less likely to express insecurity during videotaped interactions with strangers. Funder and Sneed also found that agreeable people tended to laugh more and express warmth, and were less likely to brag, exhibit condescending behavior, and express skepticism or cynicism. Other studies have demonstrated that individual differences in interpersonal behavior show substantial stability across situations (e.g., when interacting with different interaction partners who were instructed to act dominant,

submissive, agreeable, or quarrelsome; Leikas, Lönngqvist, & Verkasalo, 2012) and across time (Funder & Colvin, 1991).

As valuable as these efforts are, behaviors observed in controlled experimental settings are not representative of how people interact with the full range of interaction partners and in the variety of situations they face in their everyday lives. Psychological scientists also need a method for capturing interpersonal behavior in the moments when it naturally happens—an issue I turn to in Chapter 2. Specifically, in Chapter 2, I review methods for observing behavior in everyday life, introduce the 18 interpersonal behaviors of interest in this dissertation, and describe the dataset that I use in the four empirical chapters.

I begin the empirical component of the dissertation by examining the psychometric properties of these interpersonal behaviors (Chapters 3 and 4). In Chapter 3, I explore the factor structure of these interpersonal behaviors. Specifically, what are the broader dimensions that underlie these specific behaviors? What does this reveal about the shared psychological meaning of different interpersonal behaviors? Next, Chapter 4 addresses one of the most fundamental issues in personality psychology—behavioral consistency. To do so, I test the rank-order stability of interpersonal behaviors across one year. In other words, which interpersonal behaviors might more strongly reflect stable individual differences vs. fleeting situational influences?

In Chapter 5, I explore the correlates and consequences of interpersonal behaviors. The first goal of this chapter is to investigate how personality traits are manifested in concrete interpersonal behaviors. For example, how does an agreeable person treat others in their everyday interactions? Are “interpersonal” traits such as extraversion and agreeableness more strongly correlated with everyday interpersonal behaviors than are less interpersonal traits such as neuroticism and self-esteem? This will clarify the behavioral implications of personality traits,

as well as the psychological meaning of everyday interpersonal behaviors. Second, I will explore gender differences in everyday interpersonal behaviors. Third, I will examine whether these interpersonal behaviors have any implications for friend-reported social consequences.

Finally, Chapter 6 aims to make the interpersonal behavioral manifestations of personality traits (based on the correlates from Chapter 5) more concrete and meaningful by mapping frequencies of behaviors to otherwise arbitrary scores on personality trait measures. In doing so, I further illustrate the utility of measuring actual behavior.

Chapter 2: Measuring Interpersonal Behavior in Everyday Life

An ideal method for measuring real-world interpersonal behavior should be valid (i.e., accurately reflects actual behavior), comprehensive (i.e., capture a large amount of visual and auditory information about all interaction partners and across a range of contexts), unobtrusive (i.e., does not change the behavior of the participants and their interaction partners), scalable (i.e., such that it is feasible to achieve large samples of participants, interactions, and behaviors), and able to capture psychologically meaningful interpersonal behaviors (i.e., beyond just the quantity of social interactions). In the first part of this chapter, I review the tradeoffs involved in different methods for measuring everyday interpersonal behaviors, and argue that the Electronically Activated Recorder (EAR), an unobtrusive audio recording device, provides the best possible balance between these goals. In the second part of this chapter, I introduce the interpersonal behaviors and the dataset to be used in subsequent empirical chapters.

Methods for Measuring Everyday Interpersonal Behavior

Self-report methods. The Act Frequency Approach (AFA) was developed by Buss and Craik (1980) as an effort to clarify the behavioral implications of traits and to monitor “what people do all day” (Buss & Craik, 1983, p. 119). This method involved asking participants to provide retrospective self-reports of either whether or not they had ever engaged in each act (e.g., “I continued to apologize for the minor mistake”, “I gave advice, although none was requested”; Buss & Craik, 1983), and if so, how often they did so (e.g., rarely, sometimes, or often) within a designated time interval. The AFA fell out of fashion (but is still occasionally used, e.g., Chapman & Goldberg, 2017) after several significant criticisms emerged. Perhaps the most influential was Block’s (1989) critique that “the act frequency approach” was a misnomer, because “nowhere have acts been directly observed in the AFA” (p. 237). Block questioned the

validity of relying on retrospective self-reports of behavior across long periods of time. A further issue is that rather than providing true frequency metrics, the AFA relied on arbitrary metrics (e.g., how many acts counts as “rarely” vs. “sometimes?”).

The Experience Sampling Method (ESM; Larson & Csikszentmihalyi, 1983; Scollon, Chu, & Diener, 2003) overcomes some of the problems associated with retrospective self-reports such as the AFA. In the ESM, researchers can ask participants to complete behavioral checklists of their behaviors across a short period of time (e.g., whether or not they interacted with someone in the past hour; Sun, Harris, & Vazire, 2019), multiple times per day for several days. By simply asking participants about whether or not they engaged in a specific behavior across a relatively recent and short period of time, the ESM overcomes some of the retrospective biases associated with trying to remember one’s behaviors across a longer period of time. By averaging across all reports, it is also possible to derive true frequency metrics (e.g., the percentage of hours in which a participant interacted with someone). Despite these advantages over retrospective self-reports, ESM reports are still susceptible to other validity issues, including participants’ unwillingness or inability to accurately report their momentary behaviors (Gosling, John, Craik, & Robins, 1998; Sun & Vazire, 2019). The ESM also involves high participant burden because it requires people to actively respond to surveys multiple times each day. This limits the number of behavioral observations that can reasonably be sampled. Finally, because researchers cannot control whether or not participants complete all reports, it is possible that people systematically miss reports when they are engaging in specific interpersonal behaviors (cf. Sun, Rhemtulla, & Vazire, under review).

Behavioral observation methods.

Comprehensive in-person and video observation. In 1949, a team of eight observers took turns every half hour to physically observe and meticulously record every aspect of one boy's day (Barker & Wright, 1951). In a similar spirit, Craik's (2000) research team followed participants around with a handheld video camera. Both approaches would provide a comprehensive description of everyday interpersonal behavior observable to the eye and the ear across all contexts in a person's life, but the obtrusiveness of such methods would likely change the behavior of the participants and those around them. The resource-intensiveness of such methods also reduces the length of observations and sample sizes that would be feasible to obtain.

Scalable and less obtrusive methods. There are other more scalable methods, but existing methods tend to be limited in the type or context of social behavior that can be captured. Much of people's life takes place online nowadays, leaving plentiful behavioral traces of social interactions (especially language use) that occur via social media platforms (e.g., Facebook and Twitter; Park et al., 2015; Youyou, Kosinski, & Stillwell, 2015). However, digital footprints only capture one aspect of how people interact with one another. We also need methods for capturing what happens during in-person, face-to-face social interactions.

Mobile sensing methods provide objective and highly-scalable measures of quantitative social behaviors (e.g., calling, texting, number of social interaction partners; Harari et al., 2019), but such technologies are less well-suited for capturing qualitative aspects of social interactions (e.g., helping, gossiping, criticizing). Wearable cameras that capture pictures of people's natural environments on a fixed interval (e.g., every 30 s; Brown, Blake, & Sherman, 2017) can provide some visual information about one's social context (e.g., number and identity of interaction partners, facial expressions of interaction partners), but provide limited information about what

participants are doing during these social interactions (and no auditory information). Finally, smart-home devices contain video and microphone sensors that could be used to passively observe social behavior, but only in the context of the home (Nelson & Allen, 2018).

All of these methods validly capture actual behavior and are relatively unobtrusive (at least compared to being followed around by a team of observers) and at least somewhat scalable (i.e., less labor-intensive than hiring teams of observers and videographers, though behavioral coding of audiovisual data would still be labor-intensive). However, these methods are limited in their comprehensiveness (i.e., because they capture limited mediums or contexts of social interactions) or their ability to capture psychologically meaningful interpersonal behaviors that happen during face-to-face interactions.

The Electronically Activated Recorder. In this dissertation, I use a method that arguably achieves the best balance between the goals of validity, comprehensiveness, unobtrusiveness, scalability, and psychological meaningfulness—the EAR (Mehl, 2017; Mehl, Pennebaker, Crow, Dabbs, & Price, 2001). The EAR is an unobtrusive wearable recording device that records brief audio snippets of participants’ ambient sounds. These recordings are later coded by human observers for audible behaviors of interest, such as social behaviors. The EAR can be programmed to record snippets of different lengths (e.g., 30 seconds), at different intervals (e.g., every 10 minutes), with blackout periods (e.g., only recording during hours in which the participant is likely to be awake). This sampling strategy better preserves participant privacy and enables larger sample sizes than would be possible with a continuous recording protocol (Mehl, 2017). Because the EAR passively records behavior without requiring participants to respond (i.e., involves low participant burden), this increases the number of social interactions that can be

feasibly captured. The main constraints on scalability are the availability of human coders and time.

Previous validation studies have tested whether the EAR is in fact unobtrusive. Such studies have shown that although people reported being generally aware of the EAR, they also reported that the EAR generally did not change their behavior or the behavior of others around them, and that they generally did not feel impeded in their daily activities (Manson & Robbins, 2017; Mehl & Holleran, 2007). Participants cannot turn the EAR off and cannot tell when it is recording; however, to further preserve privacy, participants are allowed to take the EAR off when they do not want to be recorded and can choose to delete any files that they do not want the researchers to hear. In practice, participants tend to be highly compliant, wearing the EAR for more than 85% of their waking hours (Manson & Robbins, 2017; Mehl & Holleran, 2007), and deleting well under 1% of sound files (Mehl, Robbins, & Deters, 2012). Thus, the EAR generates a representative acoustic log of peoples' everyday lives, allowing researchers to eavesdrop on the everyday social interactions that people naturally engage in.

Of course, as with any method, the EAR has limitations. Some limitations are general to behavioral observation methods. First, some consequential interpersonal behaviors, such as deception, cannot be easily captured because they require that the researchers know the truth. Second, by definition, rare interpersonal acts such as punching someone in the face or saving a life would be difficult to observe using naturalistic methods, because they do not occur frequently enough. Third, the EAR observers provide an external, third-person perspective—one that is complementary to the perspective of the self, who has exclusive access to the intended interpretation of their behaviors. EAR observers have access to only the raw behavior and not the full psychological context. Still, because people regularly act as observers of others' behavior in

everyday life without full knowledge of their underlying motivations and intentions, this external perspective is socially and theoretically important (Mehl, Gosling, & Pennebaker, 2006).

Fourth, EAR-based behavioral observations are not necessarily “objective”. Whereas some behaviors can be relatively objectively coded at a “molecular” level (e.g., whether or not a person was interacting with someone), many psychologically interesting social behaviors (e.g., whether someone was helpful or sympathetic) require observers to form subjective, “molar” judgments. Finally, one EAR-specific limitation is that by capturing only audible information, the EAR is unable to capture nonverbal information (e.g., eye-rolling, smiling) that may provide important context on the nature of people’s interpersonal interactions. However, all things considered, the EAR is optimized for capturing audible social interactions, and especially verbal behaviors expressed in everyday conversations (Mehl, 2017).

A handful of studies have successfully used the EAR to measure interpersonal behaviors in everyday life. The EAR has most typically been used to measure quantitative aspects of social interactions (e.g., time spent alone, time spent on the phone, number of interaction partners; Mehl et al., 2006; Mehl & Pennebaker, 2003; Milek et al., 2018) or language use measured via transcripts of participants’ conversations (e.g., Mehl et al., 2006; Robbins, López, Weihs, & Mehl, 2014; Sun, Schwartz, Son, Kern, & Vazire, 2020; Tackman et al., 2019). Fewer studies have examined qualitative aspects of interpersonal behavior (i.e., what happens during a social interaction). Most of these studies have focused on one or two such behaviors at a time (e.g., conversational depth, Mehl et al., 2010; Milek et al., 2018; conversational depth and self-disclosure, Sun, Harris, et al., 2019; varieties of gossip, Robbins & Karan, 2020; swearing and arguing, Holtzman, Vazire, & Mehl, 2010; Mehl et al., 2006; cf. Kaplan et al., 2018), and have typically explored how such behaviors correlate with either the Big Five traits or one or two

specific traits (e.g., happiness, narcissism, mindfulness; Mehl et al., 2010; Holtzman et al., 2010; Kaplan et al., 2018).

More recently, Bollich and colleagues (2016) explored whether the EAR could be used to assess 14 behaviors that were intended to be indicative of moral character. In developing their coding scheme, they aimed to capture a range of behaviors that were a) audibly detectable, b) occurred with some regularity in everyday life, and c) could be reliably assessed based on pilot coding. These researchers aimed to capture prototypically virtuous behaviors that had a prosocial or antisocial focus (while noting that moral conduct is broader than prosocial or antisocial actions). In the resulting coding scheme, affection, gratitude, praise, sympathy, helping, apologizing, and optimism were considered to be indicative of positive moral conduct, whereas sarcasm, bragging, condescension, complaining, criticism, blame, and pessimism were considered to be indicative of negative moral conduct.

The goal of Bollich and colleagues' (2016) study was primarily psychometric, aiming to establish that their set of behaviors can be reliably coded and reflect stable individual differences. Using two small samples of participants who had worn the EAR across two weekends separated by four weeks ($N_{\text{Sample 1}} = 11$) or 10 weeks ($N_{\text{Sample 2}} = 73$), they showed that individual differences in these behaviors were moderately stable (mean $r = .47$ over 4 weeks, mean $r = .52$ over 10 weeks) and comparable to the stability of individual differences in neutral language behaviors (e.g., the use of articles, prepositions, and adverbs). In doing so, Bollich and colleagues demonstrated that more complex and subjective interpersonal behaviors can be reliably coded from EAR recordings.

Interpersonal Behaviors Included in this Dissertation

In this dissertation, I examine a set of 18 behaviors that includes almost all of the behaviors measured by Bollich and colleagues (2016, described above, except for expressing affection), as well as five additional behaviors—being funny, negative gossip, neutral or positive gossip, teasing others, and expressions of insecurity. Rather than being informed by a specific theoretical framework, this set of behaviors was intended to capture everyday interpersonal phenomena that are intuitively interesting and meaningful to psychologists and ordinary people alike, and could be captured using the EAR (i.e., are potentially audible, occur with some regularity in everyday life, and can be coded reliably). Thus, this dissertation is grounded in an approach that values comprehensive description of real-world phenomena (Rozin, 2001) and recognizes the importance of “know[ing] the thing we are trying to explain” (Asch, 1952/1987, p. 65). In this way, I take an exploratory, hypothesis-generating approach, rather than seeking to test specific hypotheses. Nevertheless, in each of the empirical chapters (Chapters 3–6), I review relevant theoretical perspectives and empirical evidence to provide a context for interpreting the results.

Towards this end, it is worth considering how the final set of behaviors relates to the major interpersonal traits of extraversion, agreeableness, and honesty-humility, as well as the IPC (introduced in Chapter 1). None of the behaviors seem central to the definition of extraversion. Most of the behaviors are most conceptually-related to high (sympathy, praise, help, optimism, gratitude, apologizing) or low Big Five agreeableness (bragging, condescension, teasing, blame, sarcasm, criticism, negativity, complaining). Bragging and condescension seemed relevant to low honesty-humility, whereas gratitude is conceptually-related to high honesty-humility. Expressions of insecurity and being funny did not seem obviously connected

with the three major interpersonal traits. Thus, overall, the set of 18 behaviors primarily reflects Big Five agreeableness-related behaviors.

In Figure 3, I present a hypothetical mapping of the 18 interpersonal behaviors onto the IPC, divided into octants (Wiggins, 1995). As summarized in Table 2, I derived these approximate locations based on a) behavioral mapping of analogous behaviors during dyadic interactions, b) analogous specific items within IPC measures, c), conceptual similarity to the constructs captured by IPC “slices” (e.g., based on the label for the slice), d) Responsibility Exchange Theory, which outlines the effects of thanking, bragging, blaming, and apologizing on perceived competence and warmth (Chaudhry & Loewenstein, 2019), and e) other empirical studies that have mapped analogous self-report measures of the behaviors measured in this study onto the IPC. Although this coding scheme was not designed with the IPC in mind, the 18 behaviors appear to capture various combinations of high and low warmth and dominance.

Thus, taken as a whole, this set of interpersonal behaviors broadly captures tendencies towards warmth, antagonism, provocativeness, self-assurance, and submissiveness that potentially have consequences for the well-being of oneself or others and are psychologically interesting in their own right. Across two additional coding tasks, the larger dataset also includes measures of other interpersonal behaviors. In this dissertation, I chose to focus only on qualitative aspects of interpersonal behavior, and thus did not include data on quantitative aspects of interpersonal behavior (e.g., being alone vs. around others, interacting vs. not, number and type of interaction partners). The larger dataset also includes codings of other qualitative aspects of social interaction (e.g., self-disclosure, conversational depth; reported in Sun, Harris, et al., 2019). I did not include these variables in this dissertation because they were coded using

scale-based judgments (based on an overall impression of all audio recordings in a given hour) rather than concrete, non-arbitrary behavioral counts based on individual audio recordings.

Figure 3. Hypothesized mapping of the 18 interpersonal behaviors onto the interpersonal circumplex.

Table 2

Justification for Approximate IPC Locations of Interpersonal Behaviors

IPC Octant and Behavior	Justification
Assured–Dominant (PA; 90°)	
Funny	Behavioral mapping (“Initiates humor”, 89°) IPC location of affiliative and self-enhancing humor styles is in the NO octant, whereas aggressive humor is in the BC octant (Markey, Suzuki, & Marino, 2014). PA falls in the middle of these two octants.
Gregarious–Extraverted (NO; 45°)	
Help/favors/support	ICL label (“Help, offer, give”; O) Behavioral mapping (“Offers advice”, 55°)
Hope/optimism	IPC location of LOT-R unipolar optimism scores (T. W. Smith, Ruiz, Cundiff, Baron, & Nealey-Moore, 2013)
Warm–Agreeable (LM; 0°)	
Neutral/positive gossip	SBI item (“I spoke favorably of someone who was not present”; Agreeableness)
Sympathy/empathy/concern	IAS item (“Sympathetic”, L) SBI item (“I showed sympathy”; Agreeableness) IPIP–IPC item (“Reassure others”, LM)
Praise/compliment	ICL label (“Affectionate, friendly actions, M)
Unassuming–Ingenuous (JK; 315°)	
Gratitude	ICL item (“Grateful”, J) REF (low competence, high warmth)
Unassured–Submissive (HI; 270°)	
Apologize	ICL item (“Apologetic, H) REF (low competence, high warmth)
Aloof–Introverted (FG; 225°)	
Insecurity/self-doubt	Behavioral mapping (“Says negative things about self”, 224°; “Expresses insecurity”, 242°)
Pessimism	ICL item (“Often gloomy”, F) IPC location of LOT-R unipolar pessimism scores (Smith et al., 2013)
Complain	ICL item (“Complaining”, F)

(table continues)

Table 2 (*continued*)

IPC Octant and Behavior	Justification
Cold-Hearted (DE; 180°)	
Negative gossip	ICL label (“Attack, unfriendly actions”, E)
Criticism	SBI item (“I criticized the other(s)”; Quarrelsomeness) ICL item (“Critical of others”, E)
Sarcasm	SBI item (“I made a sarcastic comment”; Quarrelsomeness) ICL item (“Sarcastic”, D)
Arrogant–Calculating (BC; 135°)	
Blame	REF (high competence, low warmth) ICL label (“Attack, unfriendly actions”; E) IAS item (“Uncharitable”; D) Note that one behavioral mapping study placed “Blames others” at 251°, but most other sources suggest that blame fits better in the BC octant.
Condescension	Behavioral mapping (“Exhibits condescending behavior”, 115°) IPL item (“Somewhat snobbish”, B)
Tease	IPC location of aggressive humor, which captures ridicule and teasing (Markey et al., 2014)
Brag	IPL and IAS item (“Boastful”, B) REF (high competence, low warmth)

Note. IPL = Interpersonal Checklist (Leary, 1957); IAS = Interpersonal Adjectives Scale (Wiggins, 1979); IPIP-IPC = IPIP Interpersonal Circumplex measure (Markey & Markey, 2009); LOT-R = Life Orientation Test (Scheier, Carver, & Bridges, 1994); SBI = Social Behavior Inventory (Moskowitz, 1994); Behavioral mapping = mapping behaviors expressed during a dyadic interaction onto the IPIP-IPC (P. Markey, Anderson, & Markey, 2013); REF = Responsibility Exchange Theory (Chaudhry & Loewenstein, 2019).

Method Overview

Ethics and Open Practices Statement

I used data from Wave 1 and 4 (baseline and one-year follow-up; from now on referred to as “Year 1” and “Year 2,” respectively) of the longitudinal Personality and Interpersonal Roles Study (PAIRS). These were two of the three waves (Waves 1, 4, and 7) that contained EAR codings. I did not include data from Wave 7 (two-year follow-up) because these audio files were not coded for the behaviors of interest in this dissertation. Data collection and coding procedures were approved by Institutional Review Boards at Washington University in St. Louis (IRB ID: 201206090; Study Title: Personality and Intimate Relationships Study) and University of California, Davis (IRB ID: 669518–15; Study Title: Personality and Interpersonal Roles Study).

Several published articles have used the PAIRS dataset (for a full list of citations, see <https://osf.io/3uag4>), including the Big Five trait measures used here (Colman, Vineyard, & Letzring, 2018; Finnigan & Vazire, 2018; Sun et al., 2019; Sutu, Serrano, Schultz, Jackson, & Damian, 2019; Wilson, Harris, & Vazire, 2015) and other EAR-coded interpersonal behaviors (Sun & Vazire, 2019; Sun, Harris, et al., 2019). However, this dissertation provides the first report of analyses involving the interpersonal behaviors described below.

Below, I describe the measures and procedures relevant to the current dissertation. Codebooks for all measures in the larger study are available at <https://osf.io/akbfj>. I did not preregister any of these analyses because the data were collected years ago and I was familiar with the dataset and had run some preliminary analyses before conceptualizing this dissertation. Thus, all results are exploratory and any interesting patterns should be interpreted with caution.

Participants and Procedure Overview

PAIRS involved 434 students at Washington University in St. Louis, who were recruited in 2012 and 2013 via flyers and classroom announcements across the campus. Participants completed a battery of self-report questionnaires during an initial laboratory-based assessment (\$20 compensation), and nominated friends who reported on the participants' traits and interpersonal outcomes. These self- and friend-reported measures are described in Chapter 5. For the next two weeks, participants completed experience sampling method (ESM) measures four times per day (for a 1 in 10 chance of winning \$100 if all ESM surveys were completed) that were not relevant to this dissertation. In addition, most participants ($N = 311$) wore the EAR for the first week of the ESM protocol (\$20 compensation), providing audio recordings of their everyday lives that were later coded for interpersonal behaviors (described below). One year later (Year 2), 153 (49.2%) of the participants who wore the EAR in Year 1 wore the EAR for another week. Recruitment ended at the end of the semester in which the research team had recruited at least 400 participants in Year 1.

I use behavioral data from both Years 1 and 2 to examine the factor structure and rank-order stability of these behaviors across one year. I use self-report, informant report, and behavioral data from Year 1 to answer questions about the correlates of between-person differences in interpersonal behaviors. See Table 3 for the demographic characteristics of participants included in the different analyses (inclusion criteria are described below).

Table 3

Demographic Characteristics of Participants in Each Subset

	Year 1 (<i>n</i> = 287)	Year 2 (<i>n</i> = 141)	Stability (<i>n</i> = 124)
Gender (%)			
Female	67.60	75.89	33.80
Male	31.71	24.11	9.41
Not reported			
Mean (<i>SD</i>) age in years	19.14 (1.76)	20.12 (2.06)	18.98 (1.76)
Ethnicity (%)			
White	56.10	51.06	52.42
Asian or Asian American	23.00	28.37	25.81
Black or African American	9.76	9.22	9.68
Other or Multiple	8.36	7.80	8.06
Not Reported	2.44	2.84	3.23
American Indian or Alaska Native	0.35	0.71	0.81

Note. Year 1 and Year 2 = participants who spoke enough to be included in the factor analyses. Stability = participants who spoke enough in both Year 1 and Year 2 to be included in the rank-order stability analyses. Demographics for the stability subset were based on data from Year 1.

EAR Procedures

EAR data collection. 311 participants wore the EAR for 6–8 days. The EAR was implemented through the iEAR app using an iPod Touch device. The EAR component of the study was optional, was only offered during non-summer months of the study, and was not an option when all of the researchers' iPod Touches were in use by other participants. The EAR was programmed to record 30 s audio snippets of participants' ambient sounds, every 9.5 min from 7 a.m.–2 a.m. Participants were encouraged to wear the EAR as much as possible and to wear it clipped to a waistband or the outside of their pockets (i.e., not inside a bag or pocket). Although participants had no way to tell when the device was recording, they were told that they could decide to not wear the EAR at any time for any reason. After 3–4 days, participants returned to the lab to upload their data (due to device memory limitations) and then continued wearing the device before returning it after another 3–4 days.

Upon returning the device at the end of the week, participants received a compact disk with their recordings, so that they could listen to and erase any files they did not want the researchers to hear. In Year 1, only a few participants ($n = 15$) chose to erase a total of 99 files. After deleting these files, along with files from six participants who withdrew, and one participant who only had silent recordings (suggesting that the microphone malfunctioned), 152,592 usable recordings from 304 participants remained. In Year 2, only a few participants ($n = 6$) chose to erase a total of 44 files. After deleting these files, 84,145 usable recordings from 153 participants remained.

EAR codings. Although participants could have thanked others or expressed sympathy (for example) through inaudible non-verbal behaviors or computer-mediated communications, it was only possible for us to detect these behaviors when they were spoken out loud. Thus, research assistants from the University of California, Davis first listened to the EAR recordings from Year 1 and Year 2 and transcribed the words that participants spoke, in order to identify the files that contained participants' decipherable speech. The full details of the transcription procedure are available elsewhere (Sun et al., 2020). A total of 31,209 files (Year 1) and 14,685 files (Year 2) contained at least one decipherable word spoken by the participant.

In a separate task, research assistants listened to only the files that contained participants' decipherable speech and coded the interpersonal behaviors that participants enacted in each 30 s file. The full behavioral checklist (see Table 4) included 18 interpersonal behaviors, adapted from Bollich et al. (2016). Any single file could contain and be coded for more than one interpersonal behavior (e.g., a participant could be coded as expressing sympathy and blaming someone in the same file).

Table 4

Everyday Interpersonal Behaviors and Examples from EAR Transcripts

#	Original item	Revised item	Example (adapted from EAR transcripts)
1.	*Gratitude	Expressed thanks or gratitude	Well thank you for the spare charger by the way.
2.	*Praise/compliment	Complimented or praised	It's such a pretty color on you!
3.	*Sympathy/concern	Expressed sympathy, empathy, or concern	How did you sleep? Oh. That sucks.
4.	*Help/support	Offered favors, help, or support	Are you sick too? Do you want meds for that?
5.	*Apologize	Apologized	Yeah, sorry we're taking a while.
6.	*Hope/optimism	Expressed hope or optimism	I'm really really excited. I know it's going to be great. Just need to get through this next week and a half.
7.	Funny	Did or said something funny	I had to smuggle you out like I was smuggling heroin across the border.
8.	Negative gossip	Gossiped about people in a negative way	She's just so freaking rude I just can't believe it!
9.	Neutral/positive gossip	Gossiped about people in a neutral or positive way (not negative)	Oh did Anne send you the SnapChat of Mary dancing to uh tick tock? It's so cute.
10.	*Sarcasm	Spoke sarcastically	Thank you for pointing that out, daddy. I wasn't aware of my own life goals. Thank goodness I have you here to guide me.
11.	*Brag	Bragged	I skipped medicine and society. I'm such a badass 'cause there's only 20 kids in that class!
12.	*Condescending/arrogant	Acted condescending	I don't think you quite understand the way things work around here.
13.	*Complain/whine	Complained or whined	I'm really sad because my nail polishes are arriving at different times.
14.	*Criticism	Criticized people (present or not)	What kind of mother calls her child a bitch like what is wrong with you?
15.	*Blame	Blamed someone or something	His questions were different from the practice exams and that's why they got me.
16.	*Pessimism	Expressed pessimism	I have an exam tomorrow that I'm legitimately gonna fail.
17.	Teasing others	Teased	Have you never gotten something from the deli before, huh, huh, huh?
18.	Insecurity/self-doubt	Showed insecurity in themselves or self-doubt	I am pretty damn stupid when it comes to like I can't memorize anything.

Note. EAR coders were asked to “Check all behaviors that apply to the participant during this file”. For the ESM-matched files in Year 1, coders listened for the full list of 18 behaviors. To simplify this task for the non-ESM-matched Year 1 files and all Year 2 files, I split up the task, such that one team of coders listened for behaviors 1–9, and the other team of coders listened for behaviors 10–18. Behaviors that were measured in Bollich et al. (2016) are marked with *.

Because research assistants joined and left the lab at different times (as this coding project spanned several years), each participant was coded by a different set of coders. Coders listened to each 30 s file that contained speech, recorded all behaviors (described below) that applied to the participant during the file, then moved onto the next 30 s file that contained speech. Coders listened to the files in chronological order and coded all files for a given participant before moving onto the next participant. Some coders were involved in multiple tasks (Year 1 transcription, Year 2 transcription, Year 1 interpersonal behavior coding, Year 2 interpersonal behavior coding), but they never coded the same participant across different tasks or years. If coders left the lab before they coded all speech-containing files for a given participant, their participant was re-assigned to a new coder. If a coder had coded $\geq 80\%$ of the speech-containing files, I considered their codings to be complete and retained them; otherwise, I excluded them.

For the Year 1 files, the research assistants initially only coded the 7,784 speech-containing files that occurred during the hours that matched the ESM reports (i.e., 11 a.m.–12 p.m., 2 p.m.–3 p.m., 5 p.m.–6 p.m., and 8 p.m.–9 p.m.). The original coding scheme used relatively short item labels (e.g., “gratitude”; see Original Item column in Table 4). When I added additional coders halfway through the ESM-matched-hour codings, I decided to expand the item descriptions (e.g., “expressed thanks or gratitude”; see Revised Item column in Table 4) to clarify the meaning of each behavior. For these codings, I aimed for four coders per participant and ended up with 267 participants who had four or more coders, 16 participants who had three coders, and 4 participants who had two coders.

I later decided to also obtain codings for the speech-containing files outside of these ESM-matched hours, to increase the reliability of the between-person estimates. Before coding

the non-ESM-matched hours, I modified the coding protocol again to try to improve the reliability of the codings. Instead of coding all 18 behaviors in the full checklist, I split up the list into two lists, such that one team of coders only coded nine behaviors, and the other team of coders only coded the other nine behaviors (described in Table 4). For each list of behaviors, I aimed for three coders per participant (i.e., six coders coded each participant, but two teams of three coders each coded different sets of behaviors). Due to time constraints, most participants were only double-coded. Specifically, for the first checklist, 70 participants were at least triple-coded, 181 were double-coded, and 34 had only one coder. For the second checklist, 85 participants were at least triple-coded, 181 were double-coded, and 19 had only one coder.

Coders used this split checklist to code all speech-containing files in Year 2, aiming for three coders per participant. For the first checklist, 126 participants were triple-coded, 13 were double-coded, and 2 had only one coder. For the second checklist, 121 participants were triple-coded, 16 were double-coded, and 2 had only one coder.

Inclusion criteria. Given the low base rates of these behaviors (see Table 5), to provide more reliable estimates of average interpersonal behavior, I only included participants who had at least 20 files that contained speech in the analyses for the respective year. Only participants who had enough speech-containing files in both Year 1 and Year 2 were included in the rank-order stability analyses.

Table 5

Descriptive Statistics for the Percentage of Conversations that Contained Each Behavior

Behavior	Year 1				Year 2			
	<i>M</i>	<i>SD</i>	% Never	ω	<i>M</i>	<i>SD</i>	% Never	ω
Gratitude	2.50	2.62	10.45	.77	2.73	2.57	9.93	.76
Praise/compliment	5.37	3.52	3.14	.62	5.29	3.94	2.84	.49
Sympathy/concern	9.17	5.74	0.35	.10	9.44	7.19	2.84	.00
Help/favors/support	9.58	5.93	1.39	.54	5.91	3.78	2.13	.15
Apologize	1.06	1.41	26.13	.77	1.06	1.16	28.37	.69
Hope/optimism	7.97	6.30	3.14	.23	9.16	9.49	1.42	.00
Funny	11.04	6.05	0.35	.54	9.89	5.92	0.71	.26
Negative gossip	3.99	3.20	5.23	.59	3.45	2.57	7.80	.55
Neutral/positive gossip	11.49	6.83	0.35	.62	10.36	5.25	2.13	.42
Sarcasm	2.90	2.62	8.36	.39	2.99	2.59	8.51	.24
Brag	1.28	1.47	22.30	.41	1.15	1.52	31.91	.18
Condescension	2.22	2.48	14.98	.37	1.16	1.49	26.24	.24
Complain	8.92	5.22	0.35	.62	9.60	6.48	2.13	.40
Criticism	4.50	3.13	2.79	.52	4.15	3.67	5.67	.32
Blame	1.01	1.09	21.95	.35	1.06	1.20	29.08	.12
Pessimism	2.63	2.55	11.15	.31	2.24	3.42	17.02	.06
Tease	3.62	2.91	6.27	.48	3.87	3.63	7.09	.27
Insecurity/self-doubt	3.36	2.81	6.27	.32	2.80	3.39	9.93	.18

Note. *M* = mean percent of files containing the behavior, out of all files containing participants' speech. Means were first computed for each participant, then aggregated across participants. *SD* = between-person standard deviation; % never = percentage of participants who never expressed a given behavior; ω = omega reliability estimate based on a CFA model with coders as indicators.

Aggregating codings. To estimate the percentage of speech-containing files for a given participant that contained each interpersonal behavior, I first aggregated the codings across each coder, for each participant that had at least 20 speech-containing files. These coder-participant aggregates were used as indicators when estimating the inter-reliability of the codings (described below) and for latent variables in the stability analyses (described in Chapter 4). Then, for each participant, I aggregated the codings across all available coders. These observed estimates were

used for descriptive statistics (see Table 5), factor analyses (Chapter 3), and between-person correlates (Chapters 5 and 6).

As shown in Table 5, most of these interpersonal behaviors had fairly low base rates (in the range of 1–11% of everyday conversations), and the base rates were fairly similar across the two years. The most common interpersonal behaviors involved being funny, helping others, expressing sympathy, and complaining ($> 8.9\%$ in Year 1); the rarest behaviors were blaming others, apologizing, bragging, and being condescending ($< 2.2\%$ in Year 1).

Inter-coder reliability. To estimate the extent to which sets of coders were able to reliably differentiate participants who expressed each behavior more or less often, I used the omega reliability coefficient (McDonald, 1999), estimated from a confirmatory factor analysis (CFA) model in which the indicators were aggregates based on each coder, with equal loadings and residuals. In Year 1, there were seven indicators in total, based on two “groups” of coders: 1) aggregates based on files recorded during ESM-matched hours in Year 1, and 2) aggregates based on files recorded outside of ESM-matched hours in Year 1. In Year 2, there were three indicators, which were based on coder aggregates across all files. Within each group (Year 1 ESM-matched, Year 1 non-matched, Year 2 all files), for a given participant (e.g., participant 1), all ratings from coder 1 were from the same coder (e.g., research assistant 1). However, for a different participant (e.g., participant 2), coder 1 could have been a different research assistant (e.g., research assistant 2). To model the interchangeability of coders within each group, I constrained all coder loadings to be equal (within each group), and constrained the residual variances to be equal (within each group). However, the loadings and residuals were free to vary between each group. I fit a separate CFA model for each behavior, for each year, using the R package *lavaan* (Rosseel, 2012).

As shown in Table 5, reliability estimates were generally higher in Year 1 than in Year 2. Gratitude ($\omega_{\text{Year 1}} = .77$; $\omega_{\text{Year 2}} = .76$) and apologizing ($\omega_{\text{Year 1}} = .77$; $\omega_{\text{Year 2}} = .69$) were the most reliably-coded behaviors, whereas sympathy ($\omega_{\text{Year 1}} = .10$; $\omega_{\text{Year 2}} = .00$) and optimism ($\omega_{\text{Year 1}} = .23$; $\omega_{\text{Year 2}} = .00$) were the least reliably coded. Reliabilities for other behaviors ranged between .31–.62 in Year 1, and between .06–.55 in Year 2.

Data Analyses

Models and specific analytic procedures for each research question will be described in their respective chapters. Here, I describe analytic decisions that apply across the dissertation. I conducted analyses using the R programming environment (Version 3.6.3; R Core Team, 2018). Average correlations or reliability coefficients and comparisons of correlations or factor loadings (e.g., using *t*-tests) were computed after applying the Fisher *r*-to-*z* transformation (and, where applicable, the reverse transformation to transform the mean *z* back to an *r*). Sensitivity power analyses were conducted using the R package *pwr* (Champely et al., 2017).

Because some participants had fewer coders than the intended number (e.g., 2 coders rather than 3), some participants had missing data on some coder aggregates. To handle missing data in these cases (and missing data on other variables, including scale items), I used pairwise complete observations for analyses involving observed variables, and full information maximum likelihood estimation for analyses involving latent variables. This approach makes use of all available information (compared to listwise deletion, for example).

Chapter 3: Factor Structure of Everyday Interpersonal Behaviors

Factor analyses of self-report measures have shown over and over again that large sets of psychological variables can be reduced into a smaller number of higher-order dimensions (e.g., Ashton et al., 2004; De Raad, Perugini, Hrebicková, & Szarota, 1998; Saucier & Goldberg, 2001). Following this logic, in this chapter, I explore the factor structure of the 18 interpersonal behaviors. As explained in Chapter 1, the selection of the 18 interpersonal behaviors was based on practical considerations and intuitions about which behaviors would be most interesting to capture, rather than a particular theoretical framework. In this chapter, I explore whether an informative and coherent structure will emerge from a set of behaviors that were chosen without a particular structure in mind. Rather than aiming to discover the “true” underlying structure of these interpersonal behaviors (i.e., attempting to carve nature at its joints), I had two purposes in mind in exploring whether the 18 behaviors can be summarized by a smaller set of data-driven dimensions.

First, composites of conceptually linked behaviors tend to be more reliable than individual behaviors (Block, 1989; Jaccard, 1974; Spearman, 1910). For example, Jaccard (1974) found that self-reports of dominance correlated at a mean of .20 with each of 40 self-reported behaviors, but up to .64 with the aggregate of these 40 behaviors. He therefore concluded that “measures of personality should be highly related to multiple act criteria, but not necessarily to single act criteria” (p. 362). In other words, to the extent that it is possible to derive broader and more reliable composites of conceptually linked behaviors, behavioral composites (compared to individual behaviors) should show higher rank-order stability (Chapter 4) and larger correlations with personality traits and social consequences (Chapter 5). Thus, the primary purpose of this chapter is to derive composites that would allow me to examine the rank-order stability and

correlates of not only individual behaviors but also behavioral composites in subsequent chapters.

A secondary purpose is to generate hypotheses about the more general interpersonal meanings, causes, or functions that specific interpersonal behaviors share. In other words, what are the broader traits that underlie specific behavioral manifestations? Because the set of behaviors included in this study was not designed to comprehensively sample the interpersonal space, these analyses cannot speak to the “true” factor structure of interpersonal behaviors (unlike, for example, Wiggins, 1979). However, factor analysis can at least help to reveal the psychological meaning of everyday expressions of the interpersonal behaviors included in this dissertation. For example, if gratitude loads onto the same factor (but in a reverse direction) as bragging and condescension, this might suggest that everyday expressions of gratitude primarily reflect humility. If, however, gratitude loads more strongly onto the same factor as praise, sympathy, and helping, this might suggest that gratitude reflects general interpersonal warmth.

Plausible Factor Structures

Which factors might emerge? In other words, what are the common meanings that the different interpersonal behaviors might share? Although I did not have strong theoretical predictions before conducting the analyses, it is worth considering different factor structures that might have plausibly emerged.

Intuitive groupings. One possibility is that the behaviors will simply separate into a positive cluster (gratitude, praise, sympathy, helping, apologizing, optimism, funny, neutral or positive gossip) and a negative cluster (negative gossip, sarcasm, bragging, condescension, complaining, criticism, blame, pessimism, teasing, insecurity). In reflecting on the conceptual and functional similarities between different behaviors (i.e., which behaviors might have similar

elicitors or achieve similar goals), a six-factor structure also seems plausible, with behaviors that reflect caring about others (help, sympathy), condemnation of others (negative gossip, blame, criticism), praising others (gratitude, praise, neutral or positive gossip), humility vs. self-enhancement (bragging, condescension, insecurity, apologizing), general positivity vs. negativity (optimism, complaining, pessimism), and humor (funny, teasing, and sarcasm).

Major interpersonal traits. A further possibility is that the behaviors will separate into clusters that are consistent with the major interpersonal traits of extraversion, agreeableness, and honesty-humility (reviewed in Chapter 1). As I mentioned in Chapter 2, none of the behaviors seem strongly conceptually-related to extraversion. Helping, sympathy, praise, gratitude, and apologizing are most conceptually-related to high agreeableness, whereas bragging, teasing, condescension, blame, sarcasm, criticism, complaining, and negative gossip seem to be most conceptually-related to low agreeableness. Thus, these 13 behaviors might load onto one factor (with opposing loadings for the high vs. low agreeableness behaviors), or two factors (one factor that captures the tendency to actively support others and another that captures the tendency to restrain oneself from antagonistic behaviors). It is also possible that a narrower cluster of honesty-humility-related behaviors might emerge, characterized by less bragging, less condescension, and less sarcasm.

Interpersonal circumplex. Finally, it is possible to derive predictions based on the hypothesized positions of the behaviors on the IPC (see Figure 3 and Table 2, Chapter 2). As described in Chapter 1, the circular ordering of the IPC implies that a given behavior should be a) strongly positively correlated with behaviors that have a similar circumplex location, b) uncorrelated with behaviors that are 90° apart, and c) negatively correlated with behaviors that are 180° apart. Thus, the IPC would imply, for example, that bragging a) should be more

positively associated with condescension than with blame, b) relatively uncorrelated with complaining, and c) negatively correlated with gratitude (assuming that my hypothetical mapping of the IPC locations of the 18 behaviors is correct).

Depending on the strength of the intercorrelations, the number of factors extracted, and the rotation of the factor solutions, many different factor solutions are plausible. In a two-factor solution, it should be possible to recover a warmth and dominance factor. In a four-factor solution, we may see factors that capture behaviors that fall along the aloof–extraverted (NO–FG), cold–warm (DE–LM), arrogant–unassuming (BC–JK), and dominant–submissive (PA–HI) axes, respectively (with opposite factor loadings for each opposing octant). However, many additional factor solutions are plausible because the locations of the behaviors differ as a matter of degree.

Method and Results

Intercorrelations Among Interpersonal Behaviors

In Figure 4, I report the intercorrelations among the interpersonal behaviors in Year 1 and Year 2. A sensitivity power analysis showed that I had 80% power to detect correlations $\geq |.16|$ in Year 1, and $\geq |.23|$ in Year 2. The mean absolute r s show the average extent to which each interpersonal behavior was correlated with every other behavior. In Year 1, these average $|r|$ s ranged from .06 (gratitude) to .20 (criticism). The range of the average $|r|$ s was greater in Year 2, from .07 (apologizing) to .33 (criticism). Interestingly, a paired-samples t -test showed that the intercorrelations were stronger on average in Year 2 (mean $|r| = .21$), compared to Year 1 (mean $|r| = .13$), 95% CI of mean difference = [.06, .11], $p < .001$. This could suggest either that the codings in Year 2 were more valid (despite being less reliable) or that coders were making less fine-grained distinctions between different behaviors.

Figure 4. Heatmap of observed intercorrelations among interpersonal behaviors. Behaviors are ordered by their hypothesized IPC angular location (starting at 90°). Year 1 correlations are shown in the lower diagonal; Year 2 correlations are shown in the upper diagonal. $|r|_1$ and $|r|_2$ denote the average absolute correlations for each behavior in Year 1 and Year 2, respectively. Correlations $> .12|$ (and the negative association between pessimism and neutral or positive gossip) and $> .17|$ are significant at $p < .05$ in Year 1 and Year 2, respectively (not corrected for multiple comparisons).

Exploratory Factor Analysis

To explore the factor structure of these behaviors and the replicability of this factor structure across the two years, I used exploratory factor analysis using maximum likelihood and an orthogonal (varimax) rotation method. The goal of this procedure was to identify clusters of behaviors that would form meaningful and replicable composites. I used several criteria for evaluating the optimal number of factors to extract: (a) a visual inspection of the Scree plot, (b)

the MAP test, (c) parallel analysis, (d) the very simple structure criterion, and (f) meaningfulness (i.e., extracting the number of factors that seemed most interpretable). For the meaningfulness criterion, I extracted two- to seven-factor solutions for both Year 1 and Year 2 (reported in Supplemental Tables S1–S2). As shown in Table 6, the recommended number of factors differed across criteria and between the two years, but were generally in the range of 2–4 factors.

Table 6
Number of Factors Suggested by Different Criteria

Criterion	Number of Factors	
	Year 1	Year 2
Scree Plot	4 or 5	3 or 6
MAP Test	2	2
Parallel Analysis	5	4
Empirical BIC	4	3
Very Simple Structure	3	4
Meaningfulness	4	3

A four-factor solution seemed to be most interpretable in the Year 1 data (see Table 7). Based on behaviors that had substantial factor loadings ($\geq .40$), these factors appeared to capture behaviors related to *negativity* (complaining, pessimism, blame, criticism, and negative gossip), *mockery* (being funny, sarcasm, teasing, bragging), *supportiveness* (sympathy, optimism, and helping), and *gossip* (negative gossip and positive/neutral gossip). A three-factor solution seemed to be most interpretable in the Year 2 data (see Table 7). These factors appeared to capture *antagonistic behaviors* (criticism, pessimism, insecurity, teasing, bragging, complaining, sarcasm, condescension, and blame), *positivity* (sympathy, optimism, helping, praise, and being funny), and *gossip* (negative gossip and neutral or positive gossip). In a four-factor solution, a further *respectfulness* factor was interpretable, but the behaviors had weaker factor loadings (tease = $-.47$, sarcasm = $-.33$, gratitude = $.39$, apologizing = $.27$).

Table 7

Standardized Factor Loadings for Most Interpretable Factor Solutions in Year 1 and Year 2

Behavior	Year 1				Year 2		
	F1	F2	F3	F4	F1	F2	F3
Gratitude	-.10	-.05	.21	.04	-.09	.28	.06
Praise/compliment	.08	.32	.30	.13	.04	.47	.27
Sympathy	.10	.02	.69	.04	.09	.80	.07
Help/favors/support	.01	.00	.42	-.03	-.10	.54	-.02
Apologize	.14	.04	.12	-.18	.03	.01	.09
Hope/optimism	.12	.13	.57	-.06	-.02	.76	-.02
Funny	.05	.66	.13	.00	.06	.42	.09
Negative gossip	.50	.14	.02	.83	.09	.12	.85
Neutral/positive gossip	-.07	.02	.11	.42	-.01	.22	.49
Sarcasm	.08	.64	-.04	-.04	.57	.22	.04
Brag	.01	.43	.02	.02	.65	.06	.14
Condescension	.25	.33	-.17	.15	.54	-.04	.26
Complain	.68	.10	-.01	.06	.61	-.22	.11
Criticism	.58	.30	-.07	.29	.81	-.02	.33
Blame	.65	-.01	.10	-.04	.40	-.12	.31
Pessimism	.65	.03	.01	-.12	.80	-.04	-.12
Tease	.04	.64	.02	-.02	.72	.11	.00
Insecurity/self-doubt	.21	.02	.12	-.15	.77	-.23	-.09

Note. Loadings $\geq |.40|$ are shown in **boldface**.

Comparing the two solutions, the negativity factor in Year 1 is most similar to the antagonism factor in Year 2, except that sarcasm, teasing, and bragging are factored into a separate mockery factor (along with being funny) in Year 1. The positivity factor in Year 2 is most similar to the supportiveness factor in Year 1, except that being funny loads more strongly onto the mockery factor in Year 1. The gossip factor appears to be similar in both Year 1 and Year 2.

Confirmatory Factor Analysis

Based on the above explorations, I decided to conduct a confirmatory factor analysis (CFA) of a four-factor solution that represented a blend of the above solutions (see Figure 5). The four factors comprised *negativity* (pessimism, insecurity, criticism, complaining, condescension, and blame), *supportiveness* (sympathy, optimism, helping, and praise), *mockery*

(being funny, bragging, sarcasm, and teasing), and *gossip* (negative gossip and neutral or positive gossip).

Figure 5. Initial confirmatory factor analysis model (Model 1). Dashed factor loadings depict the items that were dropped from the final model (Model 2).

Model fit and reliability estimates. The model fit indices and reliability estimates for this initial model (Model 1) are shown in Table 8. Because the CFA was informed by EFAs conducted on the same datasets, these fit statistics are overestimates and should be interpreted with caution. In any case, all five indices suggested poor model fit. Across both years, the chi-square tests rejected the null hypothesis that the model fit the data. Comparative Fit Index and Tucker-Lewis Index scores were below conventional cutoffs of .90 (Hu & Bentler, 1999), and the Root Mean Square Error of Approximation and Standardized Root Mean Square Residual

were above conventional cutoffs of .08 (Hu & Bentler, 1999). Below, I explore possible sources of misfit and whether dropping items (Model 2) improves the fit statistics.

Table 8

CFA Fit Statistics and Omega (ω) Reliability Estimates for the Factors

Fit Statistic	Model 1		Model 2	
	Year 1	Year 2	Year 1	Year 2
$\chi^2(df), p$	331.9 (99), $p < .001$	270.24 (99), $p < .001$	271.32 (72), $p < .001$	163.97 (72), $p < .001$
CFI	.740	.794	.737	.862
TLI	.685	.751	.668	.825
RMSEA	.091	.111	.098	.095
SRMR	.082	.097	.084	.078
ω				
Negativity	.65	.82	.68	.76
Supportiveness	.62	.79	.62	.79
Mockery	.70	.48	.60	.73
Gossip	.47	.52	.47	.52

Note. CFI = Comparative Fit Index, TLI = Tucker-Lewis Index, RMSEA = Root Mean Square Error of Approximation, SRMR = Standardized Root Mean Square Residual. Model 1 = model depicted in Figure 5. Model 2 = dropping the insecurity and funny items.

Comparing factor loadings. As shown in Table 9, the loadings for the gossip factor were nearly identical across the two years. The pattern of loadings for the support factor was also similar across both years, but stronger in Year 2. The biggest discrepancies between the two years were that insecurity did not load onto the negativity factor in Year 1, and being funny did not load onto the mockery factor in Year 2. Thus, one possible source of model misfit is that one item in each year did not load onto the factor specified by the model.

Table 9

Confirmatory Factor Analysis Factor Loadings and Comparisons

Factor	Model 1												Model 2		
	All Participants		Same Participants		Same Protocol		Dif. Year 1 – Year 2		Dif. All – Same Participants		Dif. All – Same Protocol		Year 1	Year 2	Dif. Year 1 – Year 2
	Year 1	Year 2	Year 1	Year 2	Year 1		All	Same	Year 1	Year 2	Year 1	Year 2			
Negativity															
Pessimism	.48	.74	.55	.70	.52		.26	.15	.07	.03	.04	.22	.46	.67	.21
Insecurity	.08	.74	.08	.71	.40		.66	.64	.00	.03	.32	.35			
Criticism	.77	.85	.53	.84	.74		.08	.30	.24	.01	.03	.11	.79	.87	.07
Complain	.61	.65	.82	.66	.59		.04	.16	.21	.01	.02	.06	.59	.63	.04
Condescension	.41	.59	.34	.58	.46		.18	.24	.07	.01	.05	.13	.42	.62	.20
Blame	.56	.47	.62	.56	.47		.09	.06	.06	.08	.09	.00	.54	.50	.04
Supportive															
Sympathy	.69	.86	.72	.88	.95		.17	.15	.03	.01	.26	.09	.71	.87	.16
Hope/optimism	.62	.75	.44	.72	.38		.14	.28	.17	.03	.23	.37	.61	.75	.14
Help	.41	.50	.51	.45	.68		.08	.06	.09	.05	.27	.18	.42	.50	.08
Praise/compliment	.32	.44	.31	.43	.18		.12	.13	.01	.01	.14	.27	.30	.44	.13
Mockery															
Funny	.70	.16	.71	.16	.31		.54	.55	.01	.00	.39	.15			
Brag	.38	.65	.14	.61	.51		.27	.47	.24	.04	.13	.14	.38	.66	.28
Sarcasm	.60	.65	.64	.69	.60		.05	.05	.04	.05	.00	.04	.60	.64	.04
Tease	.67	.78	.64	.78	.45		.12	.14	.03	.00	.22	.34	.66	.77	.10
Gossip															
Negative gossip	.98	.95	.84	.91	.93		.03	.07	.14	.04	.05	.02	.98	.96	.03
Neu./pos. gossip	.43	.47	.43	.44	.46		.04	.01	.00	.03	.03	.00	.43	.47	.04

Note. All participants = all participants who had at least 20 speech-containing files in the respective year ($N_{\text{Year1}} = 287$, $N_{\text{Year2}} = 141$); Same participants = the 124 participants who had at least 20 speech-containing files in both years; Same Protocol = only the Year 1 files ($N = 282$) that were generated by the same coding protocol as Year 2. Dif. Year 1 – Year 2 = the absolute difference in the factor loadings between Year 1 and Year 2 for all participants vs. the same participants; Dif. All – Same Participants = the absolute difference in the factor loadings between all participants or the same participants for Year 1 or Year 2; Dif. All – Same Protocol = the absolute difference in the factor loadings between the factor loadings based on only the data in Year 1 that were generated by the same coding protocol as Year 2 and all data from either Year 1 or Year 2.

To examine whether these factor loadings were significantly different across the two years, I conducted a test of metric (i.e., weak) measurement invariance using multigroup confirmatory factor analysis. Metric invariance holds if each behavior contributes to the latent construct to a similar degree across both years (i.e., factor loadings are equal; Meredith, 1993). In contrast, metric noninvariance would imply that the meaning of the constructs changed between the two years. I first estimated a model in which all parameters were freely estimated for each year, and compared the fit of this model with a model in which the factor loadings were constrained to be equal across both years. The model in which factor loadings were constrained to be equal across both years ($\chi^2[213] = 741.5, p < .001$, CFI = .694, TLI = .655, RMSEA = .108, SRMR = .120) fit significantly worse than the model in which loadings were free to vary for each year ($\chi^2[198] = 601.93, p < .001$; CFI = .766, TLI = .716, RMSEA = .098, SRMR = .088), $\Delta\chi^2(15) = 139.57, p < .001$. In other words, the factor loadings were significantly different between the two years.

Exploring sources of discrepancies. Next, I explored two potential sources of the discrepancies between the factor loadings in Year 1 and Year 2: 1) differences in participants and 2) differences in the coding procedure.

Differences in participants. The Year 1 analyses included a larger set of participants ($N = 287$) than the Year 2 analyses ($N = 141$). To examine whether the discrepancies could be accounted for by differences in participants, I re-ran the tests of weak measurement invariance on the subset of participants who had sufficient data in both years ($N = 124$), separately for each year. Once again, the model in which factor loadings were constrained to be equal across both years ($\chi^2[213] = 559.92, p < .001$; CFI = .669, TLI = .628, RMSEA = .115, SRMR = .155) fit significantly worse than the model in which loadings were free to vary for each year ($\chi^2[198] =$

447.20, $p < .001$; CFI = .763, TLI = .712, RMSEA = .101, SRMR = .096), $\Delta\chi^2(15) = 112.71$, $p < .001$).

I also used two additional methods to compare the similarity of the vectors of factor loadings: 1) correlations to compare the rank-ordering of the factor loadings, and 2) the average absolute difference in the factor loadings. The factor loadings based on all participants in Year 1 were positively correlated with those based on all participants in Year 2, $r = .69$, 95% CI [.29, .88], $p = .003$. However, after restricting the analyses to the same participants in both years, these vectors of factor loadings were not significantly correlated, $r = .36$, 95% CI [-.17, .73], $p = .173$. The factor loadings based on all vs. the same participants were also positively correlated in Year 1, $r = .73$, 95% CI [.38, .90], $p = .001$, as well as in Year 2, $r = .98$, 95% CI [.93, .99], $p < .001$. The average absolute difference in the factor loadings between Year 1 and Year 2 was .19 for all participants and .23 for the same participants. The average absolute difference in the factor loadings between all vs. the same participants was .09 in Year 1 and .03 in Year 2.

Overall, these results suggest that the size and rank-order of the factor loadings are more similar between subsets of participants (all vs. the same) than between the two years. This suggests that the differences in the pattern of factor loadings between Year 1 and Year 2 are unlikely to be due to differences in the participants included in each analysis.

Differences in coding procedures. As described in Chapter 2, the Year 1 files were coded in three phases, each with slightly different coding protocols. Only the last phase used the same coding protocol as the Year 2 files (elaborated item descriptions with separate surveys for the positive and negative behaviors). Thus, I re-ran the tests of weak measurement invariance, but only included Year 1 data that were generated using the same coding protocol as the Year 2 files. Note that this only included files that were recorded outside of the four ESM-matched

hours (except for one participant). After applying the same inclusion criteria of a minimum of 20 speech-containing files, this Year 1 subset comprised 282 participants.

The resulting factor loadings are shown in Table 9. With the coding protocol held constant, insecurity loaded onto the negativity factor (though to a lower degree than in Year 2, descriptively), and being funny loaded to a lower degree onto the mockery factor (though still to a higher degree than in Year 2, descriptively).

Next, I re-ran the tests of metric measurement invariance on a model fit to the data generated from the same coding procedure in Year 1, comparing this with a model fit to all data in Year 2. The model in which factor loadings were constrained to be equal across both years ($\chi^2[213] = 787.694, p < .001, CFI = .675, TLI = .634, RMSEA = .113, SRMR = .124$) still fit significantly worse than the model in which loadings were free to vary for each year ($\chi^2[198] = 614.681, p < .001, CFI = .765, TLI = .715, RMSEA = .10, SRMR = .088, \Delta\chi^2(15) = 173.01, p < .001$). Thus, even when the coding protocol was identical, the factor loadings were still significantly different between the two years.

As above, I used correlations and average absolute differences to further compare the similarity of the vectors of factor loadings. The factor loadings based on the final coding protocol in Year 1 were positively correlated with those based on all data in Year 1, $r = .67, 95\% CI [.26, .87], p = .005$, as well as Year 2, $r = .75, 95\% CI [.40, .91], p < .001$. The average absolute difference in the factor loadings based on the final coding protocol in Year 1 and the factor loadings based on all data was .15 for Year 1 and .16 for Year 2. In other words, holding the coding protocol constant results in factor loadings that are about halfway between the solutions from Year 1 (all data) and Year 2.

Overall, these results suggest that the overall discrepancies in the factor loadings between Year 1 and Year 2 can be partially (but not completely) explained by differences in the coding protocol. This is likely because coding for 9 (instead of all 18) behaviors at a time made the task easier for the coders, with a likely ensuing improvement in data quality. Nevertheless, because the data from the earlier coding protocol likely still have some signal (even if the data are noisier), I will include all data in the final analyses.

Removing non-invariant items. Finally, I fit an alternative model (Model 2) in which I dropped two items: insecurity (which did not load onto the negativity factor in Year 1) and being funny (which did not load onto the mockery factor in Year 2). Fit statistics and reliability estimates are reported in Table 8, and factor loadings are reported in Table 9. The average absolute difference in the factor loadings between Year 1 and Year 2 was .11, which was descriptively lower than for Model 1 (.19). These vectors of factor loadings were also descriptively more strongly correlated than in Model 1, $r = .90$, 95% CI [.71, .97], $p < .001$. The fit statistics were still poor overall, but generally reflected better fit compared to Model 1, at least for the Year 2 data. The reliability estimates were similar for the negativity, supportiveness, and gossip factors, but somewhat different for mockery (slightly lower in Year 1 and substantially higher in Year 2).

I re-ran the tests of metric measurement invariance after dropping these items. The model in which factor loadings were constrained to be equal across both years ($\chi^2[157] = 495.45$, $p < .001$; CFI = .762, TLI = .724, RMSEA = .100, SRMR = .103) still fit significantly worse than the model in which loadings were free to vary for each year ($\chi^2[144] = 435.05$, $p < .001$; CFI = .795, TLI = .741, RMSEA = .097, SRMR = .083), $\Delta\chi^2(13) = 60.41$, $p < .001$. Thus, even after dropping the two noninvariant items, the factor loadings were still significantly different between

the two years, but the between-year discrepancies in the fit statistics were less dramatic than for Model 1. Thus, for the subsequent rank-order and between-person correlate analyses (Chapters 4 and 5) involving these factors, I will use the solution from Model 2.

Intercorrelations among factors. Table 10 shows the latent intercorrelations among the four factors in Model 1 and Model 2. The intercorrelations were almost identical between Model 1 and Model 2, so I focus here on discussing the differences in effect sizes between the two years, for Model 2. The negativity and mockery factors were positively associated in both years, but the strength of these correlations suggest that negativity and mockery captured distinct constructs in Year 1 ($r = .39$), but perhaps not in Year 2 ($r = .92$). The negativity and gossip factors were also positively associated in both years, but to a much larger extent in Year 1 ($r = .67$) than in Year 2 ($.34$). Finally, there were some weaker and less consistent positive associations between mockery and gossip (Year 1 only; $r = .16$) and between supportiveness and gossip (Year 2 only; $r = .20$). All in all, despite the discrepancies in model fit between Year 1 and Year 2, the detectable pattern of intercorrelations among the factors was quite consistent.

Table 10
Latent Inter-Correlations Among the Factors

	Model 1				Model 2			
	1.	2.	3.	4.	1.	2.	3.	4.
1. Negativity		-.06	.89***	.24*		-.01	.92***	.34***
2. Supportiveness	.16		.22*	.20*	.15		.21*	.20*
3. Mockery	.33***	.21*		.15	.39***	.15		.15
4. Gossip	.66***	.13	.17*		.67***	.13	.16*	

Note. Year 1 correlations are shown in the lower diagonal; Year 2 correlations are shown in the upper diagonals. * $p < .05$, ** $p < .01$, *** $p < .001$ (not corrected for multiple comparisons).

Discussion

Personality psychology has a long tradition of discovering the factor structure of self-reported personality traits (e.g., Ashton et al., 2004; De Raad, Perugini, Hrebickova, & Szarota,

1998; Saucier & Goldberg, 2001). However, relatively few studies have attempted to uncover factor structures based on behavioral observations (cf. Kulkarni et al., 2018; Leising & Bleidorn, 2011). In this chapter, I examined the factor structure of 18 interpersonal behaviors using a combination of exploratory and confirmatory factor analyses. The best-fitting factor solution involved four factors: negativity (pessimism, criticism, complaining, condescension, and blame), supportiveness (sympathy, optimism, helping, and praise), mockery (bragging, sarcasm, and teasing), and gossip (negative and neutral or positive). Four behaviors (insecurity, being funny, gratitude, and apologizing) did not seem to fit within this four-factor structure. Here, I discuss the results with respect to the two purposes of this chapter: 1) to derive reliable behavioral composites for use in subsequent analyses, and 2) to generate hypotheses about the common interpersonal meanings of specific behaviors. I also consider the implications of these results for the next two empirical chapters.

The primary goal of this chapter was to derive reliable behavioral composites in order to supplement the analyses of the rank-order stability and correlates of individual behaviors with analyses of behavioral composites. To the extent that reliable behavioral composites can be identified, the effect sizes for analyses involving composites should be stronger than those involving individual behaviors. The negativity, supportiveness, and mockery composites showed reasonable levels of reliability in Year 1 ($\omega_s > .60$), and higher levels of reliability in Year 2 ($\omega_s > .73$). However, the gossip factor was slightly less reliable ($\omega_s \sim .50$), probably because there were only two indicators. Still, this shows that multiple behaviors can, with adequate levels of reliability, differentiate participants' general behavioral tendencies towards negativity, supportiveness, mockery, and gossip. This provides a reasonable basis for examining the rank-

order stability and correlates of behavioral composites in Chapters 4 and 5, in which I will return to the issue of whether effect sizes are stronger for composites than for individual behaviors.

The second goal of this chapter was to generate hypotheses about the shared interpersonal meanings that underlie different behaviors. I considered the possibility that the behaviors would split into 1) a positive and negative cluster, 2) an intuitive six-factor structure (caring, other-condemning, other-praising, humility vs. self-enhancement, general positivity vs. negativity, and humor), or 3) clusters based on IPC locations. The resulting four-factor structure was somewhat different from each of these possibilities.

As shown in Figure 6, the behaviors in the supportiveness factor tended to reflect a combination of warmth and moderate–high dominance. These behaviors seem to be unified by a common function of supporting and encouraging others, and making others feel good. They also seem to be most conceptually-related to high levels of Big Five agreeableness. The behaviors in the negativity factor reflected low warmth across the full range of dominance, and seem to capture a blend of condemning others (condescension, blame, criticism) as well as general negativity (complaining and pessimism). These behaviors seem to be reflective of low Big Five and HEXACO conceptualizations of agreeableness.

Notably, the supportiveness and negativity factors were relatively uncorrelated, which is consistent with the perspective that prosocial and antisocial traits are separable (Krueger, Hicks, & McGue, 2001). The behaviors in the mockery factor tended to reflect a combination of low warmth and moderate–high dominance, and perhaps have the common consequence of making other people feel worse about themselves in the service of self-enhancement and humor. This mockery factor may also reflect low Big Five agreeableness, but seems more specifically related to low honesty-humility. Finally, gossip captured the general tendency to talk about others,

regardless of the warmth of these acts. Gossip is not conceptually central to extraversion, agreeableness, or honesty-humility.

Figure 6. Hypothesized mapping of the 18 interpersonal behaviors onto the interpersonal circumplex, color-coded based on the factors.

To the extent that these behaviors do capture common interpersonal meanings, it raises the possibility that personality traits might have similar associations with behaviors in the same factor. For example, trait agreeableness may be associated with each of the behaviors in the supportiveness cluster. Similarly, to the extent that behaviors in the same cluster have similar interpersonal impacts (e.g., making others feel supported vs. derogated), they may also have similar interpersonal consequences. For example, it is possible that bragging, teasing, and sarcasm would each be related to being disliked. These results therefore generate some general

hypotheses about how traits may be related to clusters of behaviors, and how clusters of behaviors may be related to interpersonal outcomes (the focus of Chapter 5).

Far from being definitive, this factor structure should be considered highly tentative for several reasons. First, the 18 behaviors do not represent a comprehensive sampling of the interpersonal space (i.e., agreeableness-related behaviors are overrepresented and extraversion- and honesty-humility-related behaviors are underrepresented). Thus, this factor structure should not be interpreted as the true structure of interpersonal traits in general. Second, I approached these analyses in an exploratory way. The resulting factor structure would need to be confirmed in future replication studies.

Moreover, the fit statistics for the best-fitting CFA solution were still poor by conventional standards (Hu & Bentler, 1999). One source of poor model fit is the low reliabilities of the indicators, stemming from inter-coder disagreement on which participants expressed higher or lower frequencies of each behavior. Because each behavioral indicator contains noise due to measurement error, this attenuates the loadings of each behavior onto its respective factor. A second source of poor model fit could be that the factor structure seemed to change between Year 1 and Year 2. Thus, it was difficult to find a factor solution that would fit the data well in both years. The discrepancies in factor loadings between the two years could not be accounted for by differences in the subset of participants who were included in the Year 1 and Year 2 analyses. These discrepancies did seem to be partially accounted for by differences in the coding procedure (i.e., coding all 18 behaviors vs. only 9 at a time), but this did not completely explain the discrepancies. This raises the alarming possibility that coders who coded the Year 1 files (153 coders) had different concepts of these behaviors, compared to coders who coded the Year 2 files (68 coders, including 63 who had coded the files in Year 1).

Chapter 4: Rank-Order Stability Across One Year

Do people show stable individual differences in what they do to each other, or are they wildly inconsistent in how they act? In this chapter, I address one of the most central issues in personality psychology—behavioral consistency—by examining the year-to-year rank-order stability of the 18 individual behaviors as well the four factors uncovered in Chapter 3. The first, narrower purpose of this chapter is to better understand the psychometric properties of EAR-coded interpersonal behaviors; specifically, the extent to which they capture individual differences that are stable across time. The second, broader purpose is to consider the implications for the person–situation debate.

Defining Behavioral Consistency

Demonstrating behavioral consistency has been key to refuting situationist critiques and providing support for the existence of personality. It would not make much sense, for example, to say that “Mary is boastful” if she showed no consistency whatsoever in the extent to which she bragged across different situations and over time, compared to other people. However, behavioral consistency does not imply that personality is “set like plaster” (James, 1890; i.e., does not change over time). Mean-level change (e.g., people tend to become more agreeable across the lifespan) is consistent with *rank-order stability* (e.g., people who are more agreeable than others at age 18 tend to be more agreeable than others at age 40; Roberts, Wood, & Caspi, 2008).

Nor does behavioral consistency imply that people behave in the same way in every situation. People do act differently to adapt to situational demands (e.g., people talk more when interacting with a submissive partner than a dominant partner; Leikas et al., 2012), but demonstrate rank-order stability across situations (e.g., people who talk more than others when

interacting with a submissive partner still tend to talk more than others when interacting with a dominant partner; Leikas et al., 2012). Moreover, although single behavioral acts observed at one time point only weakly predict single behavioral acts at a subsequent time point (Fleeson & Gallagher, 2009), when many behavioral instances are averaged together across an extended period of time (e.g., one week), these behavioral aggregates should strongly predict behavioral aggregates at a later time point (Fleeson & Gallagher, 2009). This is the type of behavioral consistency that I focus on in this chapter.

High rank-order stability across one year would mean that people who expressed a given behavior (e.g., sympathy) more frequently across one week (compared to other people) were also more likely to express those behaviors more frequently across another week, one year later. A rank-order stability correlation of 0 would mean that a person's relative standing on a given behavior across one week provides no information on their relative standing on that behavior one year later. Analyses of rank-order stability thus provide a strong test of the extent to which observed between-person differences in interpersonal behaviors reflect stable individual differences (vs. situational factors and measurement error).

Psychometric Properties of EAR-Coded Interpersonal Behaviors

As mentioned in Chapter 2, behavioral codings based on the EAR are not necessarily “objective”. Whereas some micro-level behaviors can be coded in a fairly objective way (e.g., whether or not a participant talked or laughed), macro-level social behaviors (e.g., whether a participant bragged or expressed condescension) require a greater amount of inference. Thus, observed between-person variability in EAR-coded interpersonal behaviors will capture a blend of true individual differences, measurement error due to inter-coder disagreement, and other sources of error. In Chapter 2 (see Table 5), I reported the inter-coder reliability estimates of the

behaviors at each year. Rank-order stability provides a complementary metric of test-retest reliability.

To the extent that EAR-coded interpersonal behaviors are stable from one year to the next, I can be more confident that these codings reflect meaningful individual differences (vs. measurement error). This will provide additional context for interpreting the correlational analyses in Chapter 5. For example, null effects will be more informative to the extent that the behavioral codings adequately captured true individual differences in the first place.

As mentioned in Chapter 1, Bollich and colleagues (2016) showed that individual differences in interpersonal behaviors were moderately stable (mean $r = .47$ over 4 weeks, mean $r = .52$ over 10 weeks). Across 10 weeks, the most stable individual differences were in apologizing, helping, and complaining (r s $> .78$), followed by affection (not measured in the present dissertation), optimism, sympathy, condescension, and pessimism (r s between $.47$ to $.58$), and praise, blame, and sarcasm (r s between $.25$ to $.34$). Individual differences in gratitude and bragging had stability correlations close to 0. Because of the small sample size ($N = 73$), Bollich and colleagues' estimates were quite imprecise. However, because rank-order stability decreases as the retest interval increases (Roberts & DelVecchio, 2000), these estimates provide a sense of the upper bound on the stability correlations that might be observed across one year.

Here, I build on Bollich and colleagues (2016) in three ways, by testing rank-order stability 1) using a relatively larger sample ($N = 124$, compared to $N = 11$ and $N = 73$ in Bollich et al., 2016), 2) including additional interpersonal behaviors beyond those examined in Bollich and colleagues (being funny, negative gossip, neutral or positive gossip, teasing, and expressions of insecurity/self-doubt), and 3) across a much longer retest period—one year. More generally, to my knowledge, no other studies have examined the rank-order stability of EAR-coded everyday

behaviors across such a long retest period. Thus, the results of this chapter will provide future EAR researchers with useful information about the psychometric properties of EAR-coded interpersonal behaviors.

Implications for the Person–Situation Debate

More broadly, temporal stability analyses can shed light on deeper questions about the relative impact of persons and situations on different behaviors. If people act in a completely different way from one year to the next (assuming that the behaviors are reliably measured), this would suggest that situations have a greater influence on interpersonal behavior than do stable traits. If high levels of temporal stability exist, there are two possible explanations. The person explanation is that personality exerts a strong influence on interpersonal behavior. The situationist explanation would be that situations are still driving behavior, and people behave in similar ways over time because they end up in similar situations. In other words, temporal stability is a necessary but insufficient condition for demonstrating the influence of personality on behavior. Although the current analyses cannot distinguish between these two possibilities, rank-order stability analyses can at least reveal the extent to which a person's future interpersonal behavior can be predicted by how they treated others in the past.

Method and Results

For the four factors discovered in Chapter 3, I report latent correlations based on the models depicted in Figure 7 (Panel A), as well as zero-order correlations based on observed composites. I also attempted to fit the latent models depicted in Figure 7 (Panel B) to estimate the rank-order stability of the individual behaviors at a latent level. However, due to the very low inter-rater reliability for several of the behaviors, many of the models did not converge, and the

models that did give implausibly high estimates (e.g., $r > .80$). Thus, below, I only report zero-order correlations based on observed composites for the individual behaviors.

Figure 7. Rank-order stability models for factors (Panel A) and specific behaviors (Panel B). y_1 and y_2 denote a factor (Panel A) or specific behavior (Panel B) measured in Year 1 or Year 2, respectively. $r_{y1,y2}$ denotes the latent rank-order stability correlation. Factor loadings (λ) and residual variances (ϵ) with the same subscripts were constrained to equality. “Matched” = aggregates based on files recorded during ESM-matched hours. “Unmatched” = aggregates based on files recorded outside of ESM-matched hours. n denotes the number of behaviors used to measure that construct.

To estimate the rank-order stability of the factors at a latent level, I specified a model in which each behavior was an indicator of the latent construct measured at Year 1 or Year 2 (see Figure 7, Panel A). Factor loadings for each behavior were constrained to be equal across the two years, to ensure that the meaning of the latent factor would be equivalent across both years. The latent models for the factors (see Figure 7, Panel A) were estimated using the R package *lavaan* (Rosseel, 2012), with maximum likelihood estimation. I also planned to estimate latent models for the individual behaviors (see Figure 7, Panel B), but faced convergence difficulties that produced implausibly high estimates (probably due to the very low reliability of some of the indicators in Year 2). Thus, I only report the zero-order correlations for the rank-order stability of the individual behaviors.

Measurement Invariance for Each Factor

Prior to estimating the rank-order stability of the factors, I tested whether metric measurement invariance held for each factor. Although metric measurement invariance did not hold for the full model (as described in Chapter 3), for the purposes of assessing the rank-order stability of each factor, all that is required is that invariance holds for each factor of interest. Thus, I compared the fit of a multi-group CFA model in which all loadings were free to vary between the two years with the fit of a model in which the loadings for a given factor (e.g., negativity) were constrained to equality across both years. These analyses only included the subset of 124 participants who had sufficient data in both years and were therefore included in the stability analyses. As shown in Table 11, although the χ^2 difference tests were significant for negativity, supportiveness, and mockery, the other indices generally reflected only modest worsening of fit when loadings for a given factor were constrained to equality across both years. Thus, we can be reasonably confident that the items for each factor are contributing to the

measurement of a given construct in a similar way across both years, providing a basis for interpreting the rank-order stability correlations.

Table 11

Fit Statistics for Measurement Invariance Tests on Each Factor

Factor	$\chi^2(df), p$	CFI	TLI	RMSEA	SRMR
Free model	320.92 (144), $p < .001$	.795	.741	.100	.087
Negativity					
Constrained	344.8 (149), $p < .001$	.773	.723	.103	.110
Δ Fit	23.87 (5), $p < .001$	.022	.018	-.003	-.023
Supportiveness					
Constrained	345.44 (148), $p < .001$	.771	.719	.104	.097
Δ Fit	24.52 (4), $p < .001$	.024	.022	-.004	-.010
Mockery					
Constrained	345.12 (147), $p < .001$	.771	.716	.104	.105
Δ Fit	24.2 (3), $p < .001$	.025	.025	-.005	-.018
Gossip					
Constrained	320.92 (145), $p < .001$	.796	.744	.099	.087
Δ Fit	0 (1), $p = .998$	-.001	-.003	.001	.000

Note. These analyses only included the 124 participants who had sufficient data in both years and were thus included in the stability analyses. Free model = fit statistics for a model in which all loadings were free to vary between the two years; Constrained = fit statistics when loadings for the target factor were constrained to equality across both years; Δ Fit = change in fit statistics between free and constrained models.

Rank-Order Stability Estimates

Overall, out of 18 behaviors, 15 point estimates were positive, and 7 were statistically detectable under an uncorrected $p < .05$ threshold (see Table 12). The point estimates for helping, sarcasm, teasing, and neutral or positive gossip were between .26–.28, whereas the point estimates for negative gossip, apologizing, and condescension were slightly lower at .18–.21. The point estimates for criticism, complaining, insecurity, and gratitude were in the .16–.17 range, with 95% confidence intervals that just captured zero. In order to have 80% power to detect correlations in the .16–.17 range, I would have needed 303 participants. Thus, these four null effects could be due to low power. Finally, the point estimates for the rank-order stability of

the remaining behaviors were all $\leq .12$ (optimism, praising, bragging, being funny) or even negative (pessimism, sympathy, and blame).

Table 12

Year-to-Year Stability Correlations for Individual Behaviors and Latent Factors

Composite or Behavior	Latent			Observed		
	<i>r</i>	95% CI	<i>p</i>	<i>r</i>	95% CI	<i>p</i>
Negativity	.06	[−.17, .28]	.626	.09	[−.09, .26]	.310
Pessimism				−.12	[−.29, .05]	.169
Criticism				.17	[−.01, .34]	.060
Complain				.17	[−.01, .34]	.057
Condescension				.21	[.04, .37]	.018
Blame				−.06	[−.23, .12]	.513
Supportiveness	−.05	[−.31, .21]	.696	.08	[−.10, .25]	.388
Sympathy				−.10	[−.27, .08]	.260
Hope/optimism				.12	[−.06, .29]	.198
Help/favors/support				.27	[.10, .42]	.003
Praise/compliment				.04	[−.14, .22]	.651
Mockery	.38	[.15, .61]	.001	.28	[.11, .43]	.002
Brag				.11	[−.06, .28]	.212
Sarcasm				.26	[.09, .42]	.004
Tease				.27	[.10, .42]	.003
Gossip	.28	[.02, .54]	.036	.26	[.09, .42]	.003
Negative gossip				.18	[+.00, .34]	.049
Neutral/positive gossip				.28	[.11, .43]	.002
No factor						
Insecurity/self-doubt				.17	[−.01, .33]	.066
Funny				.09	[−.09, .26]	.320
Gratitude				.16	[−.02, .32]	.083
Apologize				.21	[.03, .37]	.022

Note. $N = 124$. Latent = latent correlations; Observed = zero-order correlations with observed composites. Estimates for which 95% CIs did not capture 0 are shown in **boldface**. Indicators for factors were individual behaviors (averaged across coders).

Turning to the behavioral composites, the mockery and gossip composites showed some behavioral stability ($r_s > .28$) from one year to the next, suggesting that they captured reliable and broader individual differences. These composites showed descriptively greater stability than the average stability of the individual behaviors (.28 vs. .22 for mockery; .26 vs. .23 for gossip)—but not by that much. Moreover, there was no evidence of stability in the negativity and

supportiveness factors (r s .08–.09), and these composites showed very similar levels of stability compared to the average stability of the individual behaviors (.09 vs. .08 for negativity, .08 vs. .08 for supportiveness). Finally, the latent correlation was somewhat higher than the observed correlation for mockery, but not for any of the other composites. Thus, overall, these results were somewhat inconsistent with the psychometric principle that composites should be more reliable than individual behaviors. This may suggest that the factors lack validity.

Discussion

In summary, these results revealed a mixed picture of the rank-order stability of EAR-coded interpersonal behaviors. As would be expected from a longer retest period, the average level of rank-order stability in this study (.13) was much lower than Bollich and colleagues' (2016) estimate of an average r of .52 across 10 weeks. In the best case, stability estimates were in the .26–.28 range (helping, sarcasm, teasing, and neutral or positive gossip); in the worst case, stability estimates were close to zero or negative (pessimism, blame, sympathy, and praise). Here, I discuss the implications of these relatively modest levels of rank-order stability for interpreting the substantive analyses in Chapter 5, and for the person–situation debate.

There are several possible reasons for these relatively modest levels of rank-order stability. Methodologically, the low levels of rank-order stability likely reflect the low reliability with which the behaviors were coded (especially in Year 2). Low inter-coder reliability means that coders were only able to approximately distinguish between people who expressed higher or lower levels of a given behavior in a given year. Indeed, the correlation between the average inter-coder reliability (across both years) and the rank-order stability correlations was .41, 95% CI [–.07, .74], $p = .088$ ($N = 18$ behaviors). Second, given that the factor structure of the interpersonal behaviors was somewhat inconsistent between the two years (see Chapter 3), rank-

order stability may have been low because the coders in Year 1 interpreted each behavior in a different way than the coders in Year 2. A third possibility is that some behaviors are truly more determined by the situation than by the person. Thus, rank-order changes in interpersonal behaviors might reflect changing situational demands from one year to the next.

For different reasons, the first and third possibilities—poor measurement reliability and the predominance of situational influences on interpersonal behavior—would both imply that it may be more difficult to detect correlates of the negativity and supportiveness composites, and the individual behaviors of pessimism, blame, sympathy, optimism, praise, bragging, and being funny (see Chapter 5). The measurement explanation is that if the observed variation in interpersonal behaviors mostly reflects noise, it would be difficult to establish behavioral manifestations of personality traits. The situationist explanation is that if the behaviors were determined more by the situation than by the person, then personality traits should not be substantively important in predicting these behaviors. On the other hand, the mockery and gossip composites and the individual behaviors of helping, sarcasm, teasing, and neutral or positive gossip show moderate amounts of stability over time. Thus, it should be easier to detect correlates of these behaviors (assuming that the observed behavioral consistency is not simply due to being in similar situations; see below), and null effects should be more interpretable (i.e., unlikely to be primarily due to poor measurement).

Although there were moderate amounts of temporal stability in some of the behaviors, these findings do not directly refute situationist critiques of personality. A situationist would argue that temporal stability in behavior can be explained by being in similar situations from one year to the next. However, the stability of situations could of course be driven by personality traits (i.e., situation selection effects). In any case, a stronger test of behavioral consistency

would require the simultaneous assessment of situations and behavior over time. This would also address the question of the extent to which low rank-order stability is due to the influence of situational variability, rather than measurement error.

In summary, although low measurement reliability, modest statistical power, and the lack of situational measurement limits the interpretability of these rank-order stability estimates, these results suggest that individual differences in several everyday interpersonal behaviors (helping, sarcasm, teasing, neutral or positive gossip, negative gossip, apologizing, and condescension) are at least somewhat stable across one year. These findings complement studies that have demonstrated cross-situational consistency in interpersonal behavior in the lab (e.g., Leikas et al., 2012; Leising & Bleidorn, 2011) by showing that individual differences in interpersonal behavior also show some stability across time even in far less controlled settings (i.e., across everyday situations in the real world). This suggests that at least some EAR-coded interpersonal behaviors reflect meaningful individual differences—an issue I will continue to explore in Chapter 5, where I examine the correlates of these interpersonal behaviors.

Chapter 5: Correlates and Consequences of Everyday Interpersonal Behavior

So far in this dissertation, I have explored the psychometric properties of the set of 18 interpersonal behaviors. Now that we have learned something about the extent to which these behaviors reflect reliable individual differences and the possible psychological meanings of different interpersonal behaviors, it is time to connect them with personality traits and informant-reported social consequences. First, I explore the personality and gender correlates of each interpersonal behavior. For example, do agreeable people express sympathy more often? Do women apologize more than men? Next, I examine whether these interpersonal behaviors have implications for being liked and having good relationships with one's friends. For example, are thankful people liked more than the ungrateful? Below, I describe the specific goals of each of these research questions.

Potential Theoretical Contributions of Correlating Traits with Behaviors

The primary purpose of correlating traits with behaviors was to clarify the interpersonal behavioral implications of personality traits. The second purpose was to further clarify the psychological meaning of everyday interpersonal behaviors, and especially whether they reflect moral character (as assumed by Bollich et al., 2016). In other words, examining the correlations between personality traits and interpersonal behaviors provides the opportunity for both types of measures to validate each other (Cronbach & Meehl, 1955). Below, I describe these two purposes in greater detail and introduce the trait correlates of interest.

Clarifying the behavioral implications of personality trait measures. In Chapter 1, I argued that behavior is the ultimate outcome of interest and that behavior is central to the conceptualization of interpersonal traits (see Table 1). However, relatively little research has examined how interpersonal traits manifest in actual interpersonal behavior in everyday life (for

a review, see Chapter 2). Here, I will use the example of agreeableness to illustrate how measuring actual behavior can shed light on the construct validity of personality measures, address theoretical debates about the link between personality traits and behavior, and potentially reveal unexpected behavioral consequences of various constructs.

Measuring actual behavior can help establish the construct validity of personality trait measures and clarify whether traits are manifested in behavior in a decontextualized manner (e.g., McAdams & Pals, 2006). For example, if acting in prosocial ways and restraining oneself from antagonistic behaviors is key to the definition of agreeableness (see Table 1), self-report measures should be positively correlated with sympathetic and considerate behaviors such as sympathy, help, and gratitude, and negatively correlated with antagonistic behaviors such as criticism, blame, and condescension. To the extent that self-report measures of agreeableness are correlated with such behaviors, we can be more confident in the validity of these measures as capturing behavior-relevant traits. If traits are strongly correlated with behavior without considering contextual factors, this might also support conceptualizations of traits as being unconditional and decontextualized (McAdams & Pals, 2006). Such associations can also help us attach more concrete behavioral meanings to statements such as, “Michelle is highly agreeable” or “Simine is extremely introverted” (an issue that I explore further in Chapter 6).

If, however, self-reports of agreeableness are only weakly or even uncorrelated with conceptually-relevant behaviors, one possibility is that self-report measures only capture a person’s self-perceptions (i.e., identity) rather than what that person is actually like (i.e., their trait; McAbee & Connelly, 2016). Alternatively, it might imply that the construct of agreeableness should be conceptualized as capturing sympathetic affect and trusting cognitions, without any behavioral implications. This seems like a dissatisfying solution because a construct

that is “all in the head” would be far less interesting and important than one that has behavioral implications. A further possibility is that agreeableness only manifests behaviorally in the presence of relevant stimuli, rather than in general. This would support conceptualizations of personality traits as being context-dependent (e.g., Allport, 1937; Tellegen, 1981; DeYoung, 2015), and would suggest that existing definitions of agreeableness should be revised to specify the relevant conditions under which agreeable thoughts and feelings are manifested as behavior.

Measuring actual behavior can also shed light on the process by which people judge their own traits, and, ultimately, whether they know themselves (Vazire & Carlson, 2011). In other words, can we find behavioral evidence for a person’s claim that they are agreeable, at least among this set of 18 interpersonal behaviors? Self-perception theory posits that people observe their behavior and infer that they must have the kind of traits that those behaviors imply (Bem, 1972). If self-reports of agreeableness (for example) are correlated with behaviors such as sympathy and helping, this might imply that people consider how often they enact such behaviors when they judge their agreeableness. If, however, none of these behaviors are correlated with self-reports of agreeableness, this might suggest that people rely more heavily on other information sources (e.g., values, self-concepts, self-enhancement) or on motivated interpretations of their behaviors than on straightforward recollections of their actual behavior when judging their agreeableness (Robins & John, 1997). Each of these factors would reduce the amount of self-insight that people have into their agreeableness.

All of the above examples consider the implications of finding or failing to find behavioral correlates that are closely matched with existing conceptualizations of constructs. A further opportunity of exploring correlations between a variety of traits and behaviors—including those that seem less obviously theoretically matched—is that we may learn something

new or unexpected about different traits. For example, by exploring the associations between personality traits and the words that people used in their Facebook posts, Park and colleagues (2015) found some theoretically-obvious correlations (e.g., extraverts used more party-related words), but also others that were less obvious (e.g., introverts talked more about fixing computers). Similarly, in a study of the connections between everyday speech and momentary emotions, unsurprisingly, people reported feeling less happy in moments when they talked more about math and classes and tests, but less intuitively, they also reported feeling happier when they talked more about clothing (Sun et al., 2020). For this reason, in this chapter, I chose to include several trait correlates, including some that seemed less obviously relevant to interpersonal behavior (described below).

Clarifying the psychological meaning of everyday interpersonal behaviors. The psychological meaning of a specific behavior cannot be simply assumed; however, examining the personality correlates of behaviors can help clarify their psychological meaning. For example, dictator game allocations are often interpreted as being indicative of altruism (Carpenter, Connolly, & Myers, 2008), but previous work on the discriminant validity of the compassion and politeness aspects of agreeableness suggests that generous dictator game allocations reflect adherence to social norms regarding fairness (i.e., good manners), rather than emotional concern (Zhao, Ferguson, & Smillie, 2017).

In Chapter 3, I began to explore the common interpersonal meanings that might underlie specific behavioral manifestations. The results of Chapter 3 revealed four factors, which I interpreted as capturing common tendencies towards negativity (condescension, blame, criticism, complaining, pessimism), supportiveness (helping, sympathy, praise, and optimism), mockery (bragging, teasing, and sarcasm), and gossip (negative and positive or neutral). Correlating these

behavioral composites with traits can further help clarify their interpersonal meanings. For example, to the extent that the supportiveness-related behaviors are correlated with trait agreeableness, this would suggest that these behaviors are motivated by common prosocial tendencies. If mockery-related behaviors are more strongly related to disagreeableness than to extraversion, this would suggest that such behaviors typically have a more antagonistic (rather than dominant and playful) meaning. More generally, if traits are more strongly correlated with behavioral composites than with individual behaviors, this would further demonstrate the validity and utility of the composites discovered in Chapter 3.

Uncovering trait correlates can also help to clarify the meaning of individual behaviors. For example, if agreeable (rather than disagreeable) people are more likely to spread negative gossip in everyday life, this would suggest that negative gossip is typically motivated by prosocial rather than spiteful motives (Feinberg, Willer, Stellar, & Keltner, 2012). Furthermore, although some forms of gratitude expressions (e.g., writing thank you cards) might be valid indicators of agreeableness, everyday expressions of gratitude (e.g., thanking a cashier) might be relatively trivial and reflect social norms and situational expectations rather than prosocial traits.

A further goal is to test the assumption that the interpersonal behaviors assessed by Bollich and colleagues (2016) are, in fact, indicative of moral character. As mentioned in Chapter 2, Bollich and colleagues aimed to capture prototypically virtuous behaviors and concluded that “there are substantially stable individual differences in these moral behaviors” (p. 19). The interpretation of these behaviors as being “moral” has important implications for longstanding debates about whether individual differences in morality exist (Doris, 2002; Fleson, Furr, Jayawickreme, Meindl, & Helzer, 2014; McAdams, 2009), or whether moral behavior can be better explained by “seemingly unimportant or irrelevant situational features”

(Ellemers, van der Toorn, Paunov, & van Leeuwen, 2019, p. 26). However, this interpretation rests on the untested assumption that the measured behaviors are actually reflective of moral character.

Here, I test this assumption by correlating several of Bollich and colleagues' (2016) "moral behaviors" with self- and informant-reports of trait morality. If Bollich and colleagues' interpretation is correct, then gratitude, praise, sympathy, helping, apologizing, and optimism should be positively associated with trait morality, whereas sarcasm, bragging, condescension, complaining, criticism, blame, and pessimism should be negatively associated with trait morality. This rests on the assumption that these self- and informant-report measures are valid criteria for trait morality, an issue I will turn to in the discussion.

Personality trait correlates of interest. I examined several personality trait correlates that were either highly interpersonally relevant or seemed conceptually relevant to one or more of the other behaviors of interest in this study. Although I approached this project in an exploratory manner, it is possible to map the interpersonal traits onto the IPC (as reviewed in Chapter 2). This provides a basis for identifying expected links between interpersonal traits and behaviors according to the circular ordering of the IPC (described in Chapter 2). In addition, Table 13 summarizes linkages that would be implied by the definitions of the traits or the empirical literature, including indirect measures of the target behaviors (e.g., self-report measures of gratitude, inferences of sarcasm's correlates based on self-reports of aggressive humor styles).

Table 13

Examples of Associations Between Traits and Interpersonal Behaviors Implied by Trait Definitions or Existing Empirical Evidence

Behavior	Trait Correlates
Pessimism	Neuroticism, (–) Extraversion, (–) Conscientiousness, (–) Agreeableness (S, Kam & Meyer, 2012; Sharpe, Martin, & Roth, 2011, L; Funder & Sneed, 1993)
Criticism	(–) Agreeableness (D)
Complain	Life satisfaction (D)
Condescension	Narcissism (S, Buss & Chiodo, 1991); (–) Agreeableness (L; Funder & Sneed, 1993)
Blame	(–) Agreeableness (D)
Sympathy	Agreeableness (D)
Hope/optimism	(–) Neuroticism, Extraversion, Agreeableness, Conscientiousness (S, Kam & Meyer, 2012; Sharpe et al., 2011)
Help/favors/support	Agreeableness (S, Chapman & Goldberg, 2017; L, Habashi, Graziano, & Hoover, 2016)
Praise/compliment	Openness (S, Chapman & Goldberg, 2017)
Brag	(–) Agreeableness (L; Funder & Sneed, 1993), Narcissism (S, Palmer, Ramsey, Morey, & Gentzler, 2016; Buss & Chiodo, 1991), (–) Conscientiousness (Funder & Sneed, 1993)
Sarcasm	(–) Agreeableness, (–) Conscientiousness (S, Mendiburo-Seguel, Páez, & Martínez-Sánchez, 2015)
Tease	(–) Agreeableness, Neuroticism, Openness (S, Chapman & Goldberg, 2017)
Negative gossip	Extraversion (EAR, Robbins & Karan, 2020)
Neutral/positive gossip	Extraversion, Agreeableness (EAR, Robbins & Karan, 2020)
Insecurity/self-doubt	Neuroticism (D)
Funny	Extraversion (S, Chapman & Goldberg, 2017, Hirsh, DeYoung, & Peterson, 2009; L, Funder & Sneed, 1993), Openness (S, Chapman & Goldberg, 2017), Funny (D)
Gratitude	Extraversion, Agreeableness, Conscientiousness, (–) Neuroticism (Wood, Joseph, & Maltby, 2009)
Apologize	Neuroticism (S, Chapman & Goldberg, 2017)

Note. EAR = Everyday behavior captured by the Electronically Activated Recorder, L = behavior measured in a lab setting, S = self-reported behavior, D = implied by the definition of the trait. Negative correlations are denoted with (–); all other correlations are positive. Correlations are approximately listed from strongest to weakest in absolute magnitude, as implied by these studies.

Interpersonal traits. I included measures of extraversion and agreeableness, the two domains within the Big Five taxonomy that are most strongly related to interpersonal behavior (as reviewed in Chapter 1). In addition, honesty-humility is another major interpersonal trait

captured by the HEXACO taxonomy. This dataset did not contain a measure of honesty-humility. However, as I describe below, narcissism is a highly interpersonal trait that falls within the honesty-humility domain (Lee & Ashton, 2005). Thus, with the inclusion of narcissism, the three major interpersonal domains captured by the Big Five and HEXACO models are represented in these analyses.

Narcissism is conceptualized by a grandiose view of oneself, beliefs about personal superiority, and a sense of entitlement (Grapsas, Brummelman, Back, & Denissen, 2020). A recent process model proposes that narcissists are driven by a dominant status motive, and engage in self-promotion or other-derogation in order to pursue status (Grapsas et al., 2020). Narcissism thus tends to be aligned near the dominance axis (i.e., 90°) in IPC space (Buss & Chiodo, 1991; Gurtman, 1992; Bradlee & Emmons, 1992; Ruiz, Smith, & Rhodewalt, 2001). Given that narcissists tend to make positive first impressions (Back, Schmukle, & Egloff, 2010) that decline over time (Leckelt, Kkufner, Nestler, & Back, 2015; Paulhus, 1998), laboratory-based observations are insufficient for understanding the interpersonal behaviors of narcissists. Instead, naturalistic behavioral observation is especially essential for understanding the behavioral portrait of a narcissist. A small previous EAR study ($N = 79$) found that narcissists tended to argue more and act in more extraverted ways (e.g., being in a group and socializing; Holtzman et al., 2010). Other EAR studies have examined associations between narcissism and language use (e.g., the use of “I” pronouns; Carey et al., 2015; Holtzman et al., 2019). In this dissertation, I build on this work by examining the self-promoting (e.g., bragging), other-derogating (e.g., teasing, condescension), and combative (e.g., criticism, blame) behaviors that seem more conceptually relevant to narcissism.

Additional traits. As mentioned above, one of the goals of this chapter was to explore correlations between a variety of traits and behaviors in order to potentially generate new hypotheses about the interpersonal implications of traits that seem less obviously relevant to interpersonal behavior. I therefore included the remaining Big Five traits of conscientiousness, neuroticism, and openness. Conscientiousness describes tendencies towards productiveness, orderliness, and responsibility. Neuroticism describes the tendency to experience negative emotions such as anxiety, depression, and anger. Finally, openness describes tendencies towards intellectual curiosity, creativity, imagination, and aesthetic appreciation. Thus, at first appearance, conscientiousness, neuroticism, and openness seem less obviously relevant to interpersonal phenomena compared to extraversion and agreeableness. McCrae and Costa (1989) did note that these traits are in one sense interpersonal because they all influence how people interact (e.g., avoiding those who are depressed, seeking out those with similar intellectual and aesthetic values, and judging friends on the basis of their reliability). Nevertheless, they argued that conscientiousness, neuroticism, and openness are not intrinsically interpersonal in that one can accomplish a task, feel unhappy, or respond to art with or without other people.

Others have argued that conscientiousness, neuroticism, and openness contain agentic and communal qualities and can thus be described within IPC space. Such efforts have placed neuroticism in the cold–submissive octant (Ansell & Pincus, 2004; Schmidt, Wagner, & Kiesler, 1999), openness in either the warm (Schmidt et al., 1999) or warm–dominant octant (Ansell & Pincus, 2004), and conscientiousness as either warm (Schmidt et al., 1999) or dominant (Ansell & Pincus, 2004). Conscientiousness, neuroticism, and openness have also been linked to various social outcomes and interpersonal behaviors (for reviews, see Harris & Vazire, 2016; Ozer & Benet-Martínez, 2006; Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007). For example,

conscientious people tend to be more satisfied with their family and romantic relationships (Ozer & Benet-Martínez, 2006), are less likely to get divorced (Roberts et al., 2007), and less frequently swear (Mehl et al., 2006). Neuroticism is correlated with a range of negative relationship outcomes, including dissatisfaction with one's family and romantic relationships, greater conflict and abuse (Ozer & Benet-Martínez, 2006), and a greater likelihood of divorce (Roberts et al., 2007). Neurotic people are also more likely to argue and less likely to socialize (Mehl et al., 2006). Finally, spouses tend to have similar openness scores (Botwin, Buss, & Shackelford, 1997; McCrae, 1996), and open people tend to delay childbirth and have fewer children (Jokela, Alvergne, Pollet, & Lummaa, 2011; van Scheppingen et al., 2016). Thus, it seems worthwhile to further explore the everyday interpersonal implications of conscientiousness, neuroticism, and openness.

I also included measures of self-esteem, satisfaction with life, religiosity, and being funny. I was interested in exploring whether self-esteem would have different interpersonal correlates compared to narcissism, given that both constructs reflect positive self-views but have opposing interpersonal implications (Campbell, Rudich, & Sedikides, 2002). I was also curious about whether satisfaction with life would have behavioral manifestations that reflected positivity (e.g., optimism) or negativity (e.g., complaining, pessimism). I included religiosity because religions seek to encourage prosocial behavior, but behavioral evidence tends to paint a more complex picture of the link between religion and prosocial behavior (Tsang, Rowatt, & Shariff, 2015). Other than these goals, these variables were included to explore the interpersonal implications of traits that fit less neatly within the Big Five model.

Exploring Gender Differences in Everyday Interpersonal Behavior

Beyond personality traits, I explored gender differences in interpersonal behaviors. Although there are many stereotypes about psychological gender differences, the gender-similarities hypothesis proposes that men and women are similar on most, but not all, psychological variables (Hyde, 2014). In general, women tend to self-report substantially higher levels of warmth, as shown by gender differences around $d = 0.45$ on traits such as compassion (Weisberg, DeYoung, & Hirsh, 2011). Men tend to self-report higher levels of agentic traits such as assertiveness, but these gender differences tend to be smaller (e.g., $d = 0.09$ in Weisberg et al., 2011). Using an event-sampling method to sample self-reported dominant, submissive, agreeable, and quarrelsome behaviors, Moskowitz, Suh, and Desaulniers (1994) found that women were more communal (especially with other women), but found no evidence that women and men differed in agentic behaviors. Gurtman and Lee (2009) found that the largest gender differences in interpersonal problems was located at approximately 322° (i.e., women tended to report greater problems related to high warmth and low dominance). They predicted that the maximal gender differences would be approximately an effect size of $r = .16$, which converts to a Cohen's d of 0.32.

Behavioral observation provides a stronger test of the extent of gender differences, compared to self-reports. For example, to the extent that there are gender differences in people's willingness to endorse different statements, or that gender stereotypes affect one's self-perceptions, estimates of gender differences based on self-report methods may be inflated. For example, a previous EAR study dispelled the popular notion that women talk more than men, showing that, instead, both genders tend to talk a similar amount (Mehl, Vazire, Ramírez-Esparza, Slatcher, & Pennebaker, 2007).

Taken as a whole, existing work implies that we should expect the largest gender differences on behaviors in the IPC octants that capture opposing blends of warmth and dominance. This would imply that women will express behaviors in the unassuming–ingenuous octant (i.e., gratitude), which represents a blend of high warmth and low dominance, more frequently than men, whereas men will express behaviors in the arrogant–calculating octant (i.e., bragging, condescension, teasing, and blaming), which represents a blend of low warmth and high dominance, more frequently than women. If women are higher on warmth than are men, women might also be more likely to gossip positively, express sympathy, and praise others more often, whereas men might be more likely to express sarcasm, criticize others, and gossip negatively. If men are higher on dominance than women, they might more frequently express behaviors in the assured–dominant octant (i.e., being funny), whereas women might more frequently express behaviors in the unassured–submissive octant (i.e., apologizing). Finally, there should be fewer gender differences in the octants that capture high warmth and high dominance (i.e., helping and optimism), or low warmth and low dominance (i.e., complaining, pessimism, and insecurity).

To give a few examples of empirical evidence that supports these ideas, Kashdan, Mishra, Breen, and Froh (2009) found that women tend to be more grateful, perhaps because they derive greater benefits from experiencing and expressing gratitude. Women self-report bragging less than men (Palmer et al., 2016), perhaps because they experience greater backlash (Miller, Cooke, Tsang, & Morgan, 1992; Rudman, 1998) and feel uncomfortable about violating modesty norms (Smith & Huntoon, 2014). Using the EAR, Robbins and Karan (2020) found that women tended to engage in more neutral gossip than did men. Men show greater humor production ability, as measured by the ability to produce funny captions for cartoons ($d = 0.38$,

Greengross & Miller, 2011). It would be worthwhile to test whether this means that men also spontaneously express more humor in everyday interpersonal interactions. Finally, using the experience sampling method, Schumann and Ross (2010) found that women reported offering more apologies than men because they have a lower threshold for what constitutes offensive behavior. Thus, overall, the goal of this research question is to provide a stronger test of the extent to which men and women differ in their everyday interpersonal behaviors.

Social Consequences

In the final section of this chapter, I turn to the second side of the lens model (see Figure 1): the links between interpersonal behavior and social consequences—specifically, being liked by and getting along with one’s friends. In two-dimensional models of person perception, being liked tends to be more strongly correlated with communal traits than with agentic traits (Wojciszke, Abele, & Baryla, 2009; Wortman & Wood, 2011). This is likely because communal traits serve the interests of others, and is in line with Pinker’s (1997) functional interpretation of liking as an emotion which is: “roughly, a willingness to offer someone a favor, and is directed to those who appear willing to offer favors back. We like people who are nice to us, and we are nice to people whom we like” (p. 404). This implies that a) behaviors in the warm–agreeable octant of the IPC should be most positively associated with being liked, whereas those in the cold–hearted octant should be most negatively associated with being liked, that b) behaviors in the gregarious–extraverted and unassuming–ingenuous octants should be positively associated with being liked, whereas those in the arrogant–calculating and aloof–introverted octants should be negatively associated with being liked, and that c) behaviors in the assured–dominant and unassured–submissive octants should be relatively uncorrelated with being liked.

Providing some support for the directions of these effects, an early study of the likeability of 555 traits found high likeability ratings (ratings > 4.43 on a 0–6 scale) for being humorous, helpful, grateful, sympathetic, and optimistic, and low likeability ratings (< 2.10) for bragging, being gossipy, critical, pessimistic, insecure, sarcastic, and complaining (Anderson, 1968; see Chandler, 2018 for a recent replication study that found very similar results). One limitation of this approach is that reports of which traits are perceived as being most likeable may not necessarily correspond to how these traits are related to being liked by others in real-world relationships.

In another study, Lott, Lott, Reed, and Crow (1970) asked participants to think of three people who they very much liked, very much disliked, and felt relatively neutral about. Participants then chose three trait adjectives (from a set of 200 adjectives) that most appropriately described each person. They found that people who were liked (vs. disliked or neutral) tended to be described as friendly, helpful, and humorous, whereas those who were disliked (vs. liked or neutral) were described as complaining, egotistical, overcritical, quarrelsome, fault-finding, boastful, conceited, ungrateful, and unsympathetic. More recently, Wortman and Wood (2011) examined the associations between 58 traits (self-reports) and being liked by one's peers (informant-reports). Insecurity and being funny were relatively uncorrelated with being liked. There was some evidence that being unsympathetic was associated with being disliked, but this was inconsistent across the two samples. The findings that humility was associated with being liked, whereas being conceited/egotistical was associated with being disliked, suggests that bragging should be associated with being disliked. Finally, gratitude was associated with being liked.

Both of these studies (Lott et al., 1970; Wortman & Wood, 2011) provide a better sense of why people are liked or disliked, at least from the perspective of informants, but do not answer the question of how much everyday behavioral manifestations of these traits are associated with being liked. Here, I use a multi-method approach to examine the association between actual behaviors (rather than self-reports) and being liked.

Method

Informant Nomination and Reports

During the initial laboratory session, participants nominated up to 10 informants of different types (e.g., friend, local best friend, local friend, hometown friend, sibling, romantic partner, ex-romantic partner, parent). Informants provided reports of the participants' traits and interpersonal outcomes (described below). Because romantic partners and parents tend to be more biased informants (Murray, Holmes, Dolderman, & Griffin, 2000; Solomon & Vazire, 2014), I only included informants who were friends (a decision I made before analyzing the data). The final analyses included 641 informants (185 men and 454 women) who were aged between 18–44 years ($M = 19.43$, $SD = 2.42$).

Measures

Means, standard deviations, reliability estimates, and intercorrelations among the personality trait and interpersonal outcome variables are shown in Table 14.

Trait measures. Participants completed self-report measures of the Big Five traits, morality, narcissism, self-esteem, satisfaction with life, religiosity, and funniness. Because previous research suggests that people may have less self-knowledge of highly-evaluative traits (Vazire, 2010), I also included informant-reports of morality and funniness, which are highly evaluative. I report the results separately for self- and informant-reports of these traits.

Table 14

Descriptive Statistics and Intercorrelation Matrix for Personality Traits and Interpersonal Correlates

	<i>M</i>	<i>SD</i>	Rel.	Intercorrelations														
				1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.
Personality traits																		
1. Extraversion	9.34	2.84	.90															
2. Agreeableness	10.29	2.18	.81	.20														
3. Narcissism	.31	.18	.66	.42	−.19													
4. Morality (self)	11.57	3.03		.04	.49	−.15												
5. Morality (informant)	11.81	2.11	.28	.00	.25	−.16	.38											
6. Conscientiousness	9.75	2.29	.85	.07	.25	.06	.38	.22										
7. Neuroticism	7.68	2.64	.85	−.26	−.31	−.24	−.07	−.05	−.17									
8. Openness	10.38	1.98	.79	.09	.09	.06	.03	−.13	−.09	−.09								
9. Self-esteem	9.37	4.18		.34	.18	.41	.11	.07	.23	−.61	.04							
10. Satisfaction with life	11.47	2.87		.36	.23	.30	.16	.08	.19	−.51	.05	.64						
11. Religiosity	3.04	1.78		.10	−.02	.14	.15	.26	.13	−.06	−.11	.11	.12					
12. Funny (self)	11.01	3.03		.39	.08	.24	.11	−.04	.14	−.17	.08	.21	.17	−.02				
13. Funny (informant)	12.29	2.02	.15	.18	−.02	.06	−.04	.24	−.05	−.07	.02	.02	−.01	−.03	.23			
Interpersonal correlates																		
14. Likeability	13.05	1.77	.27	.15	.17	−.03	.06	.40	.00	−.15	−.01	.06	.11	.08	.09	.66		
15. Liked by informants	6.34	0.79	.22	.13	.13	−.07	−.03	.36	−.07	−.07	−.00	−.00	.02	.08	.08	.51	.66	
16. Relationship quality	5.73	0.89	.90	.05	.07	.00	−.01	.41	.02	−.08	−.10	.05	.03	.07	.13	.45	.58	.70

Note. Rel. = ω reliability for composites of three or more items (i.e., the Big Five traits), α for the two-item relationship quality composite and Narcissistic Personality Inventory (which had binary items), or *ICC*(1) for inter-rater reliability of single-item informant-reported measures (morality, funny, likeability, and liking). Correlations $\geq |.13|$ (except for the association between extraversion and being liked) are significant at $p < .05$ (not corrected for multiple comparisons).

Big Five traits. Participants completed the Big Five Inventory (BFI-44; John & Srivastava, 1999), which measures trait extraversion, agreeableness, conscientiousness, neuroticism, and openness. All responses were made on a 15-point scale (1 = *Disagree strongly*, 8 = *Neither agree nor disagree*, 15 = *Agree strongly*). I computed composites by taking the mean of all items for each scale (after reverse-coding negatively-keyed items).

Narcissism. Participants completed a 16-item short version of the Narcissistic Personality Inventory (NPI-16; Ames, Rose, & Anderson, 2006). These 16 item pairs were selected from the original NPI-40 (Raskin & Terry, 1988) on the basis of face validity (i.e., being primarily representative of narcissism rather than related but distinct constructs) and coverage of domains. Each of the 16 item pairs contained a narcissistic statement (e.g., “I really like to be the center of attention”) and a non-narcissistic statement (e.g., “It makes me uncomfortable to be the center of attention”). Participants chose the statement within each item pair that best described them. The resulting composite was based on the proportion of narcissistic responses (i.e., with a possible range from 0 to 1).

Trait morality. Participants and their informants completed an ad hoc 1-item measure of global morality (“I am/[target’s name] is someone who is moral/does the right thing”; 1 = *Disagree strongly*, 8 = *Neither agree nor disagree*, 15 = *Agree strongly*).

Self-esteem. Participants rated their global self-esteem using a single item, “I am someone who has high self-esteem” (1 = *Disagree strongly*, 15 = *agree strongly*). Similar 1-item measures of self-esteem have been previously validated (Robins, Hendin, & Trzesniewski, 2001).

Satisfaction with life. Participants rated their life satisfaction using a single item, “How satisfied are you with your: life”, on a scale from 1 = *Completely dissatisfied* to 15 = *Completely*

satisfied. Similar 1-item measures of life satisfaction have been previously validated (Lucas & Donnellan, 2012).

Religiosity. Participants reported their religiosity using a single item: “How religious are you?” (1 = *Not at all religious*, 4 = *Somewhat religious*, 7 = *Very religious*).

Trait funniness. Participants and their informants completed an ad hoc 1-item measure of how funny the participant is (“I am/[target’s name] is funny”; 1 = *Disagree strongly*, 8 = *Neither agree nor disagree*, 15 = *Agree strongly*).

Interpersonal outcomes. I included informant reports of general likeability, being liked by the informants, and relationship quality. Conceptually, I consider these to be outcome variables. However, because the data are correlational, I cannot draw strong conclusions about the causal direction of the correlations. For example, it is possible that participants’ behaviors cause their informants to like or dislike them, or that participants who are liked by their friends feel more comfortable expressing certain behaviors around them (i.e., reverse causality). It is also possible that a third variable (e.g., personality traits) explains the associations between interpersonal behaviors and being liked.

Likeability. Informants completed a 1-item measure of how likeable the participant is (“[target’s name] is someone who is likeable”; 1 = *Disagree strongly*, 8 = *Neither agree nor disagree*, 15 = *Agree strongly*).

Being liked by friends. Informants also rated how much they liked the participant (“How much do you like [target’s name]?”; 1 = *Not at all*, 7 = *Very much*).

Relationship quality. Finally, informants completed two indicators of relationship quality (“How satisfied are you with your relationship with [target’s name]?”, 1 = *Not at all*, 7 = *Extremely satisfied*; “How would you rate the quality of your relationship with [target’s name]?”,

1 = *Not at all*, 7 = *Extremely good*), which I combined into a two-item composite by taking the mean of these two items.

Data Analyses

To aggregate the informant-report data, I averaged the scores across all available informants for a given target. Almost all ($n = 198$) participants had two or more informants, 53 participants had one informant, and 36 participants had no informants. I computed zero-order correlations to address the research questions about trait correlates and social consequences, and conducted *t*-tests to examine gender differences.

Sensitivity power analyses showed that I had 80% power to detect correlations $\geq |.16|$ for the self-reported measures and $\geq |.18|$ for the informant-reported measures. With 91 men and 194 women, I had 80% power to detect medium-sized gender differences (standardized *ds*) $\geq |0.36|$. For the gender differences, I report unstandardized effect sizes which represent absolute differences in the percentages of conversations in which men or women expressed a given behavior. To facilitate comparability of effect sizes with the personality trait correlations, I also report two standardized effect sizes: Cohen's *d* and the R^2 . Because the R^2 is the square of a correlation, these effect sizes can be readily converted back into a correlation.

Given the exploratory nature of this study, I used a conventional $p < .05$ threshold as a heuristic for identifying new hypotheses to be tested in future studies. This approach prioritizes reducing Type II error over Type I error. However, to provide additional context on which results might be less likely to be due to chance, I also indicate which effects survive a false discovery rate (FDR) correction (Benjamini & Hochberg, 1995). I applied this correction separately for each research question (trait correlates, gender differences, and social consequences).

To provide a complementary metric of the extent to which the overall set of results might be due to chance, I also conducted randomization tests for each research question.

Randomization tests (Block, 1960; Sherman & Funder, 2009) test whether the overall set of personality correlates, gender differences, and social consequences obtained in this dissertation would be expected to occur by chance. Whereas the FDR correction addresses the question of which specific effects should be considered significant, randomization tests answer the question of whether the set of findings, as a whole, provide evidence for the broader claims that a) traits are associated with this set of interpersonal behaviors, b) gender differences in this set of interpersonal behaviors exist, and c) this set of interpersonal behaviors is associated with interpersonal outcomes.

The randomization test procedure involves the following steps: 1) randomly re-assign scores on the outcome variables (e.g., the 18 interpersonal behaviors) from one participant to the set of predictor variables (e.g., traits) from a random participant, thus creating a new pseudo-sample; 2) compute the effect size between each predictor variable and the outcome variable; 3) record the number of uncorrected $p < .05$ effects obtained in this pseudo-sample; 4) repeat steps 1–3 many times to create approximate chance sampling distributions of the number of significant effects. Thus, this process simulates random relationships between personality and interpersonal behavior, between gender and interpersonal behavior, and between interpersonal behavior and interpersonal outcomes, while maintaining the dependencies among each set of variables (e.g., the covariations among the set of interpersonal behaviors). Calculating the proportion of pseudo-datasets that produce as many or more significant effects as the number of significant effects observed in the real dataset yields a p value that indicates the probability of obtaining the original results under chance. I conducted 10,000 such simulations for each research question.

Results

Personality Trait Correlates

Across 286 correlations (13 traits $\times$ 18 individual behaviors and 4 behavioral composites), 29 effects were detectable under a liberal $p < .05$ threshold, which I use as a heuristic for generating hypotheses to be tested in future studies (see Table 15). Out of these 29 effects, only one effect survived a FDR correction—the negative association between satisfaction with life and expressions of insecurity. However, randomization tests on 10,000 pseudo-datasets showed that the observed number of 29 correlations in the real dataset was significantly greater than the 14.28 correlations that would be expected by chance, $p = .0017$ (i.e., only 0.17% of these pseudo-datasets had a greater or equal number of correlates compared to the 29 correlates that were observed in the real dataset). Although randomization tests cannot speak to which specific effects are robust, these results do suggest that the overall set of associations is unlikely to be spurious. In other words, this provides evidence for the existence of associations between these personality traits and interpersonal behaviors.

Are interpersonal traits associated with interpersonal behavior? Out of the set of traits I considered, extraversion, agreeableness, and narcissism are the traits that are most conceptually related to interpersonal behavior. Thus, I began by examining the associations between interpersonal traits and interpersonal behaviors (see Table 15). Overall, these interpersonal traits were only weakly correlated with the set of 18 interpersonal behaviors, with average absolute correlations ranging between $r = .06$ – $.07$. The detectable correlations that emerged suggested that extraverts more frequently bragged, teased, gossiped in general, gossiped negatively, and expressed gratitude. Agreeable people were less frequently condescending, sarcastic, and funny. Apart from these findings, agreeableness had surprisingly few behavioral

correlates considering its conceptual relevance to prosocial and antagonistic behaviors. Finally, narcissists more frequently expressed condescension, mocked and teased others, and bragged. None of these associations survived a FDR correction.

To what extent are these behaviors “moral”? Another key goal of this chapter was to test Bollich and colleagues’ (2016) assumption that expressions of gratitude, praise, sympathy, help, apologizing, and optimism are indicative of positive moral conduct (i.e., “positive moral behaviors”), whereas sarcasm, bragging, condescension, complaining, criticism, blame, and pessimism are indicative of negative moral conduct (i.e., “negative moral behaviors”). I tested this assumption by examining whether these behaviors were correlated with self- and informant-reports of trait morality.

These analyses (see Table 15) provided very little evidence for the idea that these behaviors were reflective of moral character. Out of the set of “positive moral behaviors”, only apologizing was detectably positively associated with self-reported morality, and none of these behaviors were associated with informant-reported morality. Interestingly, the supportiveness composite was also weakly associated with self-reported morality, but none of the correlations with the individual behaviors of sympathy, optimism, helping, and praise were detectable. Out of the set of “negative moral behaviors”, only condescension and bragging were detectably negatively associated with informant-reported morality. In addition, blame was *positively* associated with self-reported morality, which suggests that if anything, the tendency to blame others might be an indicator of positive (rather than negative) moral character. None of these associations survived a FDR correction.

Table 15

Correlations Between Personality Traits and Interpersonal Behaviors

	Interpersonal Traits			Morality		Additional Traits								Mean behavior r
	E	A	Narc.	Self	Inf.	C	N	O	Self-esteem	SWL	Rlg.	Funny (Self)	Funny (Inf.)	
Negativity	-.01	-.08	-.01	.04	-.04	.06	.05	-.06	-.05	-.09	.02	.09	.04	.05
Pessimism	-.06	.02	-.06	.01	.05	.03	.03	.00	-.04	-.10	.00	.00	.06	.04
Criticism	.04	-.11	.07	.02	.00	.06	.03	-.11	.00	.00	.09	.15*	.00	.05
Complain	-.03	-.04	-.06	.05	.00	.06	.09	-.03	-.11	-.15*	.00	.05	.08	.06
Condescension	.03	-.14*	.12*	-.02	-.21***	.02	-.09	-.03	.09	.06	-.02	.04	-.04	.07
Blame	-.04	-.01	-.09	.13*	.05	.01	.11	-.04	-.04	.00	.00	.06	-.02	.05
Supportiveness	.04	.08	-.03	.13*	.08	.13*	.07	.08	-.04	.05	-.09	.02	-.06	.07
Sympathy	.01	.07	-.06	.10	.00	.03	.07	.03	-.09	-.02	-.14*	.05	-.08	.06
Hope/optimism	.04	.09	-.01	.10	.12	.12*	.01	.07	.00	.11	-.06	.03	-.03	.06
Help	.00	-.02	.05	.07	.04	.09	.04	.05	.01	.00	-.04	-.01	-.02	.03
Praise/compliment	.09	.07	-.09	.09	.05	.14*	.08	.07	-.03	.06	.03	-.05	-.04	.07
Mockery	.09	-.04	.13*	.02	-.04	-.01	-.07	-.09	.04	.08	.06	.15*	.08	.07
Brag	.12*	.07	.12*	.02	-.12*	.00	-.08	-.05	.06	.13*	.10	.07	-.06	.08
Sarcasm	-.04	-.13*	.05	-.02	-.03	-.03	.03	.03	-.02	-.01	.03	.08	.04	.04
Tease	.14*	.01	.13*	.04	.02	.01	-.11	-.16**	.05	.09	.03	.16**	.13*	.08
Gossip	.13*	-.05	.07	.00	-.06	.03	.07	-.09	-.07	.05	.07	-.05	.06	.06
Negative gossip	.12*	-.12	.10	.01	-.07	.07	.09	-.11	-.06	-.03	.03	.11	.03	.07
Neu./pos. gossip	.10	-.01	.04	-.01	-.04	.01	.04	-.05	-.06	.07	.07	-.11	.06	.05
No factor														
Insecurity	-.05	.02	-.11	-.04	.00	-.10	.20***	-.03	-.15**	-.22***	.02	-.13*	-.04	.09
Funny	.04	-.13*	.06	-.01	-.01	-.01	.04	-.09	-.05	.01	.04	.04	.15*	.05
Gratitude	.13*	.05	.08	.03	.03	.13*	.07	.11	.03	.04	.00	.02	.10	.06
Apologize	-.08	.06	-.05	.13*	.05	.06	.12*	-.01	-.03	-.01	.00	-.10	-.05	.06
Mean trait r	.07	.06	.07	.05	.05	.06	.07	.06	.05	.06	.04	.07	.06	

Note. * $p < .05$, ** $p < .01$, *** $p < .001$, not corrected for multiple comparisons. E = Extraversion, A = Agreeableness, Narc. = narcissism, C = Conscientiousness, N = Neuroticism, O = Openness, Rlg. = religiosity, Self = self-reported, Inf. = informant-reported. Effects that survived a false discovery rate correction are shown in **boldface**.

Which additional traits are associated with interpersonal behavior? Finally, I explored the interpersonal behavioral manifestations of an additional set of traits. The average absolute correlations for the remaining Big Five domains of conscientiousness, neuroticism, and openness were in the .06–.07 range. This suggests that these conceptually less interpersonal traits were nevertheless associated with the set of interpersonal behaviors to a similar extent as were the interpersonal domains of extraversion and agreeableness. Specifically, conscientious people tended to more frequently express supportiveness, optimism, praise, and gratitude, neurotic people expressed more insecurity and apologized more, and open people were less inclined to tease others. None of these associations survived a FDR correction.

Next, I examined the associations between interpersonal behaviors and additional traits that fit less well into the Big Five taxonomy: self-esteem, satisfaction with life, religiosity, and being funny. Self-esteem, like narcissism, reflects positive self-views. However, self-esteem was associated with less frequent expressions of insecurity, but was not detectably associated with the behavioral manifestations of narcissism (i.e., condescension, bragging, mockery, and teasing). These results therefore support the conceptual distinction between self-esteem and narcissism. Satisfaction with life—a broad indicator of personal well-being—was also manifested in several interpersonal behaviors. Specifically, people who were more satisfied with their lives tended to complain less, express less insecurity, and brag more. Adding to the mixed literature on the link between religiosity and prosocial behavior, religiosity was associated with being *less* likely to express sympathy. Self- and informant-reports of funniness were both associated with more frequently teasing others. People who saw themselves as being funny also more frequently criticized and mocked others, and less frequently expressed insecurity. Finally, people whose informants saw them as being funny more frequently did things that were funny (from the

perspective of the EAR coders). Out of these correlations, only the negative association between satisfaction with life and expressions of insecurity survived a FDR correction.

Gender Differences

Next, I examined the extent to which men and women differed in how often they expressed these interpersonal behaviors. Out of 22 gender differences (18 specific behaviors and 4 factors) that I tested, 8 were detectable at a liberal $p < .05$ threshold (see Table 16). Five of these effects survived a FDR correction: compared to men, women tended to spend a larger percentage of conversations praising others, gossiping in general, gossiping negatively, expressing insecurity, and expressing gratitude. The other detectable effects were that women tended to spend a larger percentage of conversations complaining and gossiping neutrally or positively, whereas men tended to be more condescending; however, these effects did not survive a FDR correction. Randomization tests on 10,000 pseudo-datasets showed that the observed number of 8 gender differences in the real dataset was significantly greater than the 1.10 gender differences that would be expected by chance, $p = .002$. This provides evidence for the existence of gender differences among this set of 18 interpersonal behaviors.

In terms of absolute percentages, these gender differences reflected differences in the range of 1–2% of conversations. Gender also explained at most 5% of the variance in each of these interpersonal behaviors. However, the standardized estimates were in the moderate to large range, with the largest differences in praise ($d = 0.50$) and insecurity ($d = 0.46$), and moderate differences for gossip and negative gossip ($ds = 0.36$) and gratitude ($d = 0.39$). These 95% confidence intervals typically captured effects that ranged from small to relatively large.

Table 16
Gender Differences in Frequency of Interpersonal Behaviors

Composite or Behavior	Men (<i>n</i> = 91)		Women (<i>n</i> = 194)		Unstandardized Difference		Standardized Difference		<i>p</i>	<i>R</i> ²
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	Dif.	95% CI	<i>d</i>	95% CI		
Negativity	3.75	2.12	3.88	2.02	0.14	[−0.38, 0.65]	0.07	[−0.18, 0.32]	.601	.00
Pessimism	2.75	2.91	2.57	2.37	−0.18	[−0.82, 0.46]	−0.07	[−0.32, 0.18]	.575	.00
Criticism	4.51	2.92	4.46	3.22	−0.05	[−0.84, 0.73]	−0.02	[−0.27, 0.23]	.895	.00
Complain	7.88	5.19	9.37	5.17	1.49	[0.20, 2.79]	0.29	[0.04, 0.54]	.024	.02
Condescension	2.65	2.91	2.01	2.24	−0.64	[−1.26, −0.02]	−0.26	[−0.51, −0.01]	.042	.01
Blame	0.95	1.09	1.02	1.09	0.07	[−0.21, 0.34]	0.06	[−0.19, 0.31]	.626	.00
Supportive	7.63	3.52	8.19	3.70	0.56	[−0.35, 1.47]	0.15	[−0.1, 0.4]	.228	.01
Sympathy	8.70	6.36	9.34	5.42	0.64	[−0.80, 2.07]	0.11	[−0.14, 0.36]	.382	.00
Hope/optimism	7.75	6.33	8.01	6.27	0.25	[−1.32, 1.83]	0.04	[−0.21, 0.29]	.750	.00
Help	9.88	5.71	9.51	6.02	−0.36	[−1.84, 1.12]	−0.06	[−0.31, 0.19]	.630	.00
Praise/compliment	4.18	3.13	5.89	3.57	1.71	[0.85, 2.57]	0.50	[0.24, 0.75]	< .001	.05
Mockery	2.83	1.94	2.49	1.65	−0.34	[−0.78, 0.10]	−0.19	[−0.44, 0.06]	.128	.01
Brag	1.38	1.40	1.24	1.50	−0.14	[−0.51, 0.23]	−0.10	[−0.35, 0.15]	.452	.00
Sarcasm	3.10	2.74	2.81	2.57	−0.29	[−0.94, 0.37]	−0.11	[−0.36, 0.14]	.392	.00
Tease	4.00	3.34	3.41	2.67	−0.59	[−1.32, 0.13]	−0.20	[−0.45, 0.05]	.110	.01
Gossip	6.69	4.52	8.19	3.96	1.50	[0.46, 2.54]	0.36	[0.11, 0.61]	.005	.03
Negative gossip	3.19	2.65	4.31	3.30	1.11	[0.33, 1.89]	0.36	[0.11, 0.61]	.005	.03
Neu./pos. gossip	10.18	8.25	12.07	5.98	1.89	[0.19, 3.59]	0.28	[0.03, 0.53]	.029	.02
No factor										
Insecurity	2.52	2.10	3.78	3.01	1.27	[0.58, 1.96]	0.46	[0.21, 0.71]	< .001	.04
Funny	10.61	6.40	11.19	5.87	0.58	[−0.93, 2.09]	0.10	[−0.15, 0.35]	.451	.00
Gratitude	1.82	1.96	2.82	2.83	1.01	[0.36, 1.65]	0.39	[0.14, 0.64]	.002	.03
Apologize	0.84	1.37	1.17	1.43	0.32	[−0.03, 0.68]	0.23	[−0.02, 0.48]	.073	.01
Mean gender dif.					0.69		0.20			.01

Note. Effects that survived a false discovery rate correction are shown in **boldface**. Positive effect sizes represent higher scores for women; negative effect sizes represent higher scores for men. |gender dif.| = absolute gender difference.

Social Consequences

Finally, I examined whether these interpersonal behaviors had any implications for friend-reported social consequences. The outcomes of interest were a) being seen as a likeable person, b) being liked by one's friends, and c) the quality of one's relationships with friends. As shown in Table 17, there were surprisingly few detectable correlates overall. Out of 66 correlations, only 6 were detectable under a liberal $p < .05$ threshold. Specifically, those who bragged more tended to be liked less, those who were funnier tended to be liked more, those who complained more had better relationships with their friends, and teasing was positively associated with all three interpersonal outcomes ($r_s = .16-.18$). For comparison, the strongest correlation between self-reported personality traits and interpersonal outcomes was .17 for the association between agreeableness and likeability (see Table 14). Thus, it was especially impressive that a single behavior—teasing—was as strongly correlated with these three interpersonal outcomes as were any of these personality traits (which should capture a broad range of behaviors that are relevant to being liked and getting along with one's friends).

However, none of these effects survived a FDR correction. Moreover, randomization tests on 10,000 pseudo-datasets showed that the observed number of 6 correlations in the real dataset was not detectably different from the 3.31 correlations that would be expected by chance, $p = .189$. Thus, these correlations might be spurious and provide little evidence for the idea that these interpersonal behaviors are associated with friend-reported social consequences.

Table 17

Correlations Between Interpersonal Behaviors and Friend-Reported Social Consequences

	Likeability	Liked by Friends	Relationship Quality	Mean behavior r
Negativity	.01	.03	.12	.05
Pessimism	.00	.02	.07	.03
Criticism	.08	.05	.09	.07
Complain	.01	.08	.15*	.08
Condescension	-.10	-.12	-.02	-.08
Blame	.06	-.01	.05	.03
Supportive	.00	-.05	-.03	-.03
Sympathy	-.03	-.02	-.04	-.03
Hope/optimism	.06	-.03	-.03	.00
Help/favors/support	-.04	-.08	.00	-.04
Praise/compliment	.02	.03	-.03	.01
Mockery	.09	.05	.10	.08
Brag	-.02	-.13*	-.01	-.05
Sarcasm	-.01	-.01	.01	.00
Tease	.18**	.16*	.17**	.17
Gossip	.05	.04	.03	.04
Negative gossip	-.01	.03	.05	.02
Neutral/positive gossip	.07	.04	.02	.04
No factor				
Insecurity/self-doubt	-.01	.00	.02	.00
Funny	.12	.15*	.11	.13
Gratitude	.05	.01	-.11	-.02
Apologize	.00	.00	.06	.02
Mean outcome $ r $	.05	.05	.06	

Note. * $p < .05$, ** $p < .01$, *** $p < .001$, not corrected for multiple comparisons. No effects survived a false discovery rate correction. Mean r s were computed from the raw r s for the mean behavior r s and the absolute r s for the mean behavior $|r|$ s. More positive mean behavior r s indicate stronger associations with positive interpersonal outcomes. Larger mean outcome $|r|$ s indicate which interpersonal outcomes had stronger behavioral correlates (in either direction).

Discussion

In summary, this chapter explored the correlates and social consequences of everyday interpersonal behaviors. Although I took an exploratory approach and ran many analyses, randomization tests provided evidence for the broad conclusions that personality traits are manifested in everyday interpersonal behaviors and that men and women express different frequencies of these behaviors. However, I found little evidence for the idea that these

interpersonal behaviors were associated with friend-reported interpersonal outcomes. Below, I highlight interesting patterns and discuss the implications of these results.

Behavioral Implications of Personality Traits

Do personality trait measures capture behavior-relevant tendencies? Are traits connected with behavior in a decontextualized way? Have the interpersonal implications of traits beyond the extraversion, agreeableness, and honesty-humility domains been underappreciated? These are the kinds of questions I hoped to address by exploring the associations between a large set of personality traits and interpersonal behaviors.

Overall, the detectable correlates that emerged (under a very liberal $p < .05$ threshold for exploratory findings) were generally consistent with existing conceptualizations of these traits. For example, agreeable people were less condescending and sarcastic. Neurotic people expressed more insecurity and were more apologetic. Those who were satisfied with their lives complained less frequently and expressed less insecurity. Consistent with the distinction between narcissism and self-esteem (Campbell et al., 2002), narcissism and self-esteem showed distinct detectable behavioral profiles: whereas narcissists tended to be more condescending and mocked, bragged, and teased others more often, those who were higher on self-esteem were only less likely to express insecurity.

Although these effect sizes tended to be small (in the range of $|.12-.22|$ for the detectable correlates), these results show that self-reports of personality traits are manifested in conceptually-relevant behaviors in everyday life. In other words, instead of capturing only subjective thoughts and feelings, self-reported traits also make it out of the head in the form of behaviors that can be observed by and potentially impact other people. In doing so, these results reveal the concrete behavioral implications of various personality traits—an issue I will explore

further in Chapter 6. In addition, because these associations emerged without explicitly measuring and modeling the effects of context, this suggests either that traits are manifested in behavior in a decontextualized way (McAdams & Pals, 2006) or that the relevant eliciting stimuli tend to be broad classes that are often present (DeYoung, 2015).

However, agreeableness had surprisingly few interpersonal behavioral correlates, considering its theoretical relevance to prosocial and antagonistic behaviors. Although agreeable people were less sarcastic and condescending, I found no evidence that they were more supportive, sympathetic, helpful, grateful, and apologetic, or that they were less inclined to complain, criticize, and blame others. Several of these null associations may be due to the low reliability with which the behaviors were measured. However, because other traits were detectably correlated with these behaviors despite the low reliability of these behaviors, this cannot be the only explanation.

One possible methodological explanation is that measurement at the level of the domain lacked fidelity, in that trait agreeableness was not closely matched enough with the behavioral criteria to be able to predict these criteria with high accuracy (Ashton, Jackson, Paunonen, Helmes, & Rothstein, 1995; Paunonen & Ashton, 2001). Hierarchical models of the Big Five show that trait domains subsume narrower traits, described as aspects or facets (e.g., Costa & McCrae, 1995; DeYoung et al., 2007; Soto & John, 2017). Whereas the *compassion* aspect of agreeableness captures active emotional concern for others' well-being (and should therefore be more related to behaviors such as sympathy and helping), the *politeness* or *respectfulness* aspect involves inhibiting aggressive or antagonistic impulses (and should therefore be more negatively related to behaviors such as criticism, complaining, condescension, blame, and sarcasm). Finer-

grained measurement might facilitate the discovery of the behavioral correlates of specific aspects of agreeableness.

These results might also suggest that the connection between agreeableness and interpersonal behavior is complex. Rather than being straightforwardly associated with interpersonal behavior averaged across all situations, agreeableness might only be manifested behaviorally under specific conditions. Agreeableness might also predict the thoughts and feelings that accompany interpersonal behaviors, rather than how often people engage in these behaviors. I will return to each of these possibilities in the general discussion (Chapter 7). Finally, these results may suggest that people do not infer their agreeableness from how often they enact prosocial and antagonistic behaviors and that they lack self-knowledge of their agreeableness (e.g., Sun & Vazire, 2019).

Finally, beyond confirming the behavioral implications that are consistent with existing conceptualizations of these traits, I also hoped to expand current conceptualizations of traits by revealing their less obvious behavioral implications. For example, although humor is not inherent to the definition of agreeableness, I found that agreeable people tended to be less funny in everyday life. Considering that being funny was associated with sarcasm and teasing (see Figure 4) and initially loaded onto the mockery factor in Year 1 (see Table 9), and that self-reported funniness was associated with criticizing others, the mockery factor, and teasing, this suggests that agreeable people may be less funny to the extent that they are less willing to engage in forms of humor that come at the expense of others. In addition, people who were satisfied with their lives not only complained and expressed insecurity less frequently—they also bragged more, perhaps because they have more daily satisfactions to share.

Two additional findings were somewhat more puzzling. First, religious people tended to express *less* sympathy in their everyday conversations, further adding to the mixed literature on the links between religiosity and prosocial behavior (Tsang et al., 2015). It was also unclear why openness was negatively associated with teasing. This was opposite to the positive correlation between openness and act-frequency self-reports of “Made fun of someone” (Chapman & Goldberg, 2017). Although these puzzling findings might be spurious, they nevertheless suggest that actual behavior can reveal a different picture compared to self-report measures.

Finally, these findings raise the question of what it means for a trait to be “interpersonal”. Several traits that are less inherently defined by interpersonal behavior (e.g., conscientiousness, neuroticism, satisfaction with life, self-esteem) were nevertheless manifested in a variety of conceptually-relevant interpersonal behaviors. For example, whereas there was little evidence that agreeableness was correlated with prosocial behaviors, conscientious people tended to be more supportive, express more optimism and gratitude, and praise others more often. I also computed average absolute correlations for each trait in order to explore which traits were particularly strongly associated with the full set of interpersonal behaviors. Generally, there was little evidence that traits that were conceptually more relevant to interpersonal behavior (extraversion, agreeableness, narcissism, morality, and being funny; average $|r| = .061$) were systematically more strongly associated with this set of 18 behaviors, compared with traits that were conceptually less relevant to interpersonal behavior (conscientiousness, neuroticism, self-esteem, satisfaction with life, religiosity; average $|r| = .057$). This suggests that all traits—even those that are not explicitly defined by interpersonal behaviors—have some implications for interpersonal behavior.

Clarifying the Meaning of Everyday Interpersonal Behaviors

Just as revealing the behavioral implications of personality trait measures can help to establish their construct validity, trait–behavior correlations can also clarify the meaning of everyday interpersonal behaviors. In Chapter 3, I began to explore the shared interpersonal meanings that might underlie specific behaviors. Although the four behavioral composites of negativity, supportiveness, mockery, and gossip formed reasonably reliable factors, Chapter 4 showed that only two of the four behavioral composites—mockery and gossip—showed some amount of rank-order stability over time. In this chapter, I further explored the validity and utility of the four composites by correlating them with trait measures.

One of the principles behind creating behavioral composites was that broader, more reliable composites should be more strongly associated with trait measures. However, this was not the case. The negativity factor had no detectable trait correlates, supportiveness and mockery each only had two trait correlates, and gossip only had one trait correlate. A greater number of associations emerged at the level of individual behaviors, and the correlations between traits and individual behaviors within these factors tended to be quite specific (e.g., agreeableness was negatively associated with condescension, but was not detectably associated with any of the other behaviors within the negativity factor). This suggests that even if the behaviors that load onto these factors do reflect broader shared meanings, specific behaviors might still be a more useful unit of analysis. It also suggests that human coders were able to make fine-grained distinctions between closely-related behaviors (e.g., criticism vs. blame).

I also aimed to test whether the behaviors measured by Bollich and colleagues (2016) were indeed indicative of positive and negative moral conduct. I found relatively little support for this idea, across self- and informant-reported measures of morality. Self-reported morality was positively associated with the supportiveness factor (but not detectably so with the

individual behaviors) and apologizing. This provides some indication that apologizing and a composite of supportive behaviors (sympathy, optimism, helping, and praise) may be weak indicators of positive moral behavior. In contrast, expressions of gratitude in everyday conversations may be less relevant to moral character (even if gratitude is generally seen as a moral virtue; e.g., DeSteno, Bartlett, Baumann, Williams, & Dickens, 2010; McCullough, Kilpatrick, Emmons, & Larson, 2001; Peterson & Seligman, 2004). None of these associations replicated when examining informant-reported measures of morality.

Moreover, there was no evidence that any of the “negative” behaviors (pessimism, criticism, complaining, condescension, blame, bragging, and sarcasm) were negatively associated with self-reported morality. In fact, people who saw themselves as being more moral spent a *greater* proportion of their everyday conversations blaming others. This raises the possibility that people think that holding others accountable for their misdeeds reflects positively on one’s own moral character. This is consistent with the perspective that blame is a public act of moral criticism that serves as a key mechanism for the regulation of social behavior (Malle, Guglielmo, & Monroe, 2014).

Two additional associations emerged in the informant-reports: informant-reported morality was negatively correlated with both condescension and bragging, suggesting that other-judgments of a target’s moral character might be partially shaped by perceptions of the target’s condescension and boastfulness. The finding that condescension and boastfulness may be especially diagnostic of poor moral character is consistent with the finding that negative behaviors are more diagnostic for immorality because only immoral people would be expected to engage in negative acts (whereas moral behaviors are expected to be performed by both moral and immoral people; Skowronski & Carlston, 1987). Trafimow and Trafimow (1999) further

found that people change their moral evaluations of others at a faster rate in response to violations of *perfect duties* (which are moral rules that should absolutely never be broken, e.g., dishonesty; Kant, 1797/1991) compared to violations of *imperfect duties* (flexible “duties of virtue” that do not always have to be obeyed, e.g., charity). Thus, these results could suggest that condescension and boastfulness are more diagnostic of poor moral character because they are seen as violations of perfect duties, whereas behaviors such as helpfulness and sympathy are merely imperfect duties.

Taken as a whole, there was little evidence that Bollich and colleagues’ (2016) “moral behaviors” were related to self- or informant-reports of morality, which I used as criterion measures of moral character. When two measures disagree (e.g., self-report vs. behavior), either or both of the measures could be invalid. Thus, although one interpretation is that these behaviors are not valid indicators of moral character, a second interpretation is that it was the self- and informant-report criterion measures that were invalid. This could be the case, considering that the measure was only 1 item long and that self- and informant-report measures of morality contain disturbingly high amounts of evaluative bias (up to 60%; Barranti, 2019). Evaluative bias, which captures the tendency to indiscriminately rate oneself more positively or negatively across a set of evaluative traits (rather than in a differentiated way within the set of evaluative traits), has an especially big impact on the accuracy of self-reports (Vazire, 2010). A third criterion for moral character would be needed to adjudicate between these two possibilities. Still, despite these ambiguities, these results call into question the interpretation of Bollich and colleagues’ (2016) behaviors as being moral or immoral and show that validity should not be taken for granted.

The question of whether such behaviors are indicators of positive or negative moral character is not only an empirical issue. Such an interpretation could also be justified on theoretical grounds. This requires a clear definition of what constitutes moral behavior. Functional perspectives on morality conceptualize moral systems as sets of values and rules that suppress selfishness and make cooperative social life possible (Haidt & Kesebir, 2010). These systems provide a basis for judging which behaviors are “good” or “bad”, “ethical” or “unethical”, and “moral” or “immoral.” One complexity is that people disagree on which specific principles are relevant to morality (as opposed to matters of social convention or personal preference; Graham, Haidt, & Nosek, 2009; Graham, Meindl, Beall, Johnson, & Zhang, 2016; Meindl & Graham, 2014). However, it may be possible to justify defining behaviors such as helping and sympathy as being “moral” to the extent that they are indicative of prototypical moral virtues (e.g., compassion, fairness, honesty; Goodwin et al., 2014; Peterson & Seligman, 2004).

A second consideration is whether the morality of a given act depends on the accompanying thoughts, feelings, and motives. Empirical evidence shows that moral character attributions depend not only on what people do, but also the intentionality of their morally relevant actions (Malle, 2006), how quickly they come to their moral decisions (Critcher, Inbar, & Pizarro, 2013), the selflessness of their motives (Barasch, Levine, Berman, & Small, 2014; Newman & Cain, 2014), and whether they found it easy or struggled to do the right thing (Berman & Small, 2018; Starmans & Bloom, 2016). It may well be the case that the interpretation of behaviors depends much more on thoughts and feelings for moral traits than for non-moral traits (e.g., “talkativeness” can be straightforwardly interpreted as being an extraverted act without knowledge of what the actor was thinking and feeling).

Finally, based on Aristotle's doctrine of the golden mean (Aristotle, trans. 1985, p. 44), Ng and Tay (2020) have convincingly argued that moral virtue should be conceptualized as consistently expressing thoughts, feelings, and behaviors that optimal (i.e., neither deficient nor excessive) for a given situation. In this view, the relation between behavior and moral virtue is characterized by situation-specific optimality, rather than average behavioral tendencies (as is assumed by trait conceptualizations of moral character; e.g., Fleeson et al., 2014; Peterson & Seligman, 2004). It remains an empirical question as to whether accounting for situation-specific optimality better predicts virtue-relevant outcomes, compared to simply capturing average behavioral tendencies. Nevertheless, such theoretical complexities in defining what it means for an act to be moral show that thorough conceptual analysis is a crucial first step in the development of valid behavioral measures of morality.

Relative Effect Sizes for Personality Traits vs. Gender Differences

Another goal of this chapter was to explore gender differences in interpersonal behavior. Across all 18 behaviors, the mean absolute gender difference was $d = 0.20$. Gender explained an average of 1.27% of the variance across the full set of 18 behaviors and four factors; taking the square root of the unrounded average R^2 corresponds to an effect size of $r = .11$. This is larger than the average absolute correlation for any single personality trait (range = .04 to .07; see Table 15). In other words, if we have access to only one piece of information about a person, gender provides more information about overall differences in this set of interpersonal behaviors than does any single personality trait.

Several of these findings are consistent with either past work or what would be implied by hypothetical IPC locations. The finding that women spend a greater proportion of their conversations gossiping is consistent with Robbins and Karan's (2020) EAR study, which found

that women spent a greater proportion of their waking files gossiping neutrally (with point estimates also in the direction of women more frequently engaging in negative and positive gossip). Consistent with Kashdan and colleagues (2009), women also expressed gratitude more frequently than did men. The finding that women also tended to praise others more often is also consistent with the high warmth of this behavior. Men tended to be more condescending, consistent with the perspective that men would more frequently express high-dominance, low-warmth behaviors. Finally, although the gender difference in apologizing was not detectable, the point estimate was in the direction of women apologizing more (consistent with Schumann & Ross, 2010), with a 95% CI that just captured zero.

Other gender differences, or the lack thereof, were less expected. For example, it was less clear whether there would be gender differences in complaining and insecurity, which reflect a combination of low warmth and high dominance. Yet, women tended to manifest both of these behaviors more often than did men. There were also very few detectable differences in the behaviors that I hypothesized to be located in the cold-hearted and arrogant–calculating IPC octants. For example, although women self-report bragging less than men (Palmer et al., 2016), I found no evidence that they bragged less during naturalistic conversations. The point estimate was in the direction of men bragging more, but the wide 95% confidence interval indicates that we cannot rule out effects between $d = 0.15$ (women bragging more) and -0.35 (men bragging more). There was also no evidence to suggest that men were funnier in their everyday conversations, as would be implied by the perspective that humor ability is a fitness indicator that is more highly valued in a mate by women (Greengross & Miller, 2011). If anything, the point estimate was in the opposite direction. However, it should be noted that the EAR coders may have been interpreting the item, “Did or said something funny”, as including anything that

they personally found funny (e.g., unusual behaviors), rather than only expressions of humor.

Finally, the null gender difference for sympathy can likely be attributed to the low reliability of the behavioral codings for sympathy.

Little Evidence for Friend-Reported Social Consequences

Finally, there was very little evidence that these interpersonal behaviors had implications for friend-reported social consequences. The exceptions were that teasing was positively correlated with all three outcomes (likeability, being liked by friends, relationship quality), those who bragged more were liked less, those who complained more had better relationships with their friends, and those who were funnier were liked more by their friends. The unexpectedly positive interpersonal implications of teasing might suggest that teasing can be a warm and playful (rather than antagonistic) behavior in the context of intimate relationships. I was also surprised by the positive correlation between complaining and relationship quality, considering that previous studies implied that complainers would be disliked (Anderson, 1968; Lott et al., 1970). This could suggest reverse causality, such that people with better friendships feel more comfortable complaining out loud to their friends. However, given that none of these correlations survived a FDR correction and randomization tests suggested that this set of correlations could be spurious, these findings should not be over-interpreted.

Why were there so few correlations between these interpersonal behaviors and friend-reported social consequences? One possibility is that the way that participants treated all of the people they interacted with on average was not necessarily representative of how they treated the friends who provided ratings of the outcomes. For example, participants may have been especially sympathetic or helpful towards their friend informants. In addition, because these informants were friends, there was limited variability on the measures of likeability and liking,

which reduced the ability to detect behavioral correlates. Indeed, 73% of the liking scores were ≥ 6 on a 7-point scale. There was more variability on the likeability measure, but this was also highly skewed towards positive scores. Providing some support for the impact of limited variability, these interpersonal outcomes were also barely correlated with self-reported personality traits (see Table 14).

The broader issue that this highlights is that the use of friends as informants represents an important constraint on generality. Although it might be true that interpersonal behaviors expressed across all interactions may not be that strongly connected with how much a person is liked by their friends, these results cannot speak to how much these behaviors matter for being liked *in general*. To gain a more complete picture of the social consequences of bragging, condescension, gratitude, and other such behaviors, future studies should recruit relatively neutral informants who know the participant well but are not necessarily friends with them (e.g., co-workers, friends-of-friends), as well informants who dislike the participant (Leising, Erbs, & Fritz, 2010). An ideal study would also capture the identity of the interaction partners during each social interaction. This would reveal the social consequences of how a person treats specific interaction partners.

Chapter 6: Calibration of Arbitrary Metrics

In Chapter 5, I uncovered several correlations between personality traits and everyday interpersonal behaviors. The goal of this chapter is to make the implications of those correlations more concrete by undertaking a process called *metric calibration*, which involves mapping behavioral counts onto specific scores on personality trait measures.

Unlike physical concepts (e.g., height and weight), very few psychological measures provide direct information about a person's absolute standing on the underlying psychological construct (Blanton & Jaccard, 2006; Sechrest, McKnight, & McKnight, 1996). A *metric* refers to the numbering system that is used to describe a person's standing on the construct of interest. A metric is arbitrary when a given number on a scale provides no information about where a person lies on the underlying dimension, and when we do not know how much true change is reflected by a one-unit change on the scale. For example, the Big Five Inventory (BFI; John et al., 2008) measures agreeableness using nine questions answered on a 5-point scale, averaged together. We can infer that a person who scores "5" on the BFI agreeableness scale has a higher level of agreeableness than someone who scores "3". However, the scores themselves do not meaningfully locate a person on the underlying dimension of interest. For example, we cannot infer that a person who scores "5" is at the maximum possible level of the underlying agreeableness continuum, nor that a person who scores "3" (the midpoint of the scale) is at the true "neutral" point of agreeableness. Nor do the scores reveal how a person who scores "5" treats others differently compared to a person who scores "3".

It is, however, possible to reduce the arbitrariness of such metrics by mapping them onto meaningful external referents. I argue that the most meaningful external referents are real-world behaviors, and, in the case of traits that capture interpersonal phenomena, naturally-occurring

interpersonal behaviors. By examining, for example, how a 1-point difference in BFI agreeableness translates into differences in everyday interpersonal behaviors such as helping others and expressing sympathy, we can come to understand the concrete behavioral implications of scores on otherwise arbitrary metrics. This kind of metric calibration is crucial for understanding and communicating the behavioral implications of scores on widely-used scales and the practical, real-world implications of psychological effects. Yet, to date, very few researchers have explicitly attempted to do this.

Two notable exceptions come from work by Mehl (2007) and LeBel (2011). Using the EAR, Mehl and colleagues (2006) found that extraversion was correlated with time spent alone ($r = -.27$), and conscientiousness was correlated with time spent in class ($r = .42$). Unlike scores on an arbitrary Likert scale, EAR-derived behavioral counts are intuitively meaningful and directly interpretable. Thus, Mehl (2007) translated these correlations into intuitively interpretable effect sizes by reporting unstandardized regression equations. This metric calibration process showed that people who scored a “4” on the 5-point extraversion scale spent approximately 10% less time alone than those who scored a “2” (70.8% vs. 61.4%, respectively), and that those who scored a “4” on the 5-point conscientiousness scale spent three times more time in class than those who scored a “2” (11.9% vs. 4.1%). LeBel (2011) also provided meaningful behavioral mappings for measures of gregariousness, dutifulness, and self-control. He showed, for example, that a 1-unit increase (on a 5-point scale) in gregariousness translated into attending two more parties per month. However, unlike Mehl and colleagues’ work, which used actual behavioral counts derived from naturalistic audio recordings, LeBel relied on self-reports of these behavioral correlates.

As reviewed in Chapter 1, to the extent that interpersonal behaviors have been measured, such observations have typically occurred in lab-based settings. Research using the EAR has assessed a range of social behaviors, including many of the behaviors of interest in this dissertation (Bollich et al., 2016; Kaplan et al., 2018; Robbins & Karan, 2020), but has typically not explicitly mapped them onto specific scores on personality scales (cf. Mehl, 2007). Thus, the goal of this chapter is to provide a better understanding of how arbitrary scores on personality trait measures translate into differences in real-world interpersonal behaviors. This process will impart greater meaning to these otherwise arbitrary metrics.

Method and Results

I conducted metric calibration analyses for the correlations between traits and individual behaviors that were detectable under a liberal $p < .05$ threshold in Chapter 5 (see Table 15). To do so, I used unstandardized regression models to estimate the frequencies of these behaviors for participants who scored at various points of each scale (see Table 18). Most of the original measures were on a 1–15 scale, except for narcissism (scored as a proportion) and religiosity (1–7 scale). To facilitate comparability across scales, I rescaled all trait measures to a 0–10 scale. For example, if the original measure was on a 1–15 scale, a score of 1 would be recoded as 0, a score of 8 would be recoded as 5, and a score of 15 would be recoded as 10. I report the estimated frequencies of these behaviors at the rescaled minimum, midpoint, and maximum points of the scale. Because participants did not always use the full range of each scale, I also report the estimated frequencies of these behaviors for the distribution of scores (lowest, median, highest, and 1 SD above and below the mean) that were observed in this sample. This provides a more realistic sense of the behavioral variations that are likely to be observed in practice. I report all results in Table 18, and describe and visualize a few illustrative results below (see Figure 8).

Table 18
Calibrating Trait Measures to Specific Interpersonal Behaviors

Trait and correlate	<i>b</i>	95% CI	Estimated Frequencies (%)							
			Full Range of Scale			Observed Scores				
			Min.	Mid.	Max.	Min.	−1 SD	Median	+1 SD	Max.
Extraversion						(0.36)	(3.93)	(6.07)	(7.99)	(9.82)
Brag	0.09	[0.01, 0.17]	0.74	1.19	1.64	0.77	1.09	1.29	1.46	1.63
Tease	0.20	[0.04, 0.37]	2.42	3.44	4.46	2.50	3.22	3.66	4.05	4.42
Negative gossip	0.19	[0.01, 0.37]	2.87	3.81	4.75	2.94	3.61	4.01	4.37	4.72
Gratitude	0.17	[0.02, 0.32]	1.51	2.35	3.19	1.57	2.17	2.53	2.85	3.16
Agreeableness						(0.79)	(5.08)	(6.75)	(8.19)	(9.82)
Condensation	−0.23	[−0.41, −0.05]	3.74	2.59	1.44	3.56	2.58	2.19	1.86	1.49
Sarcasm	−0.22	[−0.41, −0.03]	4.35	3.24	2.14	4.17	3.23	2.86	2.54	2.18
Funny	−0.51	[−0.96, −0.06]	14.45	11.9	9.35	14.05	11.86	11.01	10.28	9.44
Narcissism						(0)	(1.30)	(3.12)	(4.92)	(8.75)
Condensation	0.17	[0.01, 0.33]	1.70	2.53	3.37	1.70	1.91	2.22	2.52	3.16
Brag	0.10	[+0.00, 0.19]	0.98	1.46	1.94	0.98	1.10	1.28	1.45	1.82
Tease	0.22	[0.03, 0.40]	2.96	4.04	5.12	2.96	3.24	3.64	4.03	4.85
Morality (self)						(0.71)	(5.39)	(7.86)	(9.72)	(10)
Blame	0.06	[0.01, 0.12]	0.53	0.85	1.17	0.57	0.87	1.03	1.15	1.17
Apologize	0.08	[0.01, 0.16]	0.44	0.86	1.27	0.50	0.89	1.09	1.25	1.27
Morality (informant)						(1.43)	(6.21)	(7.86)	(9.23)	(10)
Condensation	−0.35	[−0.55, −0.15]	4.88	3.14	1.39	4.38	2.72	2.14	1.66	1.39
Brag	−0.12	[−0.24, −0.00]	2.13	1.54	0.95	1.97	1.40	1.20	1.04	0.95
Conscientiousness						(0.79)	(4.61)	(6.19)	(7.89)	(9.68)
Hope/optimism	0.46	[0.02, 0.91]	5.11	7.44	9.76	5.48	7.26	7.99	8.78	9.61
Praise/compliment	0.30	[0.06, 0.55]	3.48	5.00	6.52	3.72	4.88	5.36	5.88	6.42
Gratitude	0.21	[0.02, 0.39]	1.21	2.25	3.29	1.38	2.17	2.50	2.85	3.22
Neuroticism						(0)	(2.89)	(4.64)	(6.65)	(9.64)
Insecurity	0.30	[0.12, 0.47]	1.96	3.44	4.91	1.96	2.81	3.33	3.92	4.81
Apologize	0.09	[+0.00, 0.18]	0.63	1.09	1.54	0.63	0.90	1.06	1.24	1.51
Openness						(2.07)	(5.28)	(6.79)	(8.11)	(9.5)
Tease	−0.33	[−0.57, −0.10]	5.87	4.20	2.53	5.18	4.11	3.61	3.16	2.70

(table continues)

Table 18 (*continued*)

Trait and correlate	<i>b</i>	95% CI	Estimated Frequencies (%)							
			Full Range of Scale			Observed Scores				
			Min.	Mid.	Max.	Min.	−1 SD	Median	+1 SD	Max.
Self-esteem						(0)	(2.99)	(6.43)	(8.96)	(10)
Insecurity	−0.15	[−0.25, −0.04]	4.24	3.51	2.78	4.24	3.80	3.30	2.93	2.78
Life satisfaction						(0.71)	(5.43)	(7.86)	(9.53)	(10)
Complain	−0.38	[−0.67, −0.08]	11.76	9.88	7.99	11.49	9.72	8.80	8.17	7.99
Brag	0.09	[0.01, 0.18]	0.58	1.05	1.51	0.65	1.09	1.31	1.47	1.51
Insecurity	−0.30	[−0.46, −0.15]	5.63	4.12	2.61	5.42	3.99	3.25	2.75	2.61
Religiosity						(0)	(0.43)	(3.33)	(6.37)	(10)
Sympathy	−0.27	[−0.53, −0.02]	10.19	8.84	7.48	10.19	10.08	9.29	8.47	7.48
Funny (self)						(0)	(4.99)	(7.14)	(9.31)	(10)
Tease	0.21	[0.05, 0.36]	2.14	3.19	4.23	2.14	3.18	3.63	4.09	4.23
Insecurity	−0.17	[−0.32, −0.02]	4.59	3.74	2.88	4.59	3.74	3.37	3.00	2.88
Funny (informant)						(0.71)	(6.56)	(8.21)	(9.44)	(10)
Tease	0.26	[0.02, 0.51]	1.53	2.84	4.15	1.71	3.25	3.68	4.00	4.15
Funny	0.59	[0.09, 1.09]	6.36	9.32	12.28	6.79	10.25	11.23	11.95	12.28

Note. The unstandardized regression coefficient *b* represents the % increase in conversations that contained these behaviors for every 1 unit increase on the trait measure. All trait measures were rescaled to a 0–10 scale to facilitate comparability of effect sizes across trait measures. Estimated frequencies of behaviors as a percentage of conversations are based on the point estimate of the *b* coefficient for participants scoring at the lowest point (0), midpoint (5), or highest point (10) on the scale (“Full Range of Scale”), or the lowest, 1 SD below the mean (−1 SD), median, 1 SD above the mean (+1 SD), and maximum scores observed in this sample (“Observed Scores”). Values in parentheses on the same row as the trait labels represent the lowest, −1 SD, median, +1 SD, and maximum scores on each measure. Min. = minimum of the scale or minimum observed scores, Mid. = scale midpoint, Max. = maximum of the scale or maximum observed scores.

Figure 8. Examples of metric calibration of four personality measures against four behaviors, as a percentage of everyday conversations. Dashed lines indicate the estimated behavioral frequencies for participants at the highest possible point on the scale (for agreeableness). Dotted lines indicate the estimated behavioral frequencies for participants one SD above and below the mean (for narcissism and life satisfaction). Dotted–dashed lines indicate the estimated behavioral frequencies for participants who had the highest or lowest scores in this sample (for informant-reported humor). Gray ribbons show 95% confidence bands around the estimates.

These unstandardized regression estimates predict, for example, that people who are maximally disagreeable on the BFI (who express sarcasm in 4.35% of conversations) are, on average, twice as sarcastic as those who are maximally agreeable on the BFI (who express sarcasm in 2.14% of conversations). Those at the midpoint of the Narcissistic Personality Inventory bragged during 1.46% of their conversations on average. Each 1-unit increase in narcissism (on a 0–10 scale) translated into a predicted 0.10% increase in bragging, and those who had narcissism scores 1 SD above or below the mean bragged during 1.1% vs. 1.45% of

conversations on average, respectively. As another example, each 1-unit increase in life satisfaction predicted a 0.38% decrease in the number of conversations in which a person complained, and participants with life satisfaction scores 1 SD above or below the mean complained during 8.17% vs. 9.72% of their conversations on average, respectively. Finally, in this sample, those who were rated by their friends as being the most vs. least funny people did something funny during 12.28% vs. 6.78% of their conversations, respectively. These results bring trait-behavior correlations to life by translating scores on otherwise arbitrary metrics into concrete behavioral frequencies.

Discussion

The goal of this chapter was to undertake metric calibration in order to paint a more precise behavioral portrait of different scores on several personality trait measures. Whereas zero-order correlations can tell us that agreeable people tend to be less condescending ($r = -.14$), less sarcastic ($r = -.13$), and less funny ($r = -.13$), metric calibration can translate these correlations into more readily-interpretable behavioral frequencies that correspond to different agreeableness scores. For example, using this process, I showed that maximally disagreeable vs. maximally agreeable people (on the BFI measure) were predicted to be condescending in 3.74% vs. 1.44% of their conversations, sarcastic in 4.35% vs. 2.14% of their conversations, and funny in 14.45% vs. 9.35% of their conversations (on average). This provides much more information about how dramatic or modest the behavioral differences between the most or least agreeable people (according to a specific scale) are. Thus, these results illustrate the potential for metric calibration to imbue arbitrary metrics with more meaning by clarifying the tangible interpersonal behavioral implications of high and low scores on otherwise arbitrary metrics.

However, these results should be interpreted with the following limitations in mind. First, the correlations tended to be small, which means that there is a lot of behavioral variability among people who have the same score on a personality trait measure (see Figure 8). Such variability is to some extent inherent in the probabilistic nature of predicting behaviors from traits. Measurement error due to the tendency for people to interpret and respond to arbitrary scales in different ways (e.g., acquiescence and extreme response biases) further contributes to such variability. Thus, metric calibration is well-suited for revealing the behavioral implications of different scores on average, but less well-suited for predicting the behavior of a specific person, at least when using nomothetic (vs. person-specific) models.

In addition, even these noisily estimated behavioral frequencies are merely point estimates that had relatively wide 95% confidence intervals (see Table 18). For example, the difference in the percentage of boastful conversations between those who have the maximum vs. minimum scores on the BFI extraversion measure could be as large as 1.7% (if $b = 0.17$) or as small as 0.07% (if $b = 0.01$). Larger sample sizes would be required to achieve more precise estimates of average behavioral frequencies for each scale point.

A further limitation is that the estimated frequencies may be specific to North American college students, or even this sample. Thus, these frequencies should not be interpreted too literally as being definitive and generalizable estimates of the behavioral implications of specific scores on various traits. Finally, as described in Chapter 5, these correlates were discovered using an exploratory procedure that involved computing 234 correlations between traits and individual behaviors (or 286 correlations if we include correlations with the behavioral composites), of which only the negative association between life satisfaction and expressions of insecurity survived a FDR correction. Thus, these associations await future replication.

Chapter 7: General Discussion

People are defined in large part by how they relate to and interact with others. Yet, relatively little research has examined what people actually do to each other during their everyday social interactions. The goal of this dissertation was to bridge this gap and provide an in-depth, exploratory, and descriptive investigation of individual differences in 18 interpersonal behaviors (pessimism, criticism, complaining, condescension, blame, sympathy, optimism, helping, praise, bragging, sarcasm, teasing, negative gossip, neutral or positive gossip, insecurity, being funny, gratitude, and apologizing). In Table 19, I provide an overall summary of the results.

Overview of the Findings

Factor analytic methods have been central to discovering the structure of personality traits by uncovering latent factors that can explain variation in individual items on self-report measures (Ashton et al., 2004; John et al., 2008). Chapter 3 explored the factor structure of 18 interpersonal behaviors with two purposes in mind. The first purpose was to derive broader behavioral composites to complement analyses of individual behaviors. The second purpose was to generate hypotheses about the more general underlying tendencies that specific interpersonal behaviors might share. I found that 14 of the behaviors might reflect broader tendencies towards negativity (pessimism, criticism, complaining, condescension, and blame), supportiveness (sympathy, optimism, helping, and praise), mockery (bragging, sarcasm, and teasing), and gossip (negative gossip and neutral or positive gossip). Although these composites were reasonably reliable, the CFA model did not fit well and subsequent analyses in Chapters 4 and 5 provided little evidence for the validity and utility of these composites.

Table 19

Overall Summary of Psychometric Properties and Detectable Trait Correlates, Gender Differences, and Social Consequences

Composite or Behavior	Mean ω	r (Year 1, Year 2)	Trait Correlates	Gender Difs. (d)	Social Consequences
Negativity	.73	.09			
Pessimism	.18	-.12			
Criticism	.43	.18	Funny (.15*)		
Complain	.52	.17	SWL (-.15*)	W > M (0.29*)	RQ (.15*)
Condescension	.31	.21*	A (-.14*), Narc. (.12*), Morality [I] (-.21***)	M > W (-0.26*)	
Blame	.24	-.06	Morality [S] (.13*)		
Supportiveness	.72	.08	Morality [S] (.13*), C (.13*)		
Sympathy	.05	-.10	Religiosity (-.14*)		
Hope/optimism	.12	.12	C (.12*)		
Help	.36	.27**			
Praise/compliment	.56	.04	C (.14*)	W > M (0.50***)	
Mockery	.67	.28**	Narc. (.13*), Funny [S] (.15*)		
Brag	.30	.11	E (.12*), Narc. (.12*), Morality [I] (-.12*), SWL (.13*)		Liked (-.13*)
Sarcasm	.31	.26**	A (-.13*)		
Tease	.38	.27**	E (.14*), Narc. (.13*), O (-.16**), Funny [S/I] (.16**/.13*)		Likeability (.18**), Liked (.16*), RQ (.17**)
Gossip	.50	.26**	E (.13*)	W > M (0.36**)	
Negative gossip	.57	.18*	E (.12*)	W > M (0.36**)	
Neu./pos. gossip	.53	.28**		W > M (0.28*)	
No factor					
Insecurity	.25	.17	N (.20***), Self-esteem (-.15**), SWL (-.22***), Funny [S] (-.13*)	W > M (0.46***)	
Funny	.41	.09	A (-.13*), Funny [I] (.15*)		Liked (.15*)
Gratitude	.77	.16	E (.13*), C (.13*)	W > M (0.39**)	
Apologize	.73	.21*	Morality [S] (.13*), N (.12*)		

Note. Mean ω = omega reliability averaged across two years. * $p < .05$, ** $p < .01$, *** $p < .001$, not corrected for multiple comparisons. E = extraversion, A = agreeableness, Narc. = narcissism, C = conscientiousness, N = neuroticism, O = openness, SWL = satisfaction with life, [S] = self-reported, [I] = informant-reported, W = women, M = men, Liked = liked by informants. RQ = relationship quality.

The purpose of Chapter 4 was to estimate the rank-order stability of these interpersonal behaviors across one year. This addresses one of the most central issues in personality psychology—behavioral consistency—by testing the extent to which a person’s interpersonal behavior across one week predicts how they behave one year later. Across the set of 18 behaviors, the average rank-order stability was low ($r = .13$), but 7 behaviors (condescension, helping, sarcasm, teasing, negative gossip, neutral or positive gossip, apologizing) and the mockery and gossip factors showed moderate and detectable levels of stability in the range of $r = .18$ – $.28$. Whereas rank-order stability analyses of self-report measures abound (Roberts & DelVecchio, 2000), far fewer studies have examined the rank-order stability of actual behavior observed in everyday life (cf. Bollich et al., 2016; Harari et al., 2019; Mehl & Pennebaker, 2003). Moreover, to my knowledge, this is the first study to test the rank-order stability of naturalistically-observed social behaviors across such a long retest period. Thus, these findings suggest that people show some temporal stability in everyday interpersonal behaviors and shed further light on the psychometric properties of EAR-coded interpersonal behaviors.

The purpose of Chapter 5 was to explore the correlates and social consequences of interpersonal behaviors. First, I examined the interpersonal behavioral implications of personality traits. These analyses revealed some theoretically-consistent correlations, some unexpected null correlations, and some hypothesis-generating correlations (which I discussed in Chapter 5 and will further discuss below). Second, I provided a strong test of gender differences in actual interpersonal behavior. These analyses showed that there were detectable gender differences in eight interpersonal behaviors and that the average effect size for gender differences was larger than the average effect size of any given personality trait. Finally, there

was little evidence that interpersonal behaviors had implications for being seen as likeable, being liked, and having a good relationship with one's friends.

Finally, the purpose of Chapter 6 was to demonstrate how behavioral assessment can reveal the concrete behavioral implications of otherwise arbitrary scores on personality trait measures. To do so, I undertook metric calibration by mapping specific frequencies of behaviors onto specific scores on personality trait measures. This process painted more concrete behavioral portraits of different trait scores by showing, for example, that the least vs. most agreeable participants in this sample expressed condescension during 3.56% vs. 1.49% of their conversations on average.

Taken together, these results illustrate the unique value and challenges of assessing interpersonal behavior in real-world contexts. In the following sections, I discuss the contributions and challenges of observing actual interpersonal behavior, the complexities of establishing connections between traits and behaviors, what we have learned about everyday interpersonal behaviors, and limitations of the current dissertation.

Contributions of Observing Actual Interpersonal Behavior

Throughout this dissertation, I have aimed to illustrate the unique contributions of capturing actual interpersonal behavior in everyday life. Given that personality traits describe stable patterns of thoughts, feelings, and *behaviors*, a key purpose of Chapter 5 was to clarify the behavioral implications of personality traits. This revealed several behavioral correlates that are consistent with existing conceptualizations of traits. For example, agreeable people were less sarcastic and condescending, neurotic people tended to express more insecurity, people who were satisfied with their lives tended to complain less, and funny people (as rated by friends) tended to tease others more and were funnier in everyday life. In doing so, these results provided

reassuring evidence of the construct validity and behavioral relevance of self-report measures of personality traits—a goal that could only be accomplished by measuring actual behavior.

Demonstrating distinct behavioral profiles can also help clarify the distinctions between conceptually-neighboring constructs. For example, although self-esteem and narcissism are both characterized by positive self-views, they had distinct detectable behavioral profiles.

Specifically, people who were higher on self-esteem were less likely to express insecurity, whereas narcissists tended to be more condescending and boastful and more frequently mocked and teased others. This is consistent with findings that demonstrate that narcissism (unlike genuine self-esteem) is uniquely associated with aggression and antisocial behaviors (Tracy, Cheng, Robins, & Trzesniewski, 2009).

A further contribution of measuring actual behavior is that unlike scores on arbitrary metrics (e.g., a Likert-type scale from 1 = *strongly disagree* to 5 = *strongly agree*), behavioral counts are non-arbitrary. This provides the opportunity to further illuminate the concrete behavioral implications of otherwise arbitrary scores on personality trait measures. This was the goal of Chapter 6, which translated the trait–behavior correlations from Chapter 5 into concrete behavioral frequencies that would be expected for people with different scores on several personality trait measures. For example, each 10% increase on the Narcissistic Personality Inventory translated into a predicted 0.10% increase in bragging (as a percentage of everyday conversations). In doing so, Chapter 6 clarifies what it means to say that someone is highly narcissistic (vs. low on narcissism), in terms of concrete behavioral differences.

Complexities of Establishing Connections Between Traits and Behaviors

Although the trait–behavior associations that emerged were, for the most part, consistent with existing conceptualizations of the traits, the effect sizes tended to be small, ranging from

|.12–.22| for the detectable effects. Several unexpected null effects also emerged. For example, surprisingly few interpersonal behaviors were detectably associated with agreeableness, even under a liberal threshold for exploratory effects. There are many plausible and non-mutually-exclusive explanations for these weak or null correlations, including low coding reliability, lack of self-knowledge, and the importance of situational factors (which I considered in Chapter 5). Here, I would like to discuss two additional possibilities: that 1) traits are not straightforwardly manifested as behavior, but rather as behavior accompanied by the right thoughts and feelings, and 2) behavioral manifestations of traits are highly context-dependent.

A general limitation of behavioral assessment is that it does not capture the thoughts and feelings surrounding the behavior, which the self has privileged access to. One possibility is that the relation between traits and behaviors is more complex—perhaps traits may not be connected with raw behavior, but instead predict the intentions underlying the behavior and the feelings associated with the behavior. For example, agreeable and disagreeable people might be equally helpful, but for different reasons (e.g., genuine care vs. expectations of future reciprocity) and with different accompanying affective experiences (e.g., warm glow vs. resentment).

To test this idea, future studies could use multi-method approaches that combine behavioral assessment with self-reports of thoughts and feelings. This approach could also be used to more generally illuminate the connections between the ABCs of personality in everyday life (e.g., Sun, Stevenson, Kabbani, Richardson, & Smillie, 2017). For example, what are the behavioral implications of specific thoughts and feelings? Are thoughts and feelings typically manifested as behavior, or are they more often behaviorally inert? Is an agreeable person one who rarely has antagonistic thoughts and feelings about others in the first place, or is agreeableness better conceptualized as the ability to not express disagreeable thoughts and

feelings when they occur (Kammrath & Scholer, 2011)? Answers to such questions would clarify the roles of thoughts, feelings, and behaviors in the conceptualization of interpersonal traits.

A second issue is that rather than being unconditional and decontextualized (McAdams & Pals, 2006), traits might require appropriate eliciting stimuli in order to be manifested in behavior (Allport, 1937; DeYoung, 2015; Tellegen, 1981). For example, in any given interaction, not expressing gratitude may either be appropriate (because there was no benefactor) or indicative of rudeness (if a person *failed* to express gratitude in response to a favor). Without the assessment of situational factors, there is no way to distinguish between non-acts that reflect a lack of relevant eliciting stimuli vs. meaningful omissions. Thus, by averaging across all interactions without considering context, average levels of gratitude expressions might reflect a combination of dispositional tendencies as well as the frequency with which a person is in a situation in which expressing gratitude would be appropriate. This might explain some of the low correlations between traits and interpersonal behaviors.

Similarly, the correlations between interpersonal behaviors and friend-reported social consequences might have been stronger if the behavioral composites specifically reflected how the participants treated their informants (rather than their average behavior across all social interactions). Ideally, a comprehensive study of the connections between traits, behaviors, and social consequences should assess additional contextual factors such as the identity and behaviors of the interaction partner, thereby accounting for all three components of the personality triad (persons, situations, and behavior; Funder, 2006).

What Have We Learned About Interpersonal Behaviors?

Examining trait correlates and social consequences can also provide insight into the psychological meaning of everyday interpersonal behaviors. For example, the finding that

teasing was associated with likeability, being liked by one's friends, relationship quality, and the trait of being funny suggests that it might be a warm and playful behavior (rather than being necessarily antagonistic and mean-spirited).

Each behavior also had a distinct profile of trait correlates. For example, it was not the case that behaviors that loaded onto the same factor (e.g., criticism and blame) had the same trait correlates. This, along with the limited utility of the behavioral composites of negativity, supportiveness, mockery, and gossip, suggests that each behavior is conceptually distinct and that specific behaviors might be a more useful level of analysis (at least compared to the four behavioral composites examined here).

In some cases, behaviors were associated with multiple traits. For example, condescension's negative correlations with agreeableness and informant-reported morality and positive correlation with narcissism provide convergent evidence for the highly antagonistic nature of this behavior. In contrast, the presence of multiple trait correlates suggests that the same behavior might reflect different motives. For example, bragging was positively correlated with extraversion, narcissism, and satisfaction with life, and negatively correlated with informant-reported morality. Thus, among other reasons, bragging might stem from power-seeking motives (in the case of narcissism) or the desire to share one's satisfaction with life.

Finally, these results point to open questions about the psychological meaning of helping. Although individual differences in helping showed stability across a one-year retest period, helping was not detectably correlated with any of the traits that I examined. One possibility is that "helping" is too broad to be a psychologically useful category. This is because there are many different styles of supportive actions that can be mapped onto various points of the IPC (Trobst, 2000). For example, people can help in directive (offering advice while emphasizing

one's own competence and superiority), arrogant (taking control, making decisions for the recipient, and persuading them to change their behavior), nurturant (providing emotional support, affection, and patience), and engaging (showing enthusiasm, providing useful information, and expressing concern) ways (Trobst, 2000). Thus, it might be more useful to code specific helpful actions such as offering advice, offering to listen, providing practical support, persuading someone to change their behavior, and attempting to regulate someone else's emotions (Zaki & Williams, 2013).

Challenges of Measuring Actual Behavior

Despite the efforts of 158 coders across multiple years of coding, very few behaviors were assessed with adequate reliability ($> .50$) across both years (gratitude, apologizing, negative gossip), and coders found it particularly challenging to judge which participants more frequently expressed sympathy, optimism, and pessimism. Even with an average of three coders per participant, reliabilities for sympathy, optimism, sarcasm, bragging, condescension, blame, pessimism, teasing, and insecurity were $< .50$ in Year 1 and even lower in Year 2. Whereas the reasonable pattern of trait correlates for complaining, sarcasm, bragging, condescension, teasing, insecurity, and being funny suggests that these codings still had some validity, I interpret the relative lack of trait correlates for pessimism, sympathy, optimism, helping, criticism, and blame as suggesting that these codings contained little valid variance. The low inter-coder reliability of several of the interpersonal behaviors complicates the interpretation of the effect sizes and null results for the rank-order stability and substantive analyses.

This illustrates a major challenge of measuring interpersonal behavior in the real world, and specifically the kinds of interpersonal behaviors that require more subjective judgments but are arguably more interesting than micro-level behaviors such as whether or not the participant

talked or not. These results provide a sense of which behaviors can be reliably and validly assessed using behavioral observation methods such as the EAR, and the number of coders and observations that are required to achieve even modest levels of reliability. Given the low base rates of these interpersonal behaviors, an even greater number of days (e.g., two weeks) might be needed to reliably estimate between-person differences. Future research should also explore whether there are ways to improve the reliability of these kinds of behavioral codings (e.g., more precise definitions and intensive training).

Limitations

I have already discussed two issues that I cannot address with the current data: the possibilities that rather than straightforwardly predicting average behavior, traits might a) predict the thoughts and feelings that accompany behavior and b) only predict behavior in the presence of relevant eliciting stimuli. I have also discussed the fact that low inter-coder reliabilities complicate the interpretation of null results. A further limitation is that despite assessing a fairly large number of behaviors, the coding scheme was not systematically designed to comprehensively capture all interpersonal behaviors that might be of interest. There are many other potentially audible, frequent, and interesting interpersonal behaviors that were not measured in this study (e.g., expressing enthusiasm, affection, arguing, giving in, forgiving, making demands, asserting oneself, asking for help or advice, flirting), as well as many other interpersonal behaviors that would be difficult to assess using the EAR (e.g., lying, cheating, being manipulative). Some octants of the IPC were better represented (arrogant–calculating, cold-hearted, and aloof–introverted) than others (assured–dominant, unassured–submissive, unassuming–ingenuous) in the current scheme. An ideal coding scheme would include behaviors

that correspond to each octant of the IPC and each facet of major interpersonal traits such as extraversion, agreeableness, and honesty-humility.

I also chose to limit the scope of this dissertation to only consider the percentage of conversations in which each interpersonal behavior was expressed. This approach ensured that differences in interpersonal behaviors were not merely due to differences in the number of conversations that participants engaged in. A complementary approach would have been to consider the percentage of all waking files in which each interpersonal behavior was expressed. This would provide a better sense of the absolute frequency with which people engaged in each interpersonal behavior. After all, people do not merely find themselves in conversations; people can also proactively initiate conversations for the sole purpose of gossiping, criticizing others, or expressing sympathy. For example, a person who helps during a smaller percentage of their conversations might still be more frequently helpful on an absolute metric than a person who helps during a larger percentage of their conversations but rarely engages in conversations in the first place. Thus, the conclusions in this dissertation are constrained to operationalizations of average interpersonal behavior as a percentage of everyday conversations.

Finally, I took an exploratory, hypothesis-generating (rather than a confirmatory, hypothesis-driven) approach in this dissertation. I ran a very large number of analyses (e.g., 286 correlations between personality traits and interpersonal behaviors), and very few survived a FDR correction. Although randomization tests provided support for the broad conclusions that personality traits are associated with interpersonal behavior and that there are gender differences in interpersonal behavior, these analyses do not speak to which specific associations are robust. Moreover, the conclusions in this dissertation are based on one sample of young adults at a private, selective North American university. This limits my ability to generalize the current

findings to other age groups and especially to non-Western cultures. Thus, interesting patterns that emerge from this dissertation should be treated as hypotheses that await tests of replication and generalizability, rather than robust findings.

Conclusion

Personality is far from personal. Instead, some of the most important individual differences are those that describe what people do to each other. A scientific understanding of personality would thus be incomplete without understanding individual differences in interpersonal behaviors. This dissertation represents an effort to integrate personality and social psychology by painting several behavioral portraits of personality traits and exploring the social consequences of everyday interpersonal behavior. In doing so, this dissertation expands current understandings of personality traits and everyday interpersonal behaviors and illustrates the unique value and challenges of measuring actual behavior.

References

- Algoe, S. B. (2012). Find, remind, and bind: The functions of gratitude in everyday relationships. *Social and Personality Psychology Compass*, 6(6), 455–469. <https://doi.org/10.1111/j.1751-9004.2012.00439.x>
- Allport, F. H., & Allport, G. W. (1921). Personality traits: Their classification and measurement. *The Journal of Abnormal Psychology and Social Psychology*, 16(1), 6–40. <https://doi.org/10.1037/h0069790>
- Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: Holt and Company.
- Ames, D. R., Rose, P., & Anderson, C. P. (2006). The NPI-16 as a short measure of narcissism. *Journal of Research in Personality*, 40(4), 440–450. <https://doi.org/10.1016/j.jrp.2005.03.002>
- Anderson, N. H. (1968). Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology*, 9(3), 272–279. <https://doi.org/10.1037/h0025907>
- Ansell, E. B., & Pincus, A. L. (2004). Interpersonal perceptions of the five-factor model of personality: An examination using the structural summary method for circumplex data. *Multivariate Behavioral Research*, 39(2), 167–201. https://doi.org/10.1207/s15327906mbr3902_3
- Aristotle. (1985). *Nicomachean ethics*. (T. Irwin, Trans.). Indianapolis, IN: Hackett Publishing.
- Asch, S. E. (1987). *Social psychology*. New York: Oxford University Press. (Original work published 1952).
- Ashton, M. C., Jackson, D. N., Paunonen, S. V., Helmes, E., & Rothstein, M. G. (1995). The criterion validity of broad factor scales versus specific facet scales. *Journal of Research in*

- Personality*, 29(4), 432–442. <https://doi.org/10.1006/JRPE.1995.1025>
- 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. <https://doi.org/10.1177/1088868306294907>
- 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. <https://doi.org/10.1177/1088868314523838>
- 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. <https://doi.org/10.1016/j.paid.2006.10.027>
- Ashton, M. C., Lee, K., Perugini, M., Szarota, P., de Vries, R. E., Di Blas, L., ... De Raad, B. (2004). A six-factor structure of personality-descriptive adjectives: Solutions from psycholexical studies in seven languages. *Journal of Personality and Social Psychology*, 86(2), 356–366. <https://doi.org/10.1037/0022-3514.86.2.356>
- Back, M. D. (in press). Social interaction processes and personality. In J. F. Rauthmann (Ed.), *The handbook of personality dynamics and processes*. Elsevier.
- Back, M. D., Schmukle, S. C., & Egloff, B. (2010). Why are narcissists so charming at first sight? Decoding the narcissism–popularity link at zero acquaintance. *Journal of Personality and Social Psychology*, 98(1), 132–145. <https://doi.org/10.1037/a0016338>
- Back, M. D., & Vazire, S. (2015). The social consequences of personality: Six suggestions for future research. *European Journal of Personality*, 29(2), 296–307. <https://doi.org/10.1002/per.1998>
- Barasch, A., Levine, E. E., Berman, J. Z., & Small, D. A. (2014). Selfish or selfless? On the

- signal value of emotion in altruistic behavior. *Journal of Personality and Social Psychology*, 107(3), 393–413. <https://doi.org/10.1037/a0037207>
- 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. <https://doi.org/10.1016/j.paid.2015.05.038>
- Barker, R. G., & Wright, H. F. (1951). *One boy's day: A specimen record of behavior*. New York: Harper & Row.
- Barranti, M. (2019). *Shared reality or shared illusions? Evaluating moral impressions* (Unpublished doctoral dissertation). University of Toronto, Canada.
- Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, 2(4), 396–403. <https://doi.org/10.1111/j.1745-6916.2007.00051.x>
- Bem, D. J. (1972). Self-perception theory. In *Advances in experimental social psychology* (pp. 183–200). New York: Academic Press.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*. <https://doi.org/10.2307/2346101>
- Berman, J. Z., & Small, D. A. (2018). Discipline and desire: On the relative importance of willpower and purity in signaling virtue. *Journal of Experimental Social Psychology*, 76(May 2017), 220–230. <https://doi.org/10.1016/j.jesp.2018.02.007>
- Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *The American Psychologist*, 61(1), 27–41. <https://doi.org/10.1037/0003-066X.61.1.62>
- Block, J. (1960). On the number of significant findings to be expected by chance.

- Psychometrika*, 25(4), 369–380. <https://doi.org/10.1007/BF02289754>
- Block, J. (1989). Critique of the act frequency approach to personality. *Journal of Personality and Social Psychology*, 56(2), 234–245. <https://doi.org/10.1037/0022-3514.56.2.234>
- Bollich, K. L., Doris, J. M., Vazire, S., Raison, C. L., Jackson, J. J., & Mehl, M. R. (2016). Eavesdropping on character: Assessing everyday moral behaviors. *Journal of Research in Personality*, 61, 15–21. <https://doi.org/10.1016/j.jrp.2015.12.003>
- Borkenau, P., & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology*, 62(4), 645–657. <https://doi.org/10.1037/0022-3514.62.4.645>
- Botwin, M. D., Buss, D. M., & Shackelford, T. K. (1997). Personality and mate preferences: Five factors in mate selection and marital satisfaction. *Journal of Personality*, 65(1), 107–136. <https://doi.org/10.1111/j.1467-6494.1997.tb00531.x>
- Bradlee, P. M., & Emmons, R. A. (1992). Locating narcissism within the interpersonal circumplex and the five-factor model. *Personality and Individual Differences*, 13(7), 821–830. [https://doi.org/10.1016/0191-8869\(92\)90056-U](https://doi.org/10.1016/0191-8869(92)90056-U)
- Brown, N. A., Blake, A. B., & Sherman, R. A. (2017). A snapshot of the life as lived: Wearable cameras in social and personality psychological science. *Social Psychological and Personality Science*, 8(5), 592–600. <https://doi.org/10.1177/1948550617703170>
- Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley: University of California Press.
- Buss, D. M., & Chiodo, L. M. (1991). Narcissistic acts in everyday life. *Journal of Personality*, 59(2), 179–215. <https://doi.org/10.1111/j.1467-6494.1991.tb00773.x>
- Buss, D. M., & Craik, K. H. (1980). The frequency concept of disposition: dominance and

- prototypically dominant acts. *Journal of Personality*, 48(3), 379–392.
<https://doi.org/10.1111/j.1467-6494.1980.tb00840.x>
- Buss, D. M., & Craik, K. H. (1983). Act frequency approach to personality. *American Psychologist*, 20(2), 713–713. <https://doi.org/10.1037/h0021465>
- Campbell, W. K., Rudich, E. A., & Sedikides, C. (2002). Narcissism, self-esteem, and the positivity of self-views: Two portraits of self-love. *Personality and Social Psychology Bulletin*, 28(3), 358–368. <https://doi.org/10.1177/0146167202286007>
- Carey, A. L., Brucks, M. S., Küfner, A. C. P., Holtzman, N. S., große Deters, F., Back, M. D., ... Mehl, M. R. (2015). Narcissism and the use of personal pronouns revisited. *Journal of Personality and Social Psychology*, 109(3), e1–e15. <https://doi.org/10.1037/pspp0000029>
- Carpenter, J., Connolly, C., & Myers, C. K. (2008). Altruistic behavior in a representative dictator experiment. *Experimental Economics*, 11(3), 282–298.
<https://doi.org/10.1007/s10683-007-9193-x>
- Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Anandkumar, A., ... De Rosario, H. (2017). pwr: Basic functions for power analysis (Version 1.3) [R package]. Retrieved from <https://CRAN.R-project.org/package=pwr>
- Chandler, J. (2018). Likeableness and meaningfulness ratings of 555 (+487) person-descriptive words. *Journal of Research in Personality*, 72, 50–57.
<https://doi.org/10.1016/j.jrp.2016.07.005>
- Chapman, B. P., & Goldberg, L. R. (2017). Act-frequency signatures of the Big Five. *Personality and Individual Differences*, 116, 201–205.
<https://doi.org/10.1016/j.paid.2017.04.049>
- Chaudhry, S. J., & Loewenstein, G. (2019). Thanking, apologizing, bragging, and blaming:

- Responsibility exchange theory and the currency of communication. *Psychological Review*, 126(3), 313–344. <https://doi.org/10.1037/rev0000139>
- Cohen, T. R., Panter, A. T., & Turan, N. (2012). Guilt proneness and moral character. *Current Directions in Psychological Science*, 21(5), 355–359. <https://doi.org/10.1177/0963721412454874>
- Colman, D. E., Vineyard, J., & Letzring, T. D. (2018). Exploring beyond simple demographic variables: Differences between traditional laboratory samples and crowdsourced online samples on the Big Five personality traits. *Personality and Individual Differences*, 133, 41–46. <https://doi.org/10.1016/j.paid.2017.06.023>
- Costa Jr., P. T., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the revised NEO personality inventory. *Journal of Personality Assessment*, 64(1), 21–50. https://doi.org/10.1207/s15327752jpa6401_2
- Costa, P., & McCrae, R. (1989). The structure of interpersonal traits: Wiggins’s circumplex and the five-factor model. *Journal of Personality and Social Psychology*, 56, 586–595. <https://doi.org/10.1037/0022-3514.56.4.586>
- Craik, K. H. (2000). The lived day of an individual: A person-environment perspective. In W. B. Walsh, K. H. Craik, & R. H. Price (Eds.), *Person-environment psychology: New directions and perspectives (2nd ed.)* (pp. 233–266). Mahwah, NJ: Lawrence Erlbaum Associates.
- Critcher, C. R., Inbar, Y., & Pizarro, D. A. (2013). How quick decisions illuminate moral character. *Social Psychological and Personality Science*, 4(3), 308–315. <https://doi.org/10.1177/1948550612457688>
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302. <https://doi.org/10.1037/h0040957>

- Cuperman, R., & Ickes, W. (2009). Big Five predictors of behavior and perceptions in initial dyadic interactions: Personality similarity helps extraverts and introverts, but hurts “disagreeables.” *Journal of Personality and Social Psychology*, 97(4), 667–684.
<https://doi.org/10.1037/a0015741>
- De Raad, B., Perugini, M., Hrebícková, M., & Szarota, P. (1998). Lingua franca of personality: Taxonomies and structures based on the psycholexical approach. *Journal of Cross-Cultural Psychology*, 29(1), 212–232. <https://doi.org/10.1177/0022022198291011>
- DeSteno, D., Bartlett, M. Y., Baumann, J., Williams, L. A., & Dickens, L. (2010). Gratitude as moral sentiment: Emotion-guided cooperation in economic exchange. *Emotion*, 10(2), 289–293. <https://doi.org/10.1037/a0017883>
- DeYoung, C. G. (2015). Cybernetic Big Five Theory. *Journal of Research in Personality*, 56, 33–58. <https://doi.org/10.1016/j.jrp.2014.07.004>
- 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), 880–896.
<https://doi.org/10.1037/0022-3514.93.5.880>
- 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. <https://doi.org/10.1111/jopy.12020>
- Donnellan, M. B., Lucas, R. E., & Fleeson, W. (1989). Introduction to personality and assessment at age 40: Reflections on the legacy of the person-situation debate and the future of person-situation integration. *Journal of Research in Personality*, 43(2), 117–119.
<https://doi.org/10.1016/j.jrp.2009.02.010>
- Doris, J. M. (2002). *Lack of character: Personality and moral behavior*. New York: Cambridge.

- Ellemers, N., van der Toorn, J., Paunov, Y., & van Leeuwen, T. (2019). The psychology of morality: A review and analysis of empirical studies published from 1940 through 2017. *Personality and Social Psychology Review*, 23(4), 332–366.
<https://doi.org/10.1177/1088868318811759>
- Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology*, 37, 1097–1126.
<https://doi.org/10.1037/0022-3514.37.7.1097>
- Feinberg, M., Willer, R., Stellar, J., & Keltner, D. (2012). The virtues of gossip: Reputational information sharing as prosocial behavior. *Journal of Personality and Social Psychology*, 102(5), 1015–1030. <https://doi.org/10.1037/a0026650>
- Finnigan, K. M., & Vazire, S. (2018). The incremental validity of average state self-reports over global self-reports of personality. *Journal of Personality and Social Psychology*, 115(2), 321–337. <https://doi.org/10.1037/pspp0000136>
- Fleeson, W. (2004). Moving personality beyond the person-situation debate: The challenge and the opportunity of within-person variability. *Current Directions in Psychological Science*, 13(2), 83–87. <https://doi.org/10.1111/j.0963-7214.2004.00280.x>
- Fleeson, William, Furr, R. M., Jayawickreme, E., Meindl, P., & Helzer, E. G. (2014). Character: The prospects for a personality-based perspective on morality. *Social and Personality Psychology Compass*, 8(4), 178–191. <https://doi.org/10.1111/spc3.12094>
- Fleeson, William, & 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), 1097–1114.
<https://doi.org/10.1037/a0016786>

- Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*. <https://doi.org/10.1037/0033-295X.102.4.652>
- Funder, D. C. (2006). Towards a resolution of the personality triad: Persons, situations, and behaviors. *Journal of Research in Personality*, 40(1), 21–34. <https://doi.org/10.1016/J.JRP.2005.08.003>
- Funder, D. C., & Colvin, C. R. (1991). Explorations in behavioral consistency: Properties of persons, situations, and behaviors. *Journal of Personality and Social Psychology*, 60(5), 773–794. <https://doi.org/10.1037//0022-3514.60.5.773>
- Funder, D. C., & Sneed, C. D. (1993). Behavioral manifestations of personality: An ecological approach to judgmental accuracy. *Journal of Personality and Social Psychology*, 64(3), 479–490. <https://doi.org/10.1037/0022-3514.64.3.479>
- Furr, R. M. (2009). Personality psychology as a truly behavioral science. *European Journal of Personality*, 23, 369–401. <https://doi.org/10.1002/per.724>
- Furr, R. M., & Funder, D. C. (2004). Situational similarity and behavioral consistency: Subjective, objective, variable-centered, and person-centered approaches. *Journal of Research in Personality*, 38(5), 421–447. <https://doi.org/10.1016/j.jrp.2003.10.001>
- Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, 106(1), 148–168. <https://doi.org/10.1037/a0034726>
- Gosling, S. D., John, O. P., Craik, K. H., & Robins, R. W. (1998). Do people know how they behave? Self-reported act frequencies compared with on-line codings by observers. *Journal of Personality and Social Psychology*, 74(5), 1337–1349. <https://doi.org/10.1037/0022-3514.74.5.1337>

- Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, 96(5), 1029–1046.
<https://doi.org/10.1037/a0015141>
- Graham, J., Meindl, P., Beall, E., Johnson, K. M., & Zhang, L. (2016). Cultural differences in moral judgment and behavior, across and within societies. *Current Opinion in Psychology*, 8, 125–130. <https://doi.org/10.1016/j.copsyc.2015.09.007>
- Grapsas, S., Brummelman, E., Back, M. D., & Denissen, J. J. A. (2020). The “why” and “how” of narcissism: A process model of narcissistic status pursuit. *Perspectives on Psychological Science*, 15(1), 150–172. <https://doi.org/10.1177/1745691619873350>
- Greengross, G., & Miller, G. (2011). Humor ability reveals intelligence, predicts mating success, and is higher in males. *Intelligence*, 39(4), 188–192.
<https://doi.org/10.1016/j.intell.2011.03.006>
- Gurtman, M. B. (1992). Construct validity of interpersonal personality measures: The interpersonal circumplex as a nomological net. *Journal of Personality and Social Psychology*, 63(1), 105–118. <https://doi.org/10.1037/0022-3514.63.1.105>
- Gurtman, M. B. (2009). Exploring personality with the interpersonal circumplex. *Social and Personality Psychology Compass*, 3(4), 601–619. <https://doi.org/10.1111/j.1751-9004.2009.00172.x>
- Gurtman, M. B., & Lee, D. L. (2009). Sex differences in interpersonal problems: A circumplex analysis. *Psychological Assessment*, 21(4), 515–527. <https://doi.org/10.1037/a0017085>
- 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.

<https://doi.org/10.1177/0146167216652859>

- Haidt, J., & Kesebir, S. (2010). Morality. In S. Fiske, D. T. Gilbert, & Lindzey G. (Eds.), *Hand-book of social psychology* (5th ed., pp. 797–832). Hoboken, NJ: Wiley.
- Harari, G. M., Muller, Sandrine, R., Stachl, C., Wang, R., Wang, W., Bühner, M., ... Gosling, S. D. (2019). Sensing sociability: Individual differences in young adults' conversation, calling, texting and app use behaviors in daily life. *Journal of Personality and Social Psychology*.
<https://doi.org/10.1037/pspp0000245>
- Harris, K., & Vazire, S. (2016). On friendship development and the Big Five personality traits. *Social and Personality Psychology Compass*, 10(11), 647–667.
<https://doi.org/10.1111/spc3.12287>
- Hirsh, J. B., DeYoung, C. G., & Peterson, J. B. (2009). Metatraits of the Big Five differentially predict engagement and restraint of behavior. *Journal of Personality*, 77(4), 1085–1102.
<https://doi.org/10.1111/j.1467-6494.2009.00575.x>
- Holtzman, N. S., Tackman, A. M., Carey, A. L., Brucks, M. S., Küfner, A. C. P., Deters, F. G., ... Mehl, M. R. (2019). Linguistic markers of grandiose narcissism: A LIWC analysis of 15 samples. *Journal of Language and Social Psychology*, 38(5–6), 773–786.
<https://doi.org/10.1177/0261927X19871084>
- Holtzman, N. S., Vazire, S., & Mehl, M. R. (2010). Sounds like a narcissist: Behavioral manifestations of narcissism in everyday life. *Journal of Research in Personality*, 44(4), 478–484. <https://doi.org/10.1016/j.jrp.2010.06.001>
- Hu, L., & 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. <https://doi.org/10.1080/10705519909540118>

- Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology*, 65(1), 373–398. <https://doi.org/10.1146/annurev-psych-010213-115057>
- Jaccard, J. J. (1974). Predicting social behavior from personality traits. *Journal of Research in Personality*, 7, 358–367.
- James, W. (1890). *The principles of psychology*. New York: H. Holt and Company.
- John, O P, & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–138). New York: The Guilford Press.
- John, Oliver 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* (3rd ed., pp. 114–158). New York: Guilford Press.
- Jokela, M., Alvergne, A., Pollet, T. V., & Lummaa, V. (2011). Reproductive behavior and personality traits of the Five Factor Model. *European Journal of Personality*, 25(6), 487–500. <https://doi.org/10.1002/per.822>
- Kam, C., & Meyer, J. P. (2012). Do optimism and pessimism have different relationships with personality dimensions? A re-examination. *Personality and Individual Differences*, 52(2), 123–127. <https://doi.org/10.1016/J.PAID.2011.09.011>
- Kammrath, L. K., & Scholer, A. A. (2011). The pollyanna myth: How highly agreeable people judge positive and negative relational acts. *Personality and Social Psychology Bulletin*, 37(9), 1172–1184. <https://doi.org/10.1177/0146167211407641>
- Kaplan, D. M., Raison, C. L., Milek, A., Tackman, A. M., Pace, T. W. W., & Mehl, M. R. (2018). Dispositional mindfulness in daily life: A naturalistic observation study. *PLoS ONE*,

- 13(11), 1–15. <https://doi.org/10.1371/journal.pone.0206029>
- Kashdan, T. B., Mishra, A., Breen, W. E., & Froh, J. J. (2009). Gender differences in gratitude: Examining appraisals, narratives, the willingness to express emotions, and changes in psychological needs. *Journal of Personality*, 77(3), 691–730.
<https://doi.org/10.1111/j.1467-6494.2009.00562.x>
- Kenrick, D. T., & Funder, D. C. (1988). Profiting from controversy: Lessons from the person–situation debate. *American Psychologist*, 43(1), 23–34.
- Krueger, R. F., Hicks, B. M., & McGue, M. (2001). Altruism and antisocial behavior: Independent tendencies, unique personality correlates, distinct etiologies. *Psychological Science*, 12(5), 397–402.
- Kulkarni, V., Kern, M. L., Stillwell, D., Kosinski, M., Matz, S., Ungar, L., ... Schwartz, H. A. (2018). Latent human traits in the language of social media: An open-vocabulary approach. *PLoS ONE*, 13(11), e0201703. <https://doi.org/10.1371/journal.pone.0201703>
- Lapsley, D. K., & Lasky, B. (2001). Prototypic moral character. *Identity*, 1(4), 345–363.
https://doi.org/10.1207/S1532706XID0104_03
- Larson, R., & Csikszentmihalyi, M. (1983). The experience sampling method. In *Naturalistic approaches to studying social interaction* (pp. 41–56).
- Leary, T. (1957). *The interpersonal diagnosis of personality*. New York: Roland.
- LeBel, E. P. (2011). *Metric calibration of psychological instruments: Utility and feasibility*. Saarbrücken, Germany: LAP Lambert Academic Publishing.
- Leckelt, M., Kkufner, A. C. P., Nestler, S., & Back, M. D. (2015). Behavioral processes underlying the decline of narcissists' popularity. *Journal of Personality and Social Psychology*, 109(5), 856–871.

- Lee, K., & Ashton, M. C. (2005). Psychopathy, machiavellianism, and narcissism in the five-factor model and the HEXACO model of personality structure. *Personality and Individual Differences*, 38(7), 1571–1582. <https://doi.org/10.1016/J.PAID.2004.09.016>
- Lee, K., & Ashton, M. C. (2018). Psychometric properties of the HEXACO-100. *Assessment*, 25(5), 543–556. <https://doi.org/10.1177/1073191116659134>
- Leikas, S., Lönnqvist, J. E., & Verkasalo, M. (2012). Persons, situations, and behaviors: Consistency and variability of different behaviors in four interpersonal situations. *Journal of Personality and Social Psychology*, 103(6), 1007–1022. <https://doi.org/10.1037/a0030385>
- Leising, D., & Bleidorn, W. (2011). Which are the basic meaning dimensions of observable interpersonal behavior ? *Personality and Individual Differences*, 51(8), 986–990. <https://doi.org/10.1016/j.paid.2011.08.003>
- Leising, D., Erbs, J., & Fritz, U. (2010). The letter of recommendation effect in informant ratings of personality. *Journal of Personality and Social Psychology*, 98(4), 668–682. <https://doi.org/10.1037/a0018771>
- Lott, A. J., Lott, B. E., Reed, T., & Crow, T. (1970). Personality-trait descriptions of differentially liked persons. *Journal of Personality and Social Psychology*, 16(2), 284–290. <https://doi.org/10.1037/h0029851>
- Lucas, R. E., & Donnellan, M. B. (2012). Estimating the reliability of single-item life satisfaction measures: Results from four national panel studies. *Social Indicators Research*, 105(3), 323–331. <https://doi.org/10.1007/s11205-011-9783-z>
- Malle, B. F. (2006). Intentionality, morality, and their relationship in human judgment. *Journal of Cognition and Culture*, 6(1–2), 87–112. <https://doi.org/10.1163/156853706776931358>
- Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A theory of blame. *Psychological Inquiry*,

- 25(2), 147–186. <https://doi.org/10.1080/1047840X.2014.877340>
- Manson, J. H., & Robbins, M. L. (2017). New evaluation of the Electronically Activated Recorder (EAR): Obtrusiveness, compliance, and participant self-selection effects. *Frontiers in Psychology*, 8(April), 1–9. <https://doi.org/10.3389/fpsyg.2017.00658>
- Markey, P., Anderson, J. M., & Markey, C. (2013). Using behavioral mapping to examine the validity of the IPIP-IPC. *Assessment*, 20(2), 165–174. <https://doi.org/10.1177/1073191112436669>
- Markey, P. M., & Markey, C. N. (2009). A brief assessment of the interpersonal circumplex: The IPIP-IPC. *Assessment*, 16(4), 352–361. <https://doi.org/10.1177/1073191109340382>
- Markey, P. M., Suzuki, T., & Marino, D. P. (2014). The interpersonal meaning of humor styles. *Humor*, 27(1), 47–64. <https://doi.org/10.1515/humor-2013-0052>
- McAbee, S. T., & Connelly, B. S. (2016). A multi-rater framework for studying personality: The trait-reputation-identity model. *Psychological Review*, 123(5), 569–591. <https://doi.org/10.1037/rev0000035.supp>
- McAdams, D. P. (2009). The moral personality. In D. Narvaez & D. K. Lapsley (Eds.), *Personality, identity, and character: Explorations in moral personality* (pp. 11–29). Cambridge University Press.
- McAdams, D. P., & Pals, J. L. (2006). A new Big Five: Fundamental principles for an integrative science of personality. *American Psychologist*, 61(3), 204–217. <https://doi.org/10.1037/0003-066X.61.3.204>
- McCrae, R. R. (1996). Social consequences of experiential openness. *Psychological Bulletin*, 120(3), 323–337. <https://doi.org/10.1037/0033-2909.120.3.323>
- McCullough, M. E., Kilpatrick, S. D., Emmons, R. A., & Larson, D. B. (2001). Is gratitude a

- moral affect? *Psychological Bulletin*, 127(2), 249–266. <https://doi.org/10.1037/0033-2909.127.2.249>
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, N. J.: L. Erlbaum Associates.
- Mehl, M. R. (2007). Eavesdropping on health: A naturalistic observation approach for social health research. *Social and Personality Psychology Compass*, 1(1), 359–380. <https://doi.org/10.1111/j.1751-9004.2007.00034.x>
- Mehl, M. R. (2017). The Electronically Activated Recorder (EAR). *Current Directions in Psychological Science*, 26(2), 184–190. <https://doi.org/10.1177/0963721416680611>
- Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology*, 90(5), 862–877. <https://doi.org/10.1037/0022-3514.90.5.862>
- Mehl, M. R., & Holleran, S. E. (2007). An empirical analysis of the obtrusiveness of and participants' compliance with the Electronically Activated Recorder (EAR). *European Journal of Psychological Assessment*, 23(4), 248–257. <https://doi.org/10.1027/1015-5759.23.4.248>
- Mehl, M. R., & Pennebaker, J. W. (2003). The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations. *Journal of Personality and Social Psychology*, 84(4), 857–870. <https://doi.org/10.1037/0022-3514.84.4.857>
- Mehl, M. R., Pennebaker, J. W., Crow, D. M., Dabbs, J., & Price, J. H. (2001). The Electronically Activated Recorder (EAR): A device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments, & Computers*, 33(4), 517–523. <https://doi.org/10.3758/BF03195410>

- Mehl, M. R., Robbins, M. L., & Deters, F. G. (2012). Naturalistic observation of health-relevant social processes: The electronically activated recorder methodology in psychosomatics. *Psychosomatic Medicine*. <https://doi.org/10.1097/PSY.0b013e3182545470>
- Mehl, M. R., Vazire, S., Ramírez-Esparza, N., Slatcher, R. B., & Pennebaker, J. W. (2007). Are women really more talkative than men? *Science*, 317(5834), 82. <https://doi.org/10.1126/science.1139940>
- Meindl, P., & Graham, J. (2014). Know thy participant: The trouble with nomothetic assumptions in moral psychology. In H. Sarkissian & J. C. Wright (Eds.), *Advances in Experimental Moral Psychology* (pp. 233–252). London: Bloomsbury. <https://doi.org/10.5040/9781472594150.ch-012>
- Mendiburo-Seguel, A., Páez, D., & Martínez-Sánchez, F. (2015). Humor styles and personality: A meta-analysis of the relation between humor styles and the Big Five personality traits. *Scandinavian Journal of Psychology*, 56(3), 335–340. <https://doi.org/10.1111/sjop.12209>
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525–543. <https://doi.org/10.1007/bf02294825>
- Milek, A., Butler, E. A., Tackman, A. M., Kaplan, D. M., Raison, C. L., Sbarra, D. A., ... Mehl, M. R. (2018). “Eavesdropping on happiness” revisited: A pooled, multisample replication of the association between life satisfaction and observed daily conversation quantity and quality. *Psychological Science*, 29(9), 1452–1462. <https://doi.org/10.1177/0956797618774252>
- Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, 67(4), 371–378. <https://doi.org/10.1037/h0040525>
- Miller, L. C., Cooke, L. L., Tsang, J., & Morgan, F. (1992). Should I brag? Nature and impact of

- positive and boastful disclosures for women and men. *Human Communication Research*, 18, 364–399. <https://doi.org/10.1111/j.1468-2958.1992.tb00557.x>
- Mischel, W. (1968). *Personality and assessment*. New York: Wiley.
- Moskowitz, D. S. (1994). Cross-situational generality and the interpersonal circumplex. *Journal of Personality and Social Psychology*, 66(5), 921–933. <https://doi.org/10.1037//0022-3514.66.5.921>
- Moskowitz, D. S., Suh, E. J., & Desaulniers, J. (1994). Situational influences on gender differences in agency and communion. *Journal of Personality and Social Psychology*, 66(4), 753–761. <https://doi.org/10.1037/0022-3514.66.4.753>
- Murray, S. L., Holmes, J. G., Dolderman, D., & Griffin, D. W. (2000). What the motivated mind sees: Comparing friends' perspectives to married partners' views of each other. *Journal of Experimental Social Psychology*, 36(6), 600–620. <https://doi.org/10.1006/JESP.1999.1417>
- Nelson, B. W., & Allen, N. B. (2018). Extending the passive-sensing toolbox: Using smart-home technology in psychological science. *Perspectives on Psychological Science*, 13(6), 718–733. <https://doi.org/10.1177/1745691618776008>
- Newman, G. E., & Cain, D. M. (2014). Tainted altruism: When doing some good is evaluated as worse than doing no good at all. *Psychological Science*, 25(3), 648–655. <https://doi.org/10.1177/0956797613504785>
- Ng, V., & Tay, L. (2020). Lost in translation: The construct representation of character virtues. *Perspectives on Psychological Science*, 15(2), 309–326. <https://doi.org/10.1177/1745691619886014>
- Norman, W. T. (1967). *2,800 personality trait descriptors: Normative operating characteristics for a university population*. Department of Psychology, University of Michigan.

- Ozer, D. J., & Benet-Martínez, V. (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology*, 57, 401–421.
<https://doi.org/10.1146/annurev.psych.57.102904.190127>
- Palmer, C. A., Ramsey, M. A., Morey, J. N., & Gentzler, A. L. (2016). How do people share their positive events?: Individual differences in capitalizing, bragging, and mass-sharing. *Journal of Individual Differences*, 37(4), 250–259. <https://doi.org/10.1027/1614-0001/a000212>
- Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., ... Seligman, M. E. P. (2015). Automatic personality assessment through social media language. *Journal of Personality and Social Psychology*, 108(6), 1–25.
<https://doi.org/10.1037/pspp0000020>
- Paulhus, D. L. (1998). Interpersonal and intrapsychic adaptiveness of trait self-enhancement: A mixed blessing? *Journal of Personality and Social Psychology*, 74(5), 1197–1208.
<https://doi.org/10.1037/0022-3514.74.5.1197>
- Paunonen, S. V., & Ashton, M. C. (2001). Big Five factors and facets and the prediction of behavior. *Journal of Personality and Social Psychology*, 81(3), 524–539.
<https://doi.org/10.1037/0022-3514.81.3.524>
- Peterson, C., & Seligman, M. E. P. (2004). *Character strengths and virtues: A handbook and classification*. New York: Oxford University Press.
- Pincus, A. L. (2002). Constellations of dependency within the five-factor model of personality. In P. T. Costa & T. A. Widiger (Eds.), *Personality disorders and the five-factor Model of personality* (2nd ed., pp. 203–214). Washington, D.C.: American Psychological Association.

- Pinker, S. (1997). *How the mind works*. New York: Norton.
- R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Raskin, R., & Terry, H. (1988). A principal-components analysis of the Narcissistic Personality Inventory and further evidence of its construct validity. *Journal of Personality and Social Psychology*, 54, 890–902. <https://doi.org/10.1037//0022-3514.54.5.890>
- Ritter, D., & Eslea, M. (2005). Hot sauce, toy guns, and graffiti: A critical account of current laboratory aggression paradigms. *Aggressive Behavior*, 31(5), 407–419. <https://doi.org/10.1002/ab.20066>
- Robbins, M. L., & Karan, A. (2020). Who gossips and how in everyday life? *Social Psychological and Personality Science*, 11(2), 185–195. <https://doi.org/10.1177/1948550619837000>
- Robbins, M. L., López, A. M., Weihs, K. L., & Mehl, M. R. (2014). Cancer conversations in context: Naturalistic observation of couples coping with breast cancer. *Journal of Family Psychology*, 28(3), 380–390. <https://doi.org/10.1037/a0036458>
- Roberts, B. W., Wood, D., & Caspi, A. (2008). The development of personality traits in adulthood. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 375–398). New York: Guilford Press.
- Roberts, Brent W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*, 126(1), 3–25. <https://doi.org/10.1037/0033-2909.126.1.3>
- Roberts, Brent W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality. *Perspectives on Psychological Science*, 2(4), 313–345.

<https://doi.org/10.1111/j.1745-6916.2007.00047.x>

Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem:

Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale.

Personality and Social Psychology Bulletin, 27(2), 151–161.

<https://doi.org/10.1177/0146167201272002>

Robins, R. W., & John, O. P. (1997). The quest for self-insight: Theory and research on accuracy and bias in self-perception. In R. T. Hogan, J. A. Johnson, & S. R. Briggs (Eds.), *Handbook of personality psychology* (pp. 649–679). New York, NY: Academic Press.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>

Rozin, P. (2001). Social psychology and science: Some lessons from Solomon Asch. *Personality and Social Psychology Review*, 5(1), 2–14. https://doi.org/10.1207/S15327957PSPR0501_1

Rudman, L. A. (1998). Self-promotion as a risk factor for women: The costs and benefits of counterstereotypical impression management. *Journal of Personality and Social Psychology*, 74, 629–645. <https://doi.org/10.1037//0022-3514.74.3.629>

Ruiz, J. M., Smith, T. W., & Rhodewalt, F. (2001). Distinguishing narcissism and hostility: Similarities and differences in interpersonal circumplex and five-factor correlates. *Journal of Personality Assessment*, 76(3), 537–555. https://doi.org/10.1207/S15327752JPA7603_12

Saucier, G., & Goldberg, L. R. (2001). Lexical studies of indigenous personality factors: Premises, products, and prospects. *Journal of Personality*, 69(6), 847–879.

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology*, 67, 1063–1078.

<https://doi.org/10.1037//0022-3514.67.6.1063>

Schmidt, J. A., Wagner, C. C., & Kiesler, D. J. (1999). Covert reactions to Big Five personality traits: The Impact Message Inventory and the NEO-PI-R. *European Journal of Psychological Assessment*, 15, 221–232.

Schumann, K., & Ross, M. (2010). Why women apologize more than men: Gender differences in thresholds for perceiving offensive behavior. *Psychological Science*, 21(11), 1649–1655.

<https://doi.org/10.1177/0956797610384150>

Scollon, C. N., Chu, K.-P., & Diener, E. (2003). Experience sampling: Promises and pitfalls, strengths and weaknesses. *Journal of Happiness Studies*, 4(1925), 5–34.

<https://doi.org/10.1023/A:1023605205115>

Sechrest, L., McKnight, P., & McKnight, K. (1996). Calibration of measures for psychotherapy outcome studies. *The American Psychologist*, 51(10), 1065–1071.

<https://doi.org/10.1037/0003-066X.51.10.1065>

Sharpe, J. P., Martin, N. R., & Roth, K. A. (2011). Optimism and the Big Five factors of personality: Beyond neuroticism and extraversion. *Personality and Individual Differences*, 51(8), 946–951. <https://doi.org/10.1016/J.PAID.2011.07.033>

Sherman, R. A., & Funder, D. C. (2009). Evaluating correlations in studies of personality and behavior: Beyond the number of significant findings to be expected by chance. *Journal of Research in Personality*, 43(6), 1053–1063. <https://doi.org/10.1016/j.jrp.2009.05.010>

Skowronski, J. J., & Carlston, D. E. (1987). Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases. *Journal of Personality and Social Psychology*, 52(4), 689–699. <https://doi.org/10.1037/0022-3514.52.4.689>

Smith, J. L., & Huntoon, M. (2014). Women's bragging rights: Overcoming modesty norms to

- facilitate women's self-promotion. *Psychology of Women Quarterly*, 38(4), 447–459.
<https://doi.org/10.1177/0361684313515840>
- Smith, T. W., Ruiz, J. M., Cundiff, J. M., Baron, K. G., & Nealey-Moore, J. B. (2013). Optimism and pessimism in social context: An interpersonal perspective on resilience and risk. *Journal of Research in Personality*, 47(5), 553–562.
<https://doi.org/10.1016/j.jrp.2013.04.006>
- Solomon, B. C., & Vazire, S. (2014). You are so beautiful . . . to me: Seeing beyond biases and achieving accuracy in romantic relationships. *Journal of Personality and Social Psychology*, 107(3), 516–528. <https://doi.org/10.1037/a0036899>
- Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, 113(1), 117–143.
<https://doi.org/10.1037/pspp0000096>
- Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 3(3), 271–295. <https://doi.org/10.1111/j.2044-8295.1910.tb00206.x>
- Starmans, C., & Bloom, P. (2016). When the spirit is willing, but the flesh is weak: Developmental differences in judgments about inner moral conflict. *Psychological Science*, 27(11), 1498–1506. <https://doi.org/10.1177/0956797616665813>
- Sun, J., Harris, K., & Vazire, S. (2019). Is well-being associated with the quantity and quality of social interactions? *Journal of Personality and Social Psychology*. Advance online publication. <https://doi.org/10.1037/pspp0000272>
- Sun, J., Rhemtulla, M., & Vazire, S. (under review). *Eavesdropping on missing data: What are people doing when they miss experience sampling reports?* Manuscript under review.

<https://psyarxiv.com/5tcwd>

- Sun, J., Schwartz, H. A., Son, Y., Kern, M. L., & Vazire, S. (2020). The language of well-being: Tracking fluctuations in emotion experience through everyday speech. *Journal of Personality and Social Psychology*, 118(2), 364–387. <https://doi.org/10.1037/pspp0000244>
- Sun, J., Stevenson, K., Kabbani, R., Richardson, B., & Smillie, L. D. (2017). The pleasure of making a difference: Perceived social contribution explains the relation between extraverted behavior and positive affect. *Emotion*, 17(5), 794–810. <https://doi.org/10.1037/emo0000273>
- Sun, J., & Vazire, S. (2019). Do people know what they're like in the moment? *Psychological Science*, 30(3), 405–414. <https://doi.org/10.1177/0956797618818476>
- Sutu, A., Serrano, S., Schultz, L. H., Jackson, J. J., & Damian, R. I. (2019). Creating through deviancy or adjustment? The link between personality profile normativeness and creativity. *European Journal of Personality*, 33(5), 565–588. <https://doi.org/10.1002/per.2215>
- Tackman, A. M., Sbarra, D. A., Carey, A. L., Donnellan, M. B., Horn, A. B., Holtzman, N. S., ... Mehl, M. R. (2019). Depression, negative emotionality, and self-referential language: A multi-lab, multi-measure, and multi-language-task research synthesis. *Journal of Personality and Social Psychology*, 116(5), 817–834. <https://doi.org/10.1037/pspp0000187>
- Tellegen, A. (1981). Practicing the two disciplines for relaxation and enlightenment: Comment on “Role of the feedback signal in electromyograph biofeedback: The relevance of attention” by Qualls and Sheehan. *Journal of Experimental Psychology: General*, 110, 217–226. <https://doi.org/10.1037/0096-3445.110.2.217>
- Tracy, J. L., Cheng, J. T., Robins, R. W., & Trzesniewski, K. H. (2009). Authentic and hubristic pride: The affective core of self-esteem and narcissism. *Self and Identity*, 8(2–3), 196–213. <https://doi.org/10.1080/15298860802505053>

- Tracy, J. L., Robins, R. W., & Sherman, J. W. (2009). The practice of psychological science: Searching for Cronbach's two streams in social-personality psychology. *Journal of Personality and Social Psychology*, 96(6), 1206–1225. <https://doi.org/10.1037/a0015173>
- Trafimow, D., & Trafimow, S. (1999). Mapping perfect and imperfect duties onto hierarchically and partially restrictive trait dimensions. *Personality and Social Psychology Bulletin*, 25(6), 687–697. <https://doi.org/10.1177/0146167299025006004>
- Trobst, K. K. (2000). An interpersonal conceptualization and quantification of social support transactions. *Personality and Social Psychology Bulletin*, 26(8), 971–986. <https://doi.org/10.1177/01461672002610007>
- Tsang, J.-A., Rowatt, W. C., & Shariff, A. F. (2015). Religion and prosociality. In D. A. Schroeder & W. G. Graziano (Eds.), *The Oxford Handbook of Prosocial Behavior* (pp. 1–30). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195399813.013.013>
- van Scheppingen, M. A., Jackson, J. J., Specht, J., Hutteman, R., Denissen, J. J. A., & Bleidorn, W. (2016). Personality trait development during the transition to parenthood. *Social Psychological and Personality Science*, 7(5), 452–462. <https://doi.org/10.1177/1948550616630032>
- Vazire, S., & Carlson, E. N. (2011). Others sometimes know us better than we know ourselves. *Current Directions in Psychological Science*, 20(2), 104–108. <https://doi.org/10.1177/0963721411402478>
- Vazire, Simine. (2010). Who knows what about a person? The self-other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, 98(2), 281–300. <https://doi.org/10.1037/a0017908>
- 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.
<https://doi.org/10.3389/fpsyg.2011.00178>
- Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology*, 37(3), 395–412.
<https://doi.org/10.1037/0022-3514.37.3.395>
- Wiggins, J. S. (1995). *Interpersonal adjective scales professional manual*. Odessa, FL: Psychological Assessment Resources.
- Wilson, R. E., Harris, K., & Vazire, S. (2015). Personality and friendship satisfaction in daily life: Do everyday social interactions account for individual differences in friendship satisfaction? *European Journal of Personality*, 29(2), 173–186.
<https://doi.org/10.1002/per.1996>
- Wilt, J., & Revelle, W. (2015). Affect, behaviour, cognition and desire in the Big Five: An analysis of item content and structure. *European Journal of Personality*, 29(4), 478–497.
<https://doi.org/10.1002/per.2002>
- Wojciszke, B., Abele, A. E., & Baryla, W. (2009). Two dimensions of interpersonal attitudes: Liking depends on communion, respect depends on agency. *European Journal of Social Psychology*, 39(6), 973–990. <https://doi.org/10.1002/ejsp.595>
- Wood, A. M., Joseph, S., & Maltby, J. (2009). Gratitude predicts psychological well-being above the Big Five facets. *Personality and Individual Differences*, 46(4), 443–447.
<https://doi.org/10.1016/j.paid.2008.11.012>
- Wortman, J., & Wood, D. (2011). The personality traits of liked people. *Journal of Research in Personality*, 45(6), 519–528. <https://doi.org/10.1016/j.jrp.2011.06.006>
- Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are

- more accurate than those made by humans. *Proceedings of the National Academy of Sciences of the United States of America*, 112(4), 1036–1040.
<https://doi.org/10.1073/pnas.1418680112>
- Zaki, J., & Craig Williams, W. (2013). Interpersonal emotion regulation. *Emotion*, 13(5), 803–810. <https://doi.org/10.1037/a0033839>
- Zhao, K., Ferguson, E., & Smillie, L. D. (2017). Individual differences in good manners rather than compassion predict fair allocations of wealth in the dictator game. *Journal of Personality*, 85(2), 244–256. <https://doi.org/10.1111/jopy.12237>
- Zhao, K., & Smillie, L. D. (2015). The role of interpersonal traits in social decision making: Exploring sources of behavioral heterogeneity in economic games. *Personality and Social Psychology Review*, 19(3), 277–302. <https://doi.org/10.1177/1088868314553709>
- 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. <https://doi.org/10.1177/0146167202289013>

Supplemental Tables

Table S1

Factor Loadings for Exploratory Two- to Seven-Factor Solutions of Interpersonal Behaviors Observed in Year 1

Behavior	Two-Factor Solution		Three-Factor Solution			Four-Factor Solution				Five-Factor Solution				
	1	2	1	2	3	1	2	3	4	1	2	3	4	5
Gratitude	-.08	-.02	-.10	-.04	.19	-.10	-.05	.21	.04	-.12	-.06	.13	.22	.07
Praise/compliment	.13	.34	.11	.32	.29	.08	.32	.30	.13	.03	.32	.16	.58	.34
Sympathy	.10	.08	.04	.02	.68	.10	.02	.69	.04	.08	.00	.99	-.01	-.06
Help/favors/support	.03	.04	-.01	.00	.41	.01	.00	.42	-.03	-.02	.00	.37	.06	.07
Apologize	.07	.05	.05	.03	.15	.14	.04	.12	-.18	.02	.04	.06	.00	.32
Hope/optimism	.11	.16	.07	.12	.58	.12	.13	.57	-.06	.04	.12	.45	.12	.21
Funny	.04	.68	.04	.66	.14	.05	.66	.13	.00	.06	.65	.12	.05	.07
Negative gossip	.61	.16	.62	.15	.02	.50	.14	.02	.83	.74	.08	.03	.43	-.29
Neutral/positive gossip	.06	.05	.06	.04	.05	-.07	.02	.11	.42	.07	-.01	.02	.46	-.19
Sarcasm	.07	.61	.09	.62	-.02	.08	.64	-.04	-.04	.09	.63	-.04	-.02	.07
Brag	.03	.43	.04	.43	.02	.01	.43	.02	.02	.03	.43	-.02	.12	.07
Condescension	.31	.30	.34	.32	-.16	.25	.33	-.17	.15	.35	.31	-.08	-.05	-.12
Complain	.65	.08	.63	.06	.08	.68	.10	-.01	.06	.64	.06	.06	-.05	.22
Criticism	.68	.27	.72	.27	-.04	.58	.30	-.07	.29	.68	.27	-.07	.09	-.03
Blame	.63	-.02	.62	-.06	.18	.65	-.01	.10	-.04	.57	-.03	.10	-.10	.26
Pessimism	.54	.01	.52	-.01	.10	.65	.03	.01	-.12	.53	-.01	.03	-.11	.42
Tease	.06	.63	.08	.63	.02	.04	.64	.02	-.02	.09	.66	.07	-.10	-.09
Insecurity/self-doubt	.12	.03	.09	.00	.17	.21	.02	.12	-.15	.08	.00	.06	.00	.35

(table continues)

Table S1 (*continued*)

Behavior	Six-Factor Solution						Seven-Factor Solution						
	1	2	3	4	5	6	1	2	3	4	5	6	7
Gratitude	-.12	-.07	.14	.05	.20	.06	-.07	-.06	.13	-.08	.02	.04	.24
Praise/compliment	.04	.29	.18	.04	.60	.32	.30	.14	.16	.06	-.06	.21	.65
Sympathy	.11	.04	.98	-.11	-.04	-.07	.04	.09	.98	-.09	-.04	.12	-.10
Help/favors/support	-.03	-.02	.38	.05	.06	.07	-.03	-.02	.40	.07	.04	-.06	.15
Apologize	.03	.04	.06	-.01	.00	.31	.05	.21	.05	.00	-.06	-.15	.14
Hope/optimism	.06	.12	.45	-.03	.13	.20	.11	.14	.45	.04	-.05	-.02	.22
Funny	.07	.65	.10	-.01	.08	.08	.68	.09	.08	-.01	-.02	.06	.06
Negative gossip	.72	.07	.02	.06	.43	-.32	.08	.24	.02	.36	.18	.75	-.03
Neutral/positive gossip	.07	.02	.02	-.08	.43	-.22	.02	-.11	.00	-.01	-.05	.47	.17
Sarcasm	.08	.58	-.04	.17	.03	.09	.59	.08	-.05	.06	.15	-.02	.07
Brag	-.01	.38	.00	.26	.16	.08	.39	.02	-.01	-.01	.22	.02	.21
Condescension	.27	.22	.00	.91	-.07	-.19	.23	.05	-.02	.15	.96	.05	-.04
Complain	.65	.07	.05	.06	-.04	.20	.09	.63	.03	.20	.12	.21	-.13
Criticism	.67	.25	-.08	.14	.11	-.05	.21	.21	.02	.92	.17	.19	-.02
Blame	.60	-.01	.08	-.05	-.11	.25	-.02	.50	.12	.38	.00	.04	-.09
Pessimism	.53	-.02	.03	.13	-.09	.40	.01	.70	.01	.13	.14	-.01	-.01
Tease	.10	.74	.03	-.05	-.10	-.10	.70	-.04	.05	.13	-.02	.02	-.14
Insecurity/self-doubt	.10	.01	.06	-.06	.01	.35	.04	.35	.02	-.10	-.10	-.07	.10

Note. Factor loadings > .40 are shown in **boldface**.

Table S2

Factor Loadings for Exploratory Two- to Seven-Factor Solutions of Interpersonal Behaviors Observed in Year 2

Behavior	Two-Factor Solution		Three-Factor Solution			Four-Factor Solution				Five-Factor Solution				
	1	2	1	2	3	1	2	3	4	1	2	3	4	5
Gratitude	-.10	.30	-.09	.28	.06	-.09	.30	.03	.29	-.05	.24	.21	-.23	-.17
Praise/compliment	.06	.53	.04	.47	.27	.05	.49	.26	.08	.05	.35	.50	-.10	.03
Sympathy	.06	.78	.09	.80	.07	.05	.78	.09	-.02	.06	.79	.18	.06	.06
Help/favors/support	-.14	.53	-.10	.54	-.02	-.12	.55	-.03	.16	-.09	.53	.11	-.14	-.08
Apologize	.04	.04	.03	.01	.09	.04	.02	.07	.27	.05	.03	.05	.00	-.24
Hope/optimism	-.06	.72	-.02	.76	-.02	-.06	.74	.00	.00	-.04	.79	.07	.00	.02
Funny	.05	.44	.06	.42	.09	.05	.44	.10	-.11	.05	.26	.46	-.24	.31
Negative gossip	.21	.25	.09	.12	.85	.14	.11	.81	.19	.08	.05	.66	.41	-.31
Neutral/positive gossip	.06	.30	-.01	.22	.49	.02	.22	.48	.05	-.02	.08	.60	.06	-.05
Sarcasm	.57	.24	.57	.22	.04	.55	.25	.06	-.33	.51	.14	.20	.13	.44
Brag	.66	.09	.65	.06	.14	.66	.08	.10	.06	.65	.11	.03	.22	-.09
Condescension	.58	.02	.54	-.04	.26	.55	-.03	.24	-.02	.51	.01	.06	.39	-.07
Complain	.64	-.19	.61	-.22	.11	.61	-.21	.11	-.13	.56	-.17	-.07	.36	.06
Criticism	.84	.06	.81	-.02	.33	.82	.02	.28	.04	.79	-.03	.27	.25	-.03
Blame	.46	-.05	.40	-.12	.31	.40	-.13	.34	-.18	.33	-.09	.07	.56	.05
Pessimism	.75	-.05	.80	-.04	-.12	.83	.01	-.23	.20	.88	-.01	-.09	-.16	-.06
Tease	.71	.13	.72	.11	.00	.72	.15	.02	-.47	.66	.08	.08	.22	.49
Insecurity/self-doubt	.75	-.23	.77	-.23	-.09	.79	-.19	-.17	.10	.80	-.20	-.11	-.02	-.01

(table continues)

Table S2 (*continued*)

Behavior	Six-Factor Solution						Seven-Factor Solution						
	1	2	3	4	5	6	1	2	3	4	5	6	7
Gratitude	-.08	.18	.05	-.11	.16	.56	-.06	.16	.01	-.12	.17	.64	-.13
Praise/compliment	-.01	.30	.30	.00	.40	.31	.00	.32	.30	.01	.39	.30	-.05
Sympathy	.05	.79	.13	.04	.19	.01	.05	.87	.16	.06	.16	-.03	.07
Help/favors/support	-.11	.50	.00	-.07	.11	.29	-.08	.49	-.02	-.07	.09	.29	-.09
Apologize	.05	.00	.01	.09	-.07	.45	.02	.01	.05	.09	-.08	.44	.12
Hope/optimism	-.03	.81	.05	-.07	.07	.02	-.01	.74	.05	-.08	.10	.05	-.07
Funny	-.01	.23	.17	-.11	.55	.05	-.02	.21	.17	-.10	.55	.06	-.02
Negative gossip	.11	.07	.77	.26	-.02	.14	.16	.06	.72	.23	-.02	.14	-.12
Neutral/positive gossip	-.01	.08	.63	-.10	.20	-.04	-.02	.09	.69	-.09	.19	-.04	.06
Sarcasm	.45	.10	-.03	.28	.57	-.05	.46	.11	-.04	.24	.56	-.06	.04
Brag	.66	.11	.10	.14	-.03	-.01	.72	.10	.05	.08	-.04	.00	-.07
Condescension	.52	.02	.16	.33	-.03	-.01	.60	.00	.11	.26	-.04	.00	-.14
Complain	.58	-.15	.00	.35	-.03	-.10	.56	-.16	.01	.32	-.01	-.10	.14
Criticism	.81	-.03	.31	.13	.12	.01	.83	-.06	.28	.08	.14	.04	.03
Blame	.30	-.08	.08	.80	.02	.04	.33	-.08	.08	.85	.02	.04	.01
Pessimism	.84	-.02	-.14	-.09	.07	.07	.78	-.02	-.14	-.08	.07	.05	.24
Tease	.63	.08	-.02	.20	.41	-.30	.64	.05	-.03	.16	.42	-.26	.13
Insecurity/self-doubt	.80	-.21	-.14	.02	.03	.05	.71	-.21	-.10	.03	.01	.11	.65

Note. Factor loadings > .40 are shown in **boldface**.