Bridging the gap between clinical and subclinical aversive personality research:

The Dark Factor of Personality as a common core for all socially and/or ethically aversive traits

by David D. Scholz

Inaugural dissertation

submitted in partial fulfillment of the requirements for the degree

Doctor of Philosophy (Dr. phil)

at the Department of Psychology, University of Kaiserslautern-Landau (RPTU), Germany

Thesis Defense:

September 5, 2024

Chair of the PhD Committee:

Prof. Dr. Lischetzke

Thesis Reviewers:

Prof. Dr. Julia Glombiewski

Prof. Dr. Thorsten Meiser

Advisor:

Prof. Dr. Benjamin E. Hilbig

Vom Promotionsausschuss des Fachbereichs 8: Psychologie der Universität Koblenz-Landau zur Verleihung des akademischen Grades Doktor der Philosophie (Dr. phil.) genehmigte Dissertation.

"Such is professional jealousy; a scientist will never show any kindness for a theory which he did not start himself."

From *A Tramp Abroad*
by Mark Twain

Contents

Preface

This dissertation is based on three articles that have been published, accepted, or submitted for publication in peer-reviewed journals. In the following, these articles are listed in the order in which they are presented in and appended to this dissertation. The contribution of each author for each manuscript (CRediT roles) are stated after the Acknowledgments.

(1) Scholz, D. D., Hilbig, B. E., Thielmann, I., Moshagen, M., & Zettler, I. (2022). Beyond (low) Agreeableness: Towards a more comprehensive understanding of antagonistic psychopathology. *Journal of Personality, 90*(6), 956–970. https://doi.org/10.1111/jopy.12708

(2) Scholz, D. D., Zimmermann, J., Moshagen, M., Zettler, I., & Hilbig, B. E. (in press). Theoretical and empirical integration of 'dark' traits and socially aversive personality psychopathology. *Journal of Personality Disorders*.

(3) Scholz, D. D. & Hilbig, B. E. (2024). Disentangling the shared and unique aspects of clinical and subclinical socially aversive traits relevant for (interpersonal) personality dysfunctioning. Manuscript submitted for publication

Abstract

Clinical psychology and mainstream personality research have strong research traditions in socially and/or ethically aversive personality traits (especially in personality psychology's so-called "dark" trait research). However, despite their common roots, the research programs from these two areas have developed in nearly complete isolation. Previous attempts at integration have lacked consensus between the research areas and have been limited in their theoretical foundations and comprehensiveness in the number of aversive traits investigated. Thus, I propose the *Dark Factor of Personality* (D) and its underlying theoretical framework as a means to overcome the shortcomings of these previous attempts. To this end, I present three projects, designed to integrate dark traits from personality research with aversive traits from clinical research by viewing them as *flavored manifestations* of one common aversive core (i.e., D). In Project 1, we challenged the consensus in clinical psychology that aversive traits are simply expressions of low Agreeableness from the five-factor model. Next, in Project 2, we offered an alternative by showing that the aversive traits from clinical psychology can be conceptualized as flavored manifestations of D. Last, in Project 3, we illustrated that D and its framework can serve to unify research across traditions, exemplified by demonstrating that the relationships between 20 aversive traits and (interpersonal) Personality Dysfunctioning can be, by and large, parsimoniously described by D. Overall, the D-framework offers a solution for integrating research on aversive traits across research traditions. Implications within and across research areas as well as limitations of the overall approach are discussed.

1. Introduction

1.1. Socially and/or ethically aversive trait research divided across two research traditions in psychology

Socially and/or ethically aversive behavior are great challenges for society. Ideally, individuals in any community interact with each other prosocially, thus supporting the well-being of every individual. However, in everyday life, we encounter people who act in *socially and/or ethically aversive* ways, meaning that they value their own well-being much more than others', even when enhancing their personal utility involves inflicting harm on other individuals or society as a whole (Jones & Paulhus, 2023; Moshagen et al., 2018; Thomaes et al., 2017).[1] While such aversive behaviors (e.g., counterproductive work behavior, vandalism, or tax evasion) can occur for different reasons, the area of personality psychology is concerned with systemically describing such tendencies as *personality traits.* Traits are defined as "differences among individuals in a typical tendency to behave, think, or feel in some conceptually related ways, across a variety of relevant situations, and across some fairly long period of time" (Ashton, 2013, pp. 54–55). Thus, traits enable a systematic overview and investigation of differences between people (e.g., in their tendency to act in socially and/or ethically aversive ways).

Overall, there is great interest in these socially and/or ethically aversive tendencies (herein now referred to as *aversive tendencies* for short) in both clinical and mainstream personality psychology research. In this, the current consensus in clinical psychology research is to ascribe extreme personality tendencies to personality disorders (American Psychiatric Association, 2013). Therefore, extreme aversive tendencies are an integral part of the diagnosis

[1] Some researchers have a broader understanding of the term *socially aversive*, including any tendency that may lead to interpersonal problems (Zeigler-Hill & Marcus, 2016b). A prominent example is Depressivity, as it can cause negative consequences for others, i.e., induced negative mood (see Segrin & Dillard, 1992). However, Depressivity per se does not imply an acceptance of the cost to others from one's own actions, nor does an individual typically gain much utility out of inducing negative mood in others (in fact, the opposite is often the case). Therefore, tendencies that can be understood as so-called social allergens (see O'Connor, 2011) are typically, and also herein, not meant by the terms *socially and/or ethically aversive*.

of aversive personality disorders (e.g., Antisocial or Narcissistic Personality Disorder). In mainstream personality research, however, aversive tendencies are often researched as so-called *dark traits* (Jonason, 2023; Paulhus, 2014), which supplement basic models of personality, such as the *five-factor model* (FFM; McCrae & Costa, 2008; McCrae & John, 1992) or the *HEXACO model* (Ashton & Lee, 2007; Ashton et al., 2004)[2] by focusing on less desirable (i.e., aversive) traits, e.g., "subclinical" Narcissism or "everyday" Sadism (Paulhus, 2014). In contrast to the endeavors in aversive clinical psychopathology research, these dark traits have been suggested to operate specifically at a subclinical level, and thus, "rather than being incarcerated or under clinical supervision, such individuals [with high levels of these traits] manage to survive, and even flourish, in everyday society" (Paulhus, 2014, p. 421).

Although aversive clinical psychopathology and dark trait literature are therefore concerned with a similar, if not the same, phenomena, they now stand as parallel, almost completely isolated areas of research, with only sporadic joint empirical analyses and little to no conceptual integration between them (Hilbig et al., 2020; Miller & Campbell, 2008; Thomaes et al., 2017; Zeigler-Hill & Marcus, 2016a, 2016b).[3] However, there are some obvious reasons for why such parallel research programs should be avoided. First, this practice clearly enhances the likelihood of so-called *jingle-jangle fallacies* (Kelley, 1927). A *jingle fallacy* occurs when researchers use the same label for different constructs. An example is Agreeableness, which has various, somewhat inconsistent definitions (Graziano & Tobin, 2017; Horsten, Thielmann, et al., 2023), which result in notably diverging patterns in outcome predictions (Hilbig et al., 2016; Thielmann & Hilbig, 2019; Thielmann et al., 2021). A *jangle*

[2] The FFM and the HEXACO model are the two most widely accepted basic models of personality structure, based on a lexical approach (see Ashton & Lee, 2005a). Both models assign covarying adjectives (conceptualized as lower-order traits), such as Diligence and Perfectionism, systemically to broad, and largely orthogonal, higher-order factors, e.g., Conscientiousness. The main difference between the models is that the HEXACO model assumes an additional sixth factor, i.e., Honesty-Humility (Ashton & Lee, 2005b). For more details on the difference between the models, see Ashton and Lee (2007, 2020).

[3] Note that this claim specifically refers to *aversive* clinical psychopathology research and *aversive* mainstream personality research (the latter often referred to as "dark trait research"). Clinical and mainstream personality research, *in general*, do share common frameworks (see, e.g., Krueger & Hobbs, 2020)

fallacy, on the other hand, occurs when the same construct is given different labels. A prominent example is the trait Grit, which has demonstrated meta-analytic correspondence to the FFM trait Consciousness to the degree that the two constructs can be considered equivalent (Credé et al., 2017). These examples hail from *the same* research area.[4] Arguably, these fallacies become more frequent and harder to recognize across two research areas operating in parallel.

Second, even beyond jingle-jangle fallacies, advancement in one aspect of a phenomenon (e.g., antecedents, development, or outcomes of aversive tendencies) may remain completely unknown to another area. One striking example of this is the discussion of whether and which aversive traits can be considered maladaptive. Some researchers have suggested that dark traits are more adaptive than traits used in clinical research for diagnosing (aversive) personality disorders (Furnham et al., 2013; Paulhus & Williams, 2002; Pincus et al., 2009). Some researchers have subsequently used this view to justify the separation of research traditions (Jones & Paulhus, 2023; Paulhus, 2023). Consequently, different scales have been developed, e.g., for Narcissism (as a clinical construct) vs. subclinical Narcissism (as a dark trait) (Miller & Campbell, 2008; Miller et al., 2013; Miller et al., 2016). However, not many studies have investigated how dark traits are related to psychopathology (Hilbig et al., 2020; Miller & Campbell, 2008; Thomaes et al., 2017) simply *because* they are viewed as "subclinical" and are therefore typically not included in clinical psychopathology research. Thus, although the separation of clinical vs. subclinical (aversive) traits is largely based on historical reasons (outlined in the following sections), it maintains itself due to systematic differences in research topics between fields: psychopathology in clinical psychology vs. "everyday" outcomes (e.g., work behavior) in dark personality research (see Furnham et al., 2013; Thomaes et al., 2017). Consequently, the relationships between dark traits and

[4] I want to stress, however, that this issue is not specific to personality psychology. Other areas of psychology are prone to this issue as well; See, e.g., Hagger (2014) for examples in social psychology, Casper et al. (2018) for examples in work psychology, and Kongerslev et al. (2015) for examples in clinical psychology.

psychopathology remain unclear, whereas these associations are well-researched for aversive traits from clinical research.

In sum, parallel research programs clearly lead to a waste of (temporal and financial) resources and hamper scientific progress. Thus, they should be avoided. Hereby, a shared understanding of the commonalities and differences in aversive traits (within and across research areas) would be useful and would offer a first step toward integration.

Against this background, I propose a common theoretical framework for all aversive traits, integrating aversive personality research *across* research traditions. To this end, I will first provide a brief outline of the history of aversive clinical psychopathology and dark trait research (from their common roots to the nearly orthogonal research areas that exist today) and then highlight problems underlying previous attempts to integrate the two research areas. Thereafter, I will present a theoretical framework suitable for uniting research on aversive traits across research traditions and showcasing its implications across three projects.

1.2. A brief history of aversive traits in clinical vs. broad personality research

1.2.1. Common roots: The emergence of aversive "subclinical" traits

Around the turn of the current millennium, there was great interest in the assessment of aversive personality tendencies (R. M. Kowalski, 2001; Paulhus & Williams, 2002). Researchers came to suggest that individuals behave in aversive ways, which are quite similar to patterns described in clinical psychology, such as *Narcissistic Personality Disorder*, which is characterized by a pattern of grandiosity, need for admiration, and low empathy (American Psychiatric Association, 1994). However, the current clinical consensus at that time, exemplified by the current versions of the leading diagnostic systems at that time, i.e., the DSM-IV (American Psychiatric Association, 1994) and the ICD-10 (World Health Organization, 1992), stated that such disorders are conceptualized as categories, i.e., an individual either meets or does not meet the criteria of these disorders, and, as such, the individual needs (or does not

need) clinical help and treatment due to severe disturbance of the development of a "functioning" personality (see also Tyrer et al., 2015). In other words, individuals meeting the criteria of a personality disorder were seen as possessing "abnormal" characteristics that are not apparent in non-clinical populations (see, e.g., Gunderson & Zanarini, 2011). Conversely, *Subclinical Narcissism* (Raskin & Hall, 1979), defined by Grandiosity, Entitlement, Dominance, and Superiority, and later *Subclinical Psychopathy* (Ray & Ray, 1982), defined by (high) Impulsivity, (high) Thrill-seeking, (low) Empathy, and (low) Anxiety, were introduced as traits found in subclinical populations. Thus, these "subclinical" traits were conceptualized as distinct from corresponding clinical diagnoses (Paulhus & Williams, 2002). When grouped with Machiavellianism, a trait based on the strategic, though callous political strategies described by Niccolò Machiavelli (Christie & Geis, 1970), these three (subclinical) traits were subsumed under the umbrella term of the so-called *Dark Triad* (Paulhus & Williams, 2002). Thereby, one could argue that, at that time, it made sense for personality researchers to establish their own research tradition for aversive tendencies (named "dark") by reconceptualizing clinical constructs as traits that fell along a continuous spectrum ranging from less aversive to more aversive, thereby avoiding any conflict with the categorial view in clinical psychology.

1.2.2. Development in clinical psychology: Shifting toward dimensional models in clinical personality psychopathology research

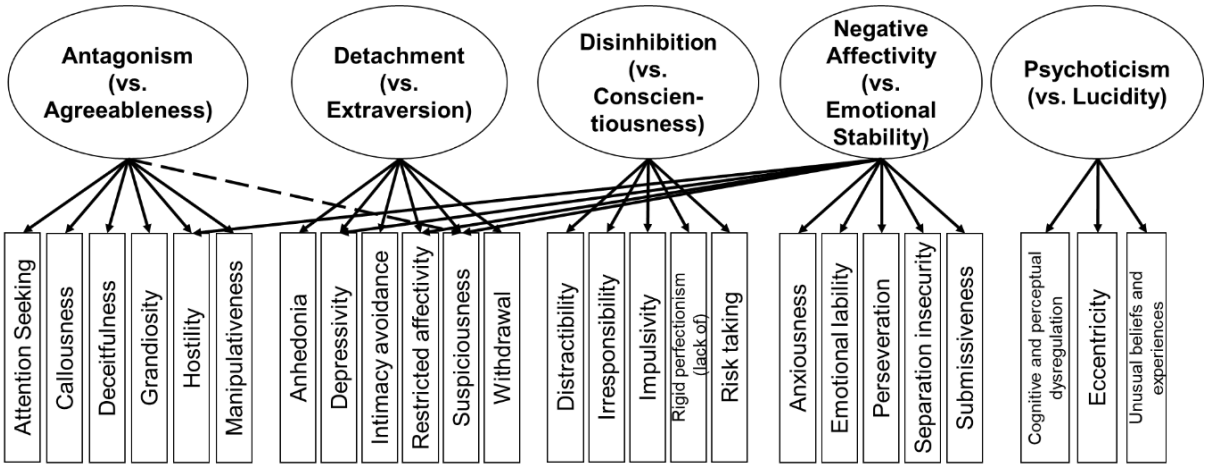
However, in the early 21st century, clinical personality psychopathology research was hotly debated. The success and wide adoption of the lexically based models in personality psychology, especially the FFM, did not come without impact. Researchers came to propose personality disorders as expressions of extreme versions of the FFM dimensions (e.g., McCrae, 1991; Miller et al., 2001; Wiggins & Pincus, 1989), e.g., Antisocial Personality Disorder ought to be characterized (mainly) by low Agreeableness and low Conscientiousness (Widiger et al., 2002). Soon, meta-analytic evidence followed, clearly indicating that the personality characteristics described in personality disorders could indeed be generally understood in terms

of extreme trait expressions of the FFM (Samuel & Widiger, 2008; Saulsman & Page, 2004). Moreover, not only was there evidence supporting a dimensional view of the features of personality disorders, but there was also a *lack of evidence* for a categorical view on personality disorders to begin with, the most important criticisms of these categorial view being a lack of reliability and validity, high comorbidity, and within-disorder heterogeneity (L. A. Clark, 2007; Trull & Durrett, 2005; Widiger & Samuel, 2005).

Following this, a proposal for a dimensional view of personality disorders was brought forward: The *Alternative Model of Personality Disorders* (AMPD) in Section 3 of the fifth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5; American Psychiatric Association, 2013). It classifies personality disorders on the basis of two main criteria: first, *Personality Dysfunctioning,* which indicates the *severity* of a personality disorder (Criterion A); second, there are 25 so-called *Maladaptive Trait Facets* mapped onto (one or more) of five higher-order dimensions (i.e., Antagonism, Detachment, Disinhibition, Negative Affectivity, and Psychoticism), ought to represent maladaptive versions of the FFM (see Figure 1), which indicate the *style* of a personality disorder (Criterion B). Although the AMPD was implemented only as an *emerging* model in the DSM-5 (American Psychiatric Association, 2013), it is now the *guiding* research model of current clinical research on personality psychopathology (L. A. Clark & Watson, 2022; Krueger & Hobbs, 2020; Zimmermann et al., 2019), leading to an ever-growing consensus to renounce the categorical view of personality disorders in clinical personality psychopathology research (Hopwood et al., 2018). Correspondingly, the newest revision of the *International Classification of Diseases* (ICD-11; World Health Organization, 2023) has adopted a completely dimensional view of personality disorders, and so has the approach used to structure psychopathology in general, not limited to *personality* psychopathology, i.e., the *Hierarchical Taxonomy of Psychopathology* (HiTOP; Kotov et al., 2017).

Figure 1

Mapping of the 25 Maladaptive Traits Facets of the Alternative Model of Personality Disorders (AMPD) onto five higher-order domains as counterparts to the dimensions from the Five Factor model (see American Psychiatric Association, 2013).



Note. The dashed line indicates that Suspiciousness is not part of Antagonism as per the AMPD, but it is part of many other conceptualizations of Antagonism (see Krueger et al., 2021; Sleep et al., 2020; Widiger & Simonsen, 2005).

1.2.3. Developments in dark trait research: Moving beyond the Dark Triad

Somewhat in parallel to this shift in clinical personality psychopathology research, in mainstream personality research, aversive trait research, still referred to as dark trait research, became more popular (Muris et al., 2017; Paulhus, 2014). However, this interest was no longer limited to the Dark Triad. Suggestions to move beyond a mere triad were brought forward, e.g., in form of the *Dark Tetrad* (Buckels et al., 2013), expanding the Dark Triad by adding the trait *Sadism* ("the enjoyment of other people's suffering" Paulhus & Dutton, 2016, p. 109); or even in the form of a "Big Tent" of dark traits (Marcus & Zeigler-Hill, 2015) by including *Spitefulness* ("a preference that would harm another but that would also entail harm to oneself"; Marcus et al., 2014, p. 566) or *Greed* ("a desire to acquire goods, status, or power for the sake of acquisition without much regard for other individuals or the common good"; Marcus & Zeigler-Hill, 2015, p. 436), among many others (see Moshagen, Zettler, & Hilbig, 2020; Zeigler-Hill & Marcus, 2016a, 2016b).

However, with the introduction of an increasing number of dark traits, questions about the distinctiveness of these traits became more pressing (C. M. Kowalski et al., 2021; Moshagen

et al., 2018), an issue that had existed since the Dark Triad's inception (Muris et al., 2017), as, e.g., Psychopathy and Machiavellianism often lack distinctive features or nomological networks (Miller et al., 2019; Vize et al., 2018). Nevertheless, there was at least the consensus that all dark traits share a common core (Jones & Figueredo, 2013; C. M. Kowalski et al., 2021; Moshagen et al., 2018; Muris et al., 2017; Vize, Collison, Miller, & Lynam, 2020). Thus, there were different suggestions for what this common core exactly is (see Horsten, 2023, for an overview), some suggesting factors from the existing FFM and HEXACO model, i.e., FFM-Agreeableness (Vize, Collison, Miller, & Lynam, 2020) and Honesty-Humility (Muris et al., 2017), respectively. However, there was no (direct) acknowledgment of the higher-order dimensions from the AMPD as candidates for the core of dark traits, let alone the suggestion that aversive AMPD traits (e.g., Deceitfulness or Manipulativeness) should be considered (potential) dark traits. One may interpret this as a consequence of the already apparent divide between aversive clinical and dark trait research at that time, resulting in a disregard for the constructs that were being studied in the respective other field.

1.2.4. The current state of "clinical" and "subclinical" aversive traits: Consensus and disagreement about the common core of aversive traits

The separation of the two research areas, however, has not gone completely unrecognized. Although only sporadically, some researchers have tried to rediscover common ground and have emphasized the importance of dark trait research for clinical psychopathology research (and vice versa; see Thomaes et al., 2017; Zeigler-Hill & Marcus, 2016b). Hereby, it has been suggested that aversive traits from the two research traditions are conceptually similar and thus share a common "antagonistic" core (Thomaes et al., 2017; Zeigler-Hill & Marcus, 2016b). In line with this idea, empirical research has indeed indicated great overlap between aversive clinical and dark traits (Anderson et al., 2014; Rose et al., 2023; Russell et al., 2017; Russell & King, 2017; Somma et al., 2020; Vize, Collison, & Lynam, 2020; Vize, Collison, Miller, & Lynam, 2020; Rose et al., in press), especially those with overlapping features, e.g.,

manipulating others to achieve one's ends is an aspect of both AMPD-Manipulativeness and Machiavellianism as part of the Dark Triad.

The details, however, remain a topic of dispute, especially regarding what actually accounts for the variance that all these aversive traits share. Featured most prominently in this debate, FFM-Agreeableness has been proposed as the core of the Dark Triad (Vize, Collison, & Lynam, 2020). Moreover, the AMPD implies that Antagonism, as the shared aspect of many aversive clinical traits, is merely a reflection of low FFM-Agreeableness (American Psychiatric Association, 2013; Krueger & Hobbs, 2020; Zimmermann et al., 2019). Thus, one could argue that FFM-Agreeableness is the common element shared by *all* aversive traits and is therefore a suitable candidate for uniting aversive traits from the two research traditions (see, e.g., Lynam & Miller, 2019).

However, there are some issues with this conclusion. First, it is largely based on Dark Triad research, which is not representative of the full spectrum of the aversive personality space (Horsten, Moshagen, et al., 2023; Marcus & Zeigler-Hill, 2015). Thus, inferring the core of all dark traits from results based on the Dark Triad (and sometimes one additional trait) is somewhat limited to begin with. Moreover, although FFM-Agreeableness has shown the largest associations (of all the FFM dimensions) with the Dark Triad, with both the individual traits and their shared variance, other dimensions are also associated with their shared variance, especially low Conscientiousness (Furnham et al., 2014; Jones & Figueredo, 2013; Paulhus, 2014; Schreiber & Marcus, 2020). Thus, the core of the Dark Triad is a mix of several broad personality domains.[5] However, a meta-analysis revealed that the core of the Dark Triad is in fact not well covered by the whole FFM space (Schreiber & Marcus, 2020), especially when compared with the coverage achieved by the HEXACO model. Hereby, Honesty-Humility

[5] Indeed, it is only when a notably broad measure of Agreeableness, beyond the FFM as suggested by Crowe et al. (2018), is used that the common core of the Dark Triad can be sufficiently approximated (Vize, Collison, Miller, & Lynam, 2020). As this measure includes facets from other broad dimensions, such as Honesty-Humility, it again underscores the idea that this core is best represented by a mix of broad traits.

shows the strongest association with this common core (Muris et al., 2017; Paulhus & Klaiber, 2020; Schreiber & Marcus, 2020), but again, other dimensions, most notably Agreeableness and Conscientiousness, are associated with this core as well (see Schreiber & Marcus, 2020, for a meta-analysis). Therefore, in terms of the HEXACO model as well, the shared variance of the Dark Triad traits is best understood as a mix of several dimensions.

1.3. Failed integration and how to move forward

Taken together, there is ample evidence against the notion that FFM-Agreeableness, or any single FFM or HEXACO dimension, constitutes the core of the Dark Triad, let alone *all* dark traits. By implication, FFM-Agreeableness (or Honesty-Humility) is also clearly insufficient to account for the aversive – or "antagonistic" (Thomaes et al., 2017; Zeigler-Hill & Marcus, 2016b) – common core of aversive traits between clinical and mainstream personality research. Thus, despite agreement on the large conceptual and empirical overlap of aversive traits identified in dark trait and clinical personality psychopathology research, the constitution of their aversive core remains unclear. Therefore, the two research fields remain theoretically unintegrated and thus have overlapping, yet largely isolated, research agendas on the same phenomenon (i.e., aversive personality). Against this background, I propose that the theoretical framework of the *Dark Factor of Personality* (D; Moshagen et al., 2018) can offer a solution for uniting the largely isolated research areas.

1.4. D as the common core of all aversive traits

1.4.1. The substantive meaning of D

As mentioned before, there is at least consensus that dark traits do share a common, aversive core (e.g., Jones & Figueredo, 2013; C. M. Kowalski et al., 2021; Muris et al., 2017; Thomaes et al., 2017; Vize, Collison, Miller, & Lynam, 2020). Moshagen et al. (2018) comprehensively defined this core as "the general tendency to maximize one's individual utility – disregarding, accepting, or malevolently provoking disutility for others – , accompanied by beliefs that serve as justifications" (p. 657) and labeled it D. Hereby, utility is meant in a very

broad sense and is not limited to visible gains (e.g., material gains or higher status), but also includes psychological utilities (e.g., pleasure). In the same manner, disutility implies anything (potentially) negative for another individual or society at large, whether they are aware of it (e.g., public vandalism) or not (e.g., undetected tax evasion). In other words, *any* behavior that serves to enhance one's own utility *at the expense of others* is accounted for by D. This conceptualization has been supported by multiple studies that have indicated that D strongly predicts a variety of aversive outcomes that refer to diverse forms of utility maximization at others' expense, both self-reported (e.g., violence, internet trolling, stalking) and consequential behaviors (e.g., dishonest, sadistic, or antisocial behavior in incentivized economic games), leaving little to no additional variance to be explained by other aversive traits (e.g., Bader, Hartung, et al., 2021; Bader, Horsten, et al., 2021; Hilbig et al., 2023; Moshagen et al., 2018; Moshagen, Zettler, & Hilbig, 2020; Scholz et al., 2023; Schrödter et al., 2021).

Moreover, the second part of the D-definition referring to justifying beliefs is based on well-established theories and empirical findings that, in general, people aim for a positive moral self-view (e.g., Allport, 1955; Bandura, 2016; Prentice et al., 2019; Tappin & McKay, 2017). When this view is threatened through their own (aversive and morally questionable) behavior, people generally, although to varying degrees (Bandura, 1999, 2002), tend to construct justifications for that behavior (Bandura, 2016; Barkan et al., 2015; Shalvi et al., 2015). For D, it is assumed that this tendency is pronounced, enabling individuals high in D to engage in selfish and malevolent behavior but still uphold a positive (moral) self-view (Hilbig et al., 2022; Moshagen et al., 2018). By implication, *any* belief that may justify such behavior is entailed by D. Note that the term "belief" in the context of D is also meant broadly to reflect "any patterns of thought than can justify aversive behavior" (Hilbig et al., 2022, p. 2731), therefore including attitudes, attributions, schemas and (world)views. In line with this idea, D has been shown to be strongly associated with a variety of beliefs that justify aversive behavior, e.g., that others

cannot be trusted (Hilbig et al., 2022; Horsten, Thielmann, et al., 2023; Rudloff et al., 2022, 2023; Scholz et al., 2022; Thielmann & Hilbig, 2023; Moshagen et al., in press; Scholz, Kraus, & Miller, in press). It has further been shown that these beliefs are indeed used by people high in D to justify corresponding aversive behavior (Hilbig et al., 2022).

Taken together, D is defined by the conjunction of two broad underlying *motivations*, i.e., first, to maximize utility at another's expense, and second, to maintain a positive moral self-view (by finding beliefs to justify aversive behavior). This is an important feature of D that sets it apart from the previously mentioned proponents of the common core of all aversive traits (i.e., FFM-AG and Honesty-Humility), which are defined by more narrow sets of (highly correlated) adjectives or traits, e.g., kind, sympathetic, undemanding, and warm for FFM-Agreeableness (Saucier, 2002), and sincere, fair-minded, unassuming, and modest for Honesty-Humility (Ashton & Lee, 2007). Consequently, these factors refer only to some *specific expressions* of utility maximization at others' expense (e.g., cheating for monetary gains) or justifying beliefs (e.g., modesty vs. grandiosity). By contrast, D refers to *any* form of utility maximization at others' expense, including gaining utility from others' disutility (e.g., Sadism or Spitefulness) and *any* belief that may justify such a behavior (e.g., that the world is a dangerous place). Neither example is covered well by FFM-Agreeableness or Honesty-Humility on a substantive level (Horsten et al., 2021; Moshagen, Zettler, Horsten, & Hilbig, 2020). Thus, similar to what was already discussed in regard to the Dark Triad, to conceptualize D in terms of broad personality models such as the FFM or HEXACO, D is a mix of dimensions, i.e., mainly low Agreeableness and low Conscientiousness (and, to a lesser degree low Extraversion as well as low Neuroticism) in the FFM, and mainly low Honesty-Humility, low Agreeableness, and low Conscientiousness (and to a lesser degree low Emotionality) in the HEXACO model (Moshagen et al., 2018). In other words, someone high in D is characterized by the presence of tendencies such as irritability, entitlement, and distrust, while lacking

tendencies, e.g., kindness, rule-following, and honesty. By implication, D cannot be equated to any single dimension from the FFM or HEXACO (Hilbig et al., 2021; Horsten, 2023; Horsten et al., 2021; Horsten, Thielmann, et al., 2023; Moshagen et al., 2018; Moshagen, Zettler, Horsten, & Hilbig, 2020; cf. Rose et al., 2022; Vize & Lynam, 2021). Crucially, D is not fully accounted for by a mix of other personality dimensions but also offers incremental value over them, e.g., in predicting justifying beliefs over and above the entire HEXACO space (Hilbig et al., 2022). Taken together, D is best *approximated* by a mix of broad personality dimensions (but still offers incremental value over them) and is compressively described by maximizing one's own utility at the expense of others, accompanied by justifying beliefs.

Importantly, it is the theoretical framework of D (herein referred to as the *D-framework*), which specifies D as the common core of *all* aversive traits, that makes it especially suited for uniting aversive personality research across research fields.

1.4.2. The theoretical framework of the aversive core: Fluidity and flavored manifestations

The D-framework includes a variety of (testable) assumptions about D. Most importantly for the current matter, D, as the common core of all aversive traits, is specified as a *fluid* construct (Moshagen et al., 2018; Moshagen, Zettler, & Hilbig, 2020). This terminology was adopted from the ability domain and specifically the two-factor theory of intelligence (Carroll, 2009; Spearman, 1927). Hereby, fluid means that D (like the g-factor of intelligence) is independent of a particular set of items. Instead, it is always approximated when there is a sufficient number of distinct enough traits or items that refer to some aspect of the D-definition (see Hilbig et al., 2020; Moshagen et al., 2018). Indeed, recent research has shown that (any) six aversive traits are typically enough to approximate D (Horsten, Moshagen, et al., 2023).

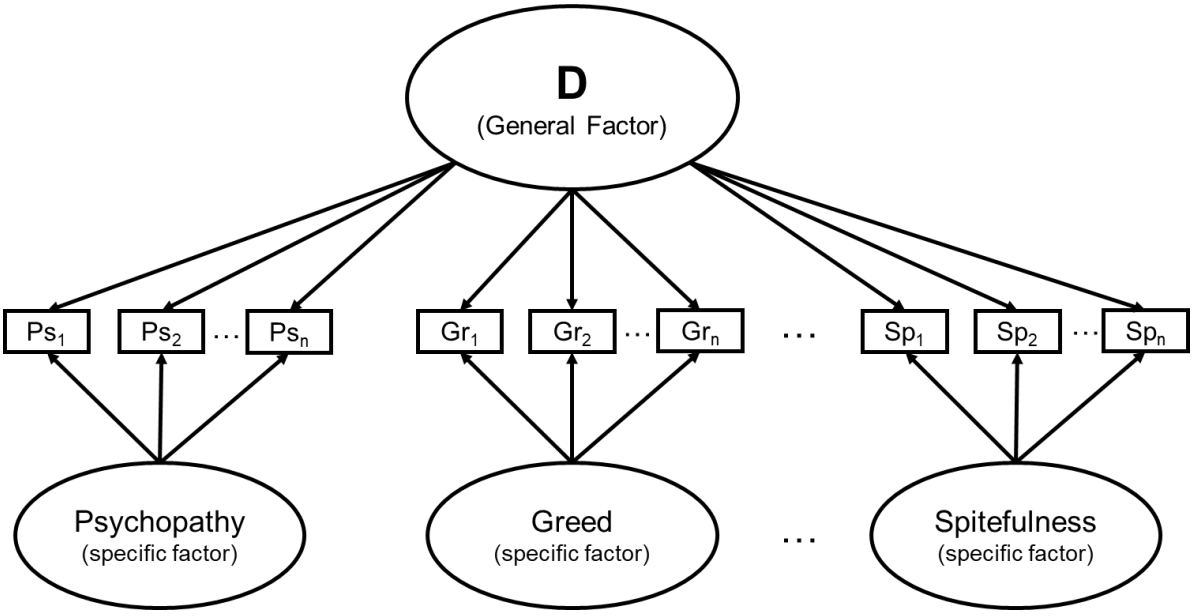
Given this fluid nature, it follows that any trait that is in line with the D-definition can be seen as a *flavored* manifestation of D (again a term borrowed from the ability domain; see Carroll, 2009). "Flavored" within the D-framework means that (a) each trait may emphasize a

different aspect of D (e.g., provoking disutility is a key feature of Sadism, whereas justifying beliefs, e.g., entitlement, are much more pronounced in Narcissism) and (b) that each trait may entail aspects beyond D, e.g., Psychopathy also entails Impulsivity, which has been dissociated from D (Moshagen et al., 2023), and thus Psychopathy can be conceptualized as a manifestation of D *flavored* by Impulsivity (Bader et al., 2023).

Thus, given any set of distinct enough aversive traits, D emerges as the *only aspect* shared by *all* traits, while each trait may entail unique "flavors" not shared by all other aversive traits. This has some important implications, some of which are directly reflected in the original statistical modeling approach for D, herein referred to as the "base model" (see Figure 2).

Hereby, D is modeled as a general factor (g-factor) in a bi-factor model over a large set of items that stem from different dark traits. Each item is allowed to load on D (as the g-factor) and its respective scale (as a specific factor). All factors are specified to be orthogonal to each other. Thus, D represents the *only* variance shared by all items (and thus traits), while the specific factors indicate the unique variance of each trait, i.e., not shared by the other traits or

Figure 2
Simplified structure of the typical bi-factor modeling approach for D (i.e., the base model).



Note. Error terms and variances have been omitted for the sake of clarity.

D. This assumption, i.e., that D is the only (aversive) aspect shared by all aversive traits, has gained a lot of empirical support. Multiple studies have demonstrated excellent fit of such a bi-factor model, even when up to 20 aversive traits are considered in tandem, and D has been identified as a strong general factor, i.e., subsuming most of the shared variance of all items of dark traits, especially those that are in line with its definition (Horsten et al., 2021; Moshagen et al., 2018; Moshagen, Zettler, & Hilbig, 2020; Scholz, Zimmermann, et al., in press). Moreover, the residualized facets (i.e., the specific factors) overall do not predict any meaningful variance beyond D in aversive outcomes (Moshagen et al., 2018). Crucially, this indicates that D is indeed the *single aversive core* of these traits, i.e., there is little to nothing shared between aversive traits besides D, and D captures the "aversive essence" of these traits (Hilbig et al., 2023). Along these lines, Zettler et al. (2021) showed that D could even longitudinally predict aversive traits that were not part of its model. For example, D was found to longitudinally predict Sadism as well as Sadism predicted itself, even when Sadism was not part of the bi-factor model indicating D. This finding emphasizes that D is an *underlying* disposition of aversive traits, from which aversive traits arise as (flavored) manifestations.

1.5. A brief overview of the dissertation projects

Taken together, the D-framework specifies some clear attributes of D, as the common aversive core, and aversive traits as its flavored manifestations. In this, in contrast to other frameworks, specifying broad and largely orthogonal factors to describe the breadth of human personality (i.e., the FFM or HEXACO model), the D-framework is concerned (both conceptually and statistically) only with disentangling aversive aspects (in the sense of the D-definition, i.e., utility maximization at the expense of others accompanied by justifying beliefs) from non-aversive ones. By offering a broad yet specific substantive definition of this common core and viewing any trait that is in line with the D-definition as a *flavored manifestation* of it, D offers a shared understanding of aversive traits, independent of their research tradition.

Thus, the main aim of the current dissertation was to test the implications of the D-framework for aversive traits specified in clinical personality research, especially the view that any trait that lines up with the D-definition is a flavored manifestation of D, thereby also outlining the D-framework's potential to function as a framework that holds *across* research traditions. To this end, we conducted three projects: Project 1 challenged the AMPD's proposition that aversive clinical traits (primarily those subsumed under the Antagonism domain) are merely a reflection of low FFM-Agreeableness, thus further stressing the need for another conception of a common aversive core. In Project 2, we tested the D-framework's claim that any aversive trait, thus also aversive AMPD traits, can be viewed as flavored manifestations of D and, thus, the question of whether D can account for the common core of aversive traits across research traditions. Last, in Project 3, we indicated that D, as the shared aspect of 20 aversive traits (from clinical and dark trait research), offers a general parsimonious account of how aversive traits are related to (interpersonal) personality psychopathology. As such, Project 3 exemplifies the potential of D and its framework to unite different research traditions, and challenges the value of separating aversive traits into clinical and subclinical variants (cf. Gunderson & Zanarini, 2011; Jones & Paulhus, 2023; Paulhus, 2023).

To this end, we relied on three different sources of data (see Table 1). Some of these data were used in multiple projects. Nonetheless, each project yields unique research questions and analyses that are not presented in the other projects. The first source of data, used in all projects, is the *Prosocial Personality Project* (PPP; see https://osf.io/m2abp/), a multi-wave online panel containing different variables related to pro- or antisocial tendencies over the course of around three years (2019 to 2022). The second source of data, used in Projects 2 and 3, constitutes an assessment of aversive traits, personality psychopathology, and aversive behavior. These different variables are distributed over two waves of assessment (one-week gap between measurement occasions). The third source of data, used only in Project 3, is

darkfactor.org, which is a website containing information about D. People interested in their level of D can answer an online questionnaire about different constructs and receive feedback on their individual D score. Data overlap across the projects is transparently reported in all manuscripts as well as the present dissertation (see Table 1). Moreover, all data sets, analysis scripts, as well as all materials used in the manuscripts included in this dissertation are publicly and freely available on the Open Science Framework (OSF). The respective repositories and – whenever applicable – preregistrations are referred to and linked in the respective manuscripts.

Table 1

Data sources for the three projects included in the current dissertation.

	Prosocial Personality Project (PPP)	Two-wave sample	darkfactor.org
Project 1	Data used from waves T1 and T2; N = 3,396	-	-
Project 2	Data used from waves T2, T3, T5, and T7b for the pilot; N = 1,781	Data used from both waves for the main study; N = 2,006	-
Project 3	Data used from waves T1, T2, T3, and T7b for Study 1; N = 604	Data used from both waves for Study 2; N = 2,006	Data used from three waves for Studies 3a-c; N = 639–888

2. Projects

2.1. Project 1 – Challenging the (clinical) consensus: Antagonistic traits as merely low FFM-Agreeableness?

Manuscript: Scholz, D. D., Hilbig, B. E., Thielmann, I., Moshagen, M., & Zettler, I. (2022). Beyond (low) Agreeableness: Towards a more comprehensive understanding of antagonistic psychopathology. *Journal of Personality, 90*(6), 956–970. https://doi.org/10.1111/jopy.12708

As described above, one core issue that has hindered integration between aversive psychopathology and dark trait research is the view upheld in clinical psychology that aversive clinical traits share low FFM-Agreeableness as a common core (American Psychiatric Association, 2013; Krueger & Hobbs, 2020; Zimmermann et al., 2019). While it has been shown that FFM-Agreeableness is not sufficient to account for the core of dark traits (e.g., Horsten, 2023; Muris et al., 2017), the view that it reflects aversive traits from clinical research, especially those from the Antagonism domain (see Figure 1) – hereafter referred to as *antagonistic traits* – , remains unchallenged. Whereas some of these antagonistic traits are indeed covered well by FFM-Agreeableness on a substantive level (e.g., Hostility), others are not captured the substantive definition of FFM-Agreeableness (e.g., Manipulativeness) but are covered (better) by other factors, also suggested to approximate or directly represent the aversive core, i.e., Honesty-Humility (Schreiber & Marcus, 2020) and D (Moshagen et al., 2018) respectively. Thus, we compared the substantive overlap of these three "candidates" (FFM-Agreeableness, Honesty-Humility, D) with the antagonistic traits (see Table 2).

In short, both FFM-Agreeableness and Honesty-Humility refer to some, but not all, aspects of the antagonistic traits. FFM-Agreeableness mainly shows substantive overlap with the primarily affective antagonistic traits, whereas Honesty-Humility mainly shows substantive overlap with the primarily behavioral antagonistic traits. In other words, many of the antagonistic traits do not overlap with the substantive definitions of either candidate. Thus, similar to what was discussed in regard to dark traits, antagonistic traits appear to be accounted

Table 2

Substantive overlap of the antagonistic traits with FFM-Agreeableness, Honesty-Humility, and the Dark Factor of Personality (D).

Primary aspect	Antagonistic trait	Overlap FFM-Agreeableness	Overlap Honesty-Humility	Overlap D
Affect	Callousness	+++		+
	Hostility	+++		+
Behavior	Manipulativeness		+++	+
	Deceitfulness		+++	+
	Attention Seeking		+++	
Cognition	Grandiosity		+	+++
	Suspiciousness	+		+++

Note. +++ indicates the strongest substantive overlap of one of the three candidates with the respective trait; + indicates that there is substantive overlap with the respective trait, but it is not the strongest of all the candidates.

for by a mix of dimensions. By contrast, D overlaps with all antagonistic traits on a substantive level (with the exception of Attention Seeking); thus, it offers the most balanced representation of antagonistic traits. Consequently, we expected D to account for incremental variance in the traits that were not covered well by FFM-Agreeableness (i.e., the primarily behavioral and cognitive traits) or Honesty-Humility (i.e., the primarily affective and cognitive traits).

To test these hypotheses empirically, we applied structural equation modeling (SEM) to a sample from the PPP (N = 3,396; see Table 1). Hereby, we regressed the antagonistic traits on the three candidates (i.e., FFM-AG, Honesty-Humility, and D) in separate latent regression models (to obtain zero-order effects). We also predicted the antagonistic traits from two or all three candidates at once (to obtain unique and incremental effects). As expected, all candidates, accounted sufficiently ($|r| > .50$) for the traits with which they had substantive overlap. The only exception was Attention Seeking, which was only weakly to moderately correlated with the three candidates ($|r| = .14 – 35$). Crucially, as predicted, D had incremental variance over and above FFM-Agreeableness and Honesty-Humility in their less pronounced trait groups.

Moreover, D explained the highest median variance and offered the most balanced distribution of explained variance across all antagonistic traits.

In conclusion, some of the antagonistic traits were *related* to Agreeableness but are clearly more than just an expression of low FFM-Agreeableness. In fact, many traits are best accounted for by other personality factors (e.g., Deceitfulness by Honesty-Humility and D, or Suspiciousness only by D), and yet other traits show only a very limited overlap with FFM-Agreeableness to begin with (i.e., Attention Seeking). Therefore, concluding that antagonistic traits are merely low FFM-Agreeableness is not supported by its substantive definition or by empirical overlap.

2.2. Project 2 – Offering an alternative: D as the common core of *all* aversive traits

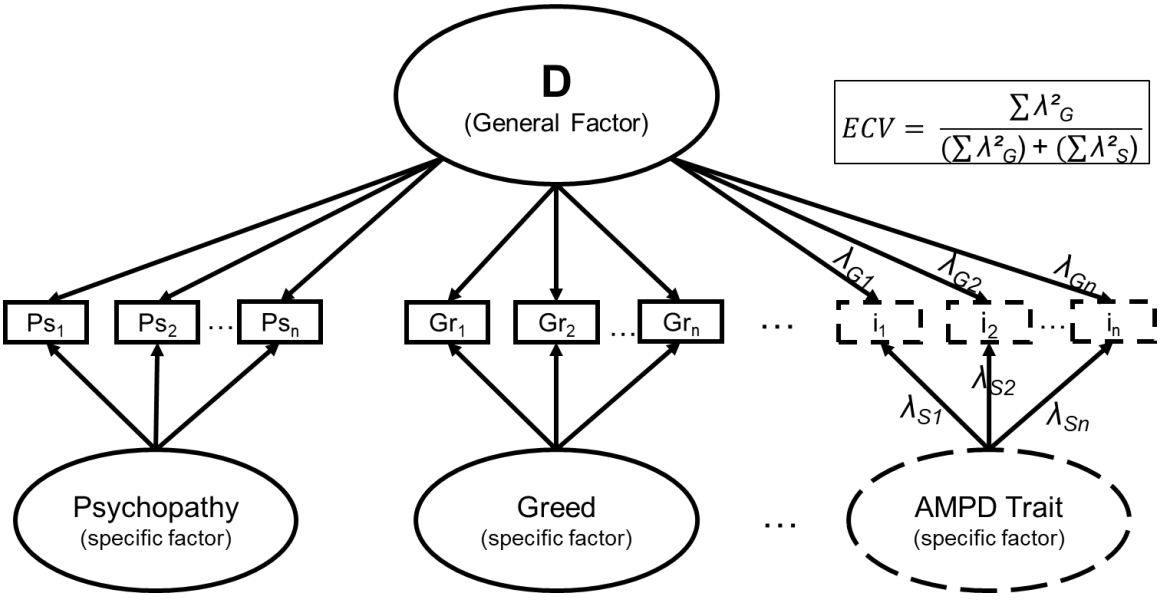
Manuscript: Scholz, D. D., Zimmermann, J., Moshagen, M., Zettler, I., & Hilbig, B. E. (in press). Theoretical and empirical integration of 'dark' traits and socially aversive personality psychopathology. *Journal of Personality Disorders*.[6]

So far, it is only clear that antagonistic traits are *not* (merely) low FFM-Agreeableness. However, the question that still remains is how aversive traits from clinical research can (instead) be conceptualized in a framework that is able to unify accounts from research on aversive clinical and dark traits. Previous approaches that have attempted to specify a common core (see Horsten, 2023, for an overview) have been limited in that the common element of aversive traits is specified on aspects shared by *some*, but not *all* aversive traits (e.g., low Agreeableness), and empirical investigations have typically been limited to the Dark Triad, plus one other dark trait at best. However, D is conceptually defined as the core of *all* aversive traits, and thus, the implications of its theoretical framework *must* apply to aversive traits from clinical psychology as well.

[6] To be more attentive to the criticism regarding the usage of the term "dark" in the context of clinical psychopathology research (see, e.g., Miller et al., 2023; Thomaes et al., 2017), we refer to dark traits as *aversive traits from mainstream personality research* in the respective manuscript.

Therefore, we applied the D-framework to the AMPD by viewing the aversive AMPD traits as flavored manifestations of D and tested different implications of the framework. First, we tested whether each aversive AMPD trait could be incorporated into the base model of D separately (see Figure 3). Thereby, any trait in line with the D-definition should be sufficiently accounted for by D (*ECV* > .20)[7] and should not distort model fit. Second, D is proposed to capture anything aversive in the sense of utility maximization at others' expense. Thus, none of its manifestations should be able to account for non-trivial explained variance (i.e., $R^2 \geq 2\%$; see Cohen, 2013) over and above D in predicting this very tendency. For this tendency, the *Social Value Orientation* (SVO) task (Murphy et al., 2011) is a prime manifestation (see Hilbig et al., 2023), as the amount of money participants can earn in this task completely depends on their decisions to value their own profit more than others'.

Figure 3
The typical bi-factor model of the Dark Factor of Personality (D) extended by an aversive trait from the Alternative Model of Personality Disorders (AMPD).



Note. The dashed lines indicate that each aversive AMPD trait was introduced into the base model (see Figure 2) separately as an additional specific factor, i.e., each trait had its own model for evaluating the hypotheses. Error terms and variances have been omitted for the sake of clarity.

[7] The *Explained Common Variance* (*ECV*) is a measure in latent modeling of how well a g-factor accounts for the shared variance in a given set of items, e.g., the items of one trait. Thus, its interpretation is similar to that of R^2 from linear regression, i.e., as the proportion of variance "explained" by the g-factor.

To test these propositions, we relied on another sample from the PPP (N = 1,781) for a pilot study and on the sample from the two-wave study (N = 2,006) for the main study (see Table 1). We aimed for a broad and diverse reflection of aversive traits in both studies. Therefore, we concurrently assessed twelve dark traits and eight aversive AMPD traits, i.e., the seven antagonistic traits (see Project 1) plus Irresponsibility, which is usually not associated with Antagonism but clearly entails aspects in line with the D-definition, i.e., utility maximization at the cost of others. Moreover, we used incentivized SVO tasks in both studies.

SEM showed that the D-framework's implications held for all (aversive) AMPD traits, in line with the D-definition. Thus, the aversive clinical traits could be satisfactorily integrated into the bi-factor model, indicating that they could indeed be conceptualized as flavored manifestations of D. Moreover, none of the aversive traits predicted non-trivial variance over and above D in SVO (all R^2s ≤ 1.3%), indicating that D was their underlying *aversive* core. In conclusion, based on the D-framework, it is clear that *all* aversive traits – including AMPD traits – share a common core, representing the tendency to maximize one's own individual utility at the expense of others accompanied by justifying beliefs.

2.3. Project 3 – Exemplifying the potential of D across research traditions: The relationship between aversive traits and personality psychopathology

Manuscript: Scholz, D. D. & Hilbig, B. E. (2024). Disentangling the shared and unique aspects of clinical and subclinical socially aversive traits relevant for (interpersonal) personality dysfunctioning. Manuscript submitted for publication.[8]

With D established as the common core of all aversive traits, it can now serve to guide research questions concerning aversive traits from both fields. As an example and as mentioned before, there is an ongoing debate on whether the separation between clinical aversive traits and dark traits (as supposedly "subclinical" traits) can be justified. In short, some researchers

[8] In the respective manuscript, we used the label *aversive subclinical traits* to refer to dark traits, so as to again avoid using the term "dark" but also to contrast the two research traditions (aversive subclinical research with dark traits, clinical research with aversive AMPD traits). Hence, the labels *dark traits*, *aversive traits from mainstream personality research*, and *aversive subclinical traits* are used interchangeably in the current dissertation.

argue that clinical traits are associated with impairment in personality functioning (Morey et al., 2020; Sharp, 2022), whereas the same does not apply to dark traits, or at least to a much lesser degree, as they ought to operate on a subclinical level (Gunderson & Zanarini, 2011; Paulhus, 2023; Paulhus & Klaiber, 2020). However, studies investigating dark traits with any form of psychopathology are scarce (Thomaes et al., 2017), and importantly, these studies typically do not specifically address Personality Dysfunctioning (Criterion A in the AMPD), which ought to be the key criterion for distinguishing personality disorders from other forms of psychopathology (Morey et al., 2020; Sharp & Wall, 2021; Wright et al., 2022). Given these limitations, the separation of aversive traits in clinical vs. subclinical traits is arguably somewhat premature.

Ultimately, the question of whether dark traits are (at least somewhat) related to personality psychopathology *because* they are aversive traits can be answered by investigating whether D, and thus the variance shared by all aversive traits, is associated with Personality Dysfunctioning. Typically, Personality Dysfunctioning can be distinguished into two broad (correlated) factors, i.e., dysfunctioning related to the self (e.g., the lack of a stable identity or self-esteem) and dysfunctioning related to interpersonal matters (e.g., a lack of empathy or positive interactions with others) (American Psychiatric Association, 2013; Weekers et al., 2019). Given that aversive traits primarily indicate problems with interpersonal matters, we concentrated on *Interpersonal Personality Dysfunctioning* (IPD) rather than dysfunction related to the self. Importantly, even if D, and thus any of its manifestations, is related to IPD, this relationship could of course still be more pronounced for clinical manifestations of D as compared with the (supposedly) subclinical manifestations (i.e., dark traits). Indeed, such a finding would be fully in line with D-framework conceptually, as clinical manifestations of D could be flavored by (general) psychopathology; thus, a flavor that would be unique to them or at least more pronounced than for other manifestations.

To this end, we investigated how 20 aversive traits (the same as in Project 2) and their common core (i.e., D) were associated with IPD over the course of five studies. In Study 1 (N = 604; another sample from the PPP) and Study 2 (N = 2,006; the same sample as in Project 2), we used SEM to investigate whether and which aversive traits were notably related to IPD beyond their common core. Results revealed that this was the case for only three (AMPD traits) out of the 20 aversive traits (i.e., Hostility, Irresponsibility, and Suspiciousness). Crucially, on average, aversive AMPD traits and dark traits were similarly correlated with IPD (median difference between aversive AMPD vs. dark traits: $|\Delta r|$ = .01 - .11) and showed near identical test information curves. Both these results contradicted the view of dark traits as less maladaptive and thus the distinction between clinical and "subclinical" (aversive) traits.

To further identify the unique flavor of the three traits relevant for IPD beyond D, we conducted three further studies, i.e., Study 3a for Hostility (N = 710), Study 3b for Irresponsibility (N = 639), and Study 3c for Suspiciousness (N = 888). As predicted, we could identify the flavor of Hostility as a *lack of Emotion Regulation*, the flavor of Irresponsibility as a *lack of Self-Control*, and the flavor of Suspiciousness as *Delusional Paranoid Thoughts*.

In conclusion, the separation of aversive clinical and subclinical traits is not justified. All aversive traits are related to IPD, mainly because they are manifestations of D. Only some traits entailed flavors (beyond D) that indicated that they were related to IPD beyond their common core, but this was the case only for three of the eight aversive AMPD traits. In other words, the relation of 20 aversive traits (independent of their research tradition) with IPD could be parsimoniously described by four aspects, i.e., utility maximization at the expense of others accompanied by justifying beliefs (i.e., D), (lack of) Emotion Regulation, (lack of) Self-Control, and Delusional Paranoid Thoughts.

3. Discussion

3.1. Summary

Despite their common roots, research on aversive tendencies in clinical and mainstream personality research (the latter often referred to as dark trait research) has developed in almost completely parallel research programs. Previous attempts were limited by theoretical issues (e.g., no personality dimension was explicitly conceptualized as the core of all aversive traits, let alone within a suitable framework) and research designs (e.g., the limited number of aversive traits investigated). To overcome these limitations, I proposed the D-framework and its conceptualization of a common core of all aversive traits as a unifying framework *across* research areas. In this, the D-framework offers testable assumptions about and a comprehensive definition of the common core of all aversive traits (i.e., D), as utility maximization at the expense of others accompanied by justifying beliefs.

Specifically, in three projects investigating up to 20 aversive traits at once, we showed that most antagonistic traits are related to FFM-Agreeableness but are certainly not just a reflection of its low pole, a view upheld by clinical research and hindering further integration with mainstream personality research (Project 1). Instead, *any* aversive trait can be seen as a flavored manifestation of D, and thus the D-framework provides a unifying framework for aversive traits across research traditions (Project 2). Thus, the relationships between aversive traits and IPD, a main criterion for distinguishing between clinical and subclinical traits, is, by and large, due to D, indicating that there are no meaningful differences between dark and aversive AMPD traits on the group level (Project 3). This finding calls into question the separation of research traditions upheld by some and, crucially, also exemplifies the potential of D as a unifying approach for addressing overarching research questions concerning aversive traits. This conclusion has important implications for dark and aversive clinical psychopathology research alike, as well as for their further unification.

3.2. Implications

3.2.1 Implications for aversive personality research (within and across research traditions)

In clinical psychopathology research, debates about the comprehensiveness of the personality space covered by the AMPD are still ongoing (L. A. Clark & Watson, 2022; Sharp & Miller, 2022). On the basis of the findings presented herein, different conclusions about aversive personality can be drawn. On the one hand, given that dark and aversive AMPD traits share a common core and are similarly related to IPD, the findings of the current dissertation give confidence to the idea that dark traits may be valuable additions for a possible revision of the AMPD (see L. A. Clark & Watson, 2022; Thomaes et al., 2017). In this, covering many different instances of aversive personality traits will enable clinicians in practice to describe the personality of a specific individual more precisely. On the other hand, the current findings also indicate that aversive traits are highly redundant in accounting for the main feature of personality disorders (i.e., Criterion A). From a parsimonious viewpoint, one may argue that the AMPD should include largely distinct traits that can account for unique variance especially in Criterion A. As is often the case, there is no right or wrong answer to the different affordances of theoretical work and clinical practice.

No matter how these findings are integrated into clinical research, dark trait research indeed offers valuable insights for aversive clinical psychopathology research (and vice versa), as has been argued by some before (Miller & Campbell, 2008; Thomaes et al., 2017; Zeigler-Hill & Marcus, 2016b). Hereby, previous findings and future research efforts from the two areas could be integrated via the D-framework. Specifically, this framework offers a way to disentangle what is shared by all aversive traits (i.e., D) from their unique aspects, thus offering a clearer understanding of antecedents, development, and outcomes of *aversive* features and how to distinguish them from other features (e.g., Disinhibition).

In this, the unification of dark and aversive clinical psychopathology research may profit if certain labels are no longer used. For one, using the term “subclinical” for dark traits,

e.g., "Subclinical Sadism" (e.g., Jones & Paulhus, 2023) is misleading in my view. Of course, it is vital to develop different measures that are sensitive to different levels of (mal)adaptivity. However, this sensitivity is a property of the measure and not the construct itself. Moreover, the findings presented herein indicate that current popular measures of (supposedly subclinical) dark and (clinical) AMPD traits have similar relationships to IPD and differentiate at the same level of (mal)adaptivity. Thus, I suggest that researchers come to a shared understanding and definitions of aversive traits, just as with the aversive core, and specifically develop measures of these traits that differentiate different levels of (mal)adaptivity.

Another issue regarding labels is the term "dark." This term has been criticized as stigmatizing, vague, and problematic for other reasons in the context of clinical psychopathology research (Miller et al., 2023; Thomaes et al., 2017). Many researchers agree that dark traits essentially represent *socially and/or ethically aversive tendencies* (C. M. Kowalski et al., 2021; Moshagen et al., 2018; Thomaes et al., 2017), which could be a more suitable term for both research areas. Therefore, to enhance unification, researchers interested in aversive personality traits might adopt this term. For the sake of clarity and consistency, I have used the term "dark" throughout this dissertation but have provided some suggestions for other labels in the manuscripts. Independent of these labels, it is important to note that individuals high in aversive tendencies should not be stigmatized as "mentally ill" or "evil." Importantly, maladaptively is not reflected (only) in extreme traits levels but also in the appropriateness of a behavior given a certain context (Wright & Hopwood, 2022). Thus, in harsh and antagonistic environments, (e.g., corrupt political systems), the (justifying) beliefs underlying D might be correct (e.g., that others cannot be trusted); hence, engaging in more aversive behavior, which will be reflected in higher levels of D, might be necessary and adaptive (to a certain degree).

3.2.2. Practical Implications

The justifying beliefs aspect of D may also provide some interesting implications for clinical practitioners. Of course, the current dissertation was concerned with basic – as opposed to applied – research, and as such, cannot directly provide practical implications. Nevertheless, some hypotheses can be generated from this work, which could be investigated in applied settings. For example, in many forms of psychotherapy, some forms of life lessons, scripts, schemas worldviews, expectancies, or core beliefs are assumed to (implicitly) guide behavior (see Kriz, 2017, 2023; Rief et al., 2015). Most prominently, *Cognitive Behavioral Therapy* (CBT) aims to change such cognitive structures through *cognitive restructuring* (see Beck & Dozois, 2011; D. A. Clark, 2013), which is then supposed to lead to subsequent changes in behavior. Hereby, D has some interesting aspects that could be considered in clinical settings. As per the very definition of D, aversive behavior is proposed to be justified by beliefs that are designed to uphold a positive, moral self-view (Hilbig et al., 2022; Moshagen et al., 2018). If such beliefs related to D (for an overview and taxonomy, see Hilbig et al., 2022) are challenged in a therapeutic context, the justification for aversive behavior could be taken away, making the aversive behavior less likely to occur. Moreover, D is also strongly associated with *Moral Disengagement*, the tendency to construct (any) belief that may justify a behavior (Bandura, 1999, 2016). Thus, one could also focus on this tendency, e.g., by searching for reasons for why someone has an elevated tendency to construct beliefs to justify their own aversive behavior rather than questioning the behavior.

In the end, these are some ideas that were derived from the present work and need further empirical investigation. Furthermore, as clinical practice operates with categorial diagnoses, it will also be useful to determine which diagnoses are related to D. Along these lines, the study by Hilbig et al. (2020) already indicated that D longitudinally predicts instances of aversive personality psychopathology (i.e., narcissistic, antisocial, paranoid, and borderline tendencies). It will be beneficial to expand these findings to personality disorders and other mental disorders.

3.3. General Limitations

Some limitations to the findings presented herein should be noted. First, the latent models used herein were ultimately based on mere covariation. However, traits can covary for a multitude of reasons (e.g., common covariates), as has been argued for the association of Risk Taking and D (Tiwari et al., 2021). Although Risk Taking and D do not overlap on a substantive level, they are correlated. Thereby, Grandiosity has been suggested as the common denominator that is responsible for their covariation (Tiwari et al., 2021). Thus, even though the aversive clinical traits investigated herein showed both substantial and empirical overlap with D, a more critical test of this covariation will support an understanding of the underlying processes and the development of aversive personality.

Second, as is common in the clinical field, we relied only on the AMPD as a framework for clinical personality traits and only on one inventory to measure its traits, i.e., the *Personality Inventory for DSM-5* (PID-5; Krueger et al., 2012) across all studies presented herein, therefore raising some concerns about the generalizability of the findings. We chose the PID-5 as it is the most widely accepted operationalization of the AMPD (Krueger & Hobbs, 2020; Zimmermann et al., 2019), some even go so far as to equate research on the AMPD with the PID-5 (L. A. Clark & Watson, 2022). Thus, we chose this measure to enhance the comparability of our research with the vast majority of clinical personality psychopathology research. Along these lines, the AMPD is currently the guiding research model in clinical personality psychopathology research (Krueger & Hobbs, 2020; Zimmermann et al., 2019). Although there are other clinical personality models in addition to the AMPD, most prominently the ICD-11 (World Health Organization, 2023), validated measures for such other models are still in development or are less researched (see L. A. Clark et al., 2021).

Third, the studies presented herein relied primarily on self-reports, thus raising concerns about biased estimates due to common method variance, e.g., response styles (Ashton et al.,

2017; Plieninger, 2017). This concern can be at least somewhat alleviated, as we also used an incentivized economic game (i.e., the SVO task) in Project 2. In this, it should be noted that other measures of aversive traits or psychopathology, e.g., informant reports or interviews (Ro et al., 2017), did not seem feasible for achieving the necessary statistical power, as huge sample sizes would have been needed to test the present hypotheses and run the statistical tests. Nevertheless, future work should consider using other forms of assessment for these constructs. Last, we relied on German-speaking samples across studies, therefore limiting the generalizability of the results mostly to Western, educated, industrialized, rich, and democratic (WEIRD; see Henrich et al., 2010) populations. In sum, replicating the findings in more diverse samples, with other models of psychopathology beyond the AMPD, and methods other than self-reports will lend confidence to the conclusions drawn herein.

3.4. Conclusion

The D-framework serves as a unifying framework that accounts for the variance shared across all aversive traits, independent of their research traditions, and allows testable predictions to be made about their shared aversive core (i.e., D). As such, it offers the potential to unite research agendas on aversive personality across research traditions, alleviating the burdens of parallel and largely unintegrated areas of research on this topic.

References

Allport, G. W. (1955). *Becoming: Basic considerations for a psychology of personality. Terry lectures*. Yale University Press.

American Psychiatric Association. (1994). *DSM-IV: Diagnostic and Statistical Manual of Mental Disorders*. Author.

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

Anderson, J. L., Sellbom, M., Wygant, D. B., Salekin, R. T., & Krueger, R. F. (2014). Examining the associations between DSM-5 section III antisocial personality disorder traits and psychopathy in community and university samples. *Journal of Personality Disorders*, *28*(5), 675–697. https://doi.org/10.1521/pedi_2014_28_134

Ashton, M. C. (2013). *Individual differences and personality* (Second edition). Academic Press.

Ashton, M. C., & Lee, K. (2005a). A defence of the lexical approach to the study of personality structure. *European Journal of Personality*, *19*(1), 5–24. https://doi.org/10.1002/per.541

Ashton, M. C., & Lee, K. (2005b). Honesty-humility, the big five, and the five-factor model. *Journal of Personality*, *73*(5), 1321–1353. https://doi.org/10.1111/j.1467-6494.2005.00351.x

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc*, *11*(2), 150–166. https://doi.org/10.1177/1088868306294907

Ashton, M. C., & Lee, K. (2020). Objections to the HEXACO Model of Personality Structure—And Why Those Objections Fail. *European Journal of Personality*, *34*(4), 492–510. https://doi.org/10.1002/per.2242

Ashton, M. C., Lee, K., Perugini, M., Szarota, P., Vries, R. E. de, Di Blas, L., Boies, K., & Raad, B. de (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

Ashton, M. C., Vries, R. E. de, & Lee, K. (2017). Trait Variance and Response Style Variance in the Scales of the Personality Inventory for DSM-5 (PID-5). *Journal of Personality Assessment*, *99*(2), 192–203. https://doi.org/10.1080/00223891.2016.1208210

Bader, M., Hartung, J., Hilbig, B. E., Zettler, I., Moshagen, M., & Wilhelm, O. (2021). Themes of the dark core of personality. *Psychological Assessment*.

Bader, M., Hilbig, B. E., Zettler, I., & Moshagen, M. (2023). Rethinking Aversive Personality: Decomposing the Dark Triad Traits into their Common Core and Unique Flavors. *Journal of Personality.* Advance online publication. https://doi.org/10.1111/jopy.12785

Bader, M., Horsten, L. K., Hilbig, B. E., Zettler, I., & Moshagen, M. (2021). Measuring the Dark Core of Personality in German: Psychometric Properties, Measurement Invariance, Predictive Validity, and Self-Other Agreement. *Journal of Personality Assessment*, 1–14. https://doi.org/10.1080/00223891.2021.1984931

Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review : An Official Journal of the Society for Personality and Social Psychology, Inc*, *3*(3), 193–209. https://doi.org/10.1207/s15327957pspr0303_3

Bandura, A. (2002). Selective Moral Disengagement in the Exercise of Moral Agency. *Journal of Moral Education*, *31*(2), 101–119. https://doi.org/10.1080/0305724022014322

Bandura, A. (2016). *Moral Disengagement: How People Do Harm and Live with Themselves, by Albert Bandura. New York: Macmillan, 2016. 544 pp. ISBN: 978-1*. https://www.cambridge.org/core/services/aop-cambridge-core/content/view/s1052150x16000373 evelopments in Criminological Theory

Barkan, R., Ayal, S., & Ariely, D. (2015). Ethical dissonance, justifications, and moral behavior. *Current Opinion in Psychology*, *6*, 157–161. https://doi.org/10.1016/j.copsyc.2015.08.001

Beck, A. T., & Dozois, D. J. A. (2011). Cognitive therapy: Current status and future directions. *Annual Review of Medicine*, *62*, 397–409. https://doi.org/10.1146/annurev-med-052209-100032

Buckels, E. E., Jones, D. N [Daniel Nelson], & Paulhus, D. L. (2013). Behavioral confirmation of everyday sadism. *Psychological Science*, *24*(11), 2201–2209. https://doi.org/10.1177/0956797613490749

Carroll, J. B. (2009). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.

Casper, W. J., Vaziri, H., Wayne, J. H., DeHauw, S., & Greenhaus, J. (2018). The jingle-jangle of work-nonwork balance: A comprehensive and meta-analytic review of its meaning and measurement. *The Journal of Applied Psychology*, *103*(2), 182–214. https://doi.org/10.1037/apl0000259

Christie, R., & Geis, F. L. (1970). *Studies in Machiavellianism*. Academic Press.

Clark, D. A. (2013). Cognitive Restructuring. In S. G. Hofmann (Ed.), *The Wiley Handbook of Cognitive Behavioral Therapy* (pp. 1–22). Wiley. https://doi.org/10.1002/9781118528563.wbcbt02

Clark, L. A. (2007). Assessment and diagnosis of personality disorder: Perennial issues and an emerging reconceptualization. *Annual Review of Psychology*, *58*, 227–257. https://doi.org/10.1146/annurev.psych.57.102904.190200

Clark, L. A., Corona-Espinosa, A., Khoo, S., Kotelnikova, Y., Levin-Aspenson, H. F., Serapio-García, G., & Watson, D. (2021). Preliminary Scales for ICD-11 Personality Disorder: Self and Interpersonal Dysfunction Plus Five Personality Disorder Trait Domains. *Frontiers in Psychology*, *12*, 668724. https://doi.org/10.3389/fpsyg.2021.668724

Clark, L. A., & Watson, D. (2022). The trait model of the DSM-5 alternative model of personality disorder (AMPD): A structural review. *Personality Disorders: Theory, Research, and Treatment*, *13*(4), 328–336. https://doi.org/10.1037/per0000568

Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences*. Elsevier Science.

Credé, M., Tynan, M. C., & Harms, P. D. (2017). Much ado about grit: A meta-analytic synthesis of the grit literature. *Journal of Personality and Social Psychology*, *113*(3), 492–511. https://doi.org/10.1037/pspp0000102

Crowe, M. L., Lynam, D. R., & Miller, J. D. (2018). Uncovering the structure of agreeableness from self-report measures. *Journal of Personality*, *86*(5), 771–787. https://doi.org/10.1111/jopy.12358

Furnham, A., Richards, S. C., & Paulhus, D. L. (2013). The Dark Triad of Personality: A 10 Year Review. *Social and Personality Psychology Compass*, *7*(3), 199–216. https://doi.org/10.1111/spc3.12018

Furnham, A., Richards, S., Rangel, L., & Jones, D. N [Daniel N.] (2014). Measuring malevolence: Quantitative issues surrounding the Dark Triad of personality. *Personality and Individual Differences*, *67*, 114–121. https://doi.org/10.1016/j.paid.2014.02.001

Graziano, W. G., & Tobin, R. M. (2017). Agreeableness and the five factor model. In T. A. Widiger (Ed.), *Oxford handbook of the five factor model* (pp. 105–132). Oxford University Press.

Gunderson, J. G., & Zanarini, M. C. (2011). Deceptively simple – or radical shift? *Personality and Mental Health*, *5*(4), 260–262. https://doi.org/10.1002/pmh.181

Hagger, M. S. (2014). Avoiding the "déjà-variable" phenomenon: Social psychology needs more guides to constructs. *Frontiers in Psychology*, *5*, 52. https://doi.org/10.3389/fpsyg.2014.00052

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, *33*(2-3), 61-83; discussion 83-135. https://doi.org/10.1017/S0140525X0999152X

Hilbig, B. E., Moshagen, M., Horsten, L. K., & Zettler, I. (2021). Agreeableness is dead. Long live Agreeableness? Reply to Vize and Lynam. *Journal of Research in Personality*, *91*, 104074. https://doi.org/10.1016/j.jrp.2021.104074

Hilbig, B. E., Moshagen, M., Thielmann, I., & Zettler, I. (2022). Making rights from wrongs: The crucial role of beliefs and justifications for the expression of aversive personality. *Journal of Experimental Psychology: General.* Advance online publication. https://doi.org/10.1037/xge0001232

Hilbig, B. E., Moshagen, M., & Zettler, I. (2016). Prediction Consistency: A Test of the Equivalence Assumption across Different Indicators of the Same Construct. *European Journal of Personality*, *30*(6), 637–647. https://doi.org/10.1002/per.2085

Hilbig, B. E., Thielmann, I., Klein, S. A., Moshagen, M., & Zettler, I. (2020). The dark core of personality and socially aversive psychopathology. *Journal of Personality.* Advance online publication. https://doi.org/10.1111/jopy.12577

Hilbig, B. E., Thielmann, I., Zettler, I., & Moshagen, M. (2023). The Dispositional Essence of Proactive Social Preferences: The Dark Core of Personality vis-à-vis 58 Traits. *Psychological Science*, *34*(2), 201–220. https://doi.org/10.1177/09567976221116893

Horsten, L. K. (2023). *Manoeuvering through the jangle jungle. What is the common core of aversive personality traits (not)?* [Doctoral dissertation, RPTU Kaiserslautern-Landau]. RIS. https://kluedo.ub.rptu.de/files/7300/dissertation_horsten_2023.pdf

Horsten, L. K., Moshagen, M., Thielmann, I., Zettler, I., & Hilbig, B. E. (2023). How many (and which) aversive traits does it take to approximate the common core of aversive personality? [Manuscript submitted for publication].

Horsten, L. K., Moshagen, M., Zettler, I., & Hilbig, B. E. (2021). Theoretical and empirical dissociations between the Dark Factor of Personality and low Honesty-Humility. *Journal of Research in Personality*, *95*, 104154. https://doi.org/10.1016/j.jrp.2021.104154

Horsten, L. K., Thielmann, I., Moshagen, M., Zettler, I., Scholz, D., & Hilbig, B. E. (2023). Testing the equivalence of the aversive core of personality and a blend of agreeableness(-related) items. *Journal of Personality.* Advance online publication. https://doi.org/10.1111/jopy.12830

Jonason, P. K. (2023). Introduction. In P. K. Jonason (Ed.), *Shining light on the dark side of personality: Measurement properties and theoretical advances* (pp. 3–8). Hogrefe.

Jones, D. N [Daniel Nelson], & Figueredo, A. J. (2013). The Core of Darkness: Uncovering the Heart of the Dark Triad. *European Journal of Personality*, *27*(6), 521–531. https://doi.org/10.1002/per.1893

Jones, D. N [Daniel Nelson], & Paulhus, D. L. (2023). Short Dark Triad and Short Dark Tetrad. In P. K. Jonason (Ed.), *Shining light on the dark side of personality: Measurement properties and theoretical advances* (pp. 215–230). Hogrefe.

Kelley, T. L. (1927). *Interpretation of educational measurements.* World Book.

Kongerslev, M., Simonsen, S., & Bo, S. (2015). The quest for tailored treatments: A meta-discussion of six social cognitive therapies. *Journal of Clinical Psychology*, *71*(2), 188–198. https://doi.org/10.1002/jclp.22154

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K. E., . . . Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454–477. https://doi.org/10.1037/abn0000258

Kowalski, C. M., Rogoza, R., Saklofske, D. H., & Schermer, J. A. (2021). Dark triads, tetrads, tents, and cores: Why navigate (research) the jungle of dark personality models without a compass (criterion)? *Acta Psychologica*, *221*, 103455. https://doi.org/10.1016/j.actpsy.2021.103455

Kowalski, R. M. (Ed.). (2001). *Behaving badly aversive behaviors in interpersonal relationships 1st ed*. American Psychological Association. https://doi.org/10.1037/10365-000

Kriz, J. (2017). *Subjekt und Lebenswelt: Personzentrierte Systemtheorie für Psychotherapie, Beratung und Coaching* (1. Auflage). Vandenhoeck & Ruprecht.

Kriz, J. (2023). *Grundkonzepte der Psychotherapie: Mit Online-Material* (8. Auflage). Julius Beltz GmbH & Co. KG.

Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2012). Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychological Medicine*, *42*(9), 1879–1890. https://doi.org/10.1017/S0033291711002674

Krueger, R. F., & Hobbs, K. A. (2020). An Overview of the DSM-5 Alternative Model of Personality Disorders. *Psychopathology*, *53*(3-4), 126–132. https://doi.org/10.1159/000508538

Krueger, R. F., Hobbs, K. A., Conway, C. C., Dick, D. M., Dretsch, M. N., Eaton, N. R., Forbes, M. K., Forbush, K. T., Keyes, K. M., Latzman, R. D., Michelini, G., Patrick, C. J., Sellbom, M., Slade, T., South, S. C., Sunderland, M., Tackett, J., Waldman, I., Waszczuk, M. A., . . . Kotov, R. (2021). Validity and utility of Hierarchical Taxonomy of Psychopathology (HiTOP): Ii. Externalizing superspectrum. *World Psychiatry : Official Journal of the World Psychiatric Association (WPA)*, *20*(2), 171–193. https://doi.org/10.1002/wps.20844

Lynam, D. R., & Miller, J. D. (2019). The basic trait of Antagonism: An unfortunately underappreciated construct. *Journal of Research in Personality*, *81*, 118–126. https://doi.org/10.1016/j.jrp.2019.05.012

Marcus, D. K., & Zeigler-Hill, V. (2015). A Big Tent of Dark Personality Traits. *Social and Personality Psychology Compass*, *9*(8), 434–446. https://doi.org/10.1111/spc3.12185

Marcus, D. K., Zeigler-Hill, V., Mercer, S. H., & Norris, A. L. (2014). The psychology of spite and the measurement of spitefulness. *Psychological Assessment*, *26*(2), 563–574. https://doi.org/10.1037/a0036039

McCrae, R. R. (1991). The five-factor model and its assessment in clinical settings. *Journal of Personality Assessment*, *57*(3), 399-14. https://doi.org/10.1207/s15327752jpa5703_2

McCrae, R. R., & Costa, P. T [P. T.]. (2008). The five-factor theory of personality. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research.* The Guilford Press.

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

Miller, J. D., & Campbell, W. K. (2008). Comparing clinical and social-personality conceptualizations of narcissism. *Journal of Personality*, *76*(3), 449–476. https://doi.org/10.1111/j.1467-6494.2008.00492.x

Miller, J. D., Gentile, B., Wilson, L., & Campbell, W. K. (2013). Grandiose and vulnerable narcissism and the DSM-5 pathological personality trait model. *Journal of Personality Assessment*, *95*(3), 284–290. https://doi.org/10.1080/00223891.2012.685907

Miller, J. D., Lynam, D. R., & Campbell, W. K. (2016). Measures of Narcissism and Their Relations to DSM-5 Pathological Traits: A Critical Reappraisal. *Assessment*, *23*(1), 3–9. https://doi.org/10.1177/1073191114522909

Miller, J. D., Lynam, D. R., Widiger, T. A., & Leukefeld, C. (2001). Personality disorders as extreme variants of common personality dimensions: Can the Five-Factor Model adequately represent psychopathy? *Journal of Personality*, *69*(2), 253–276. https://doi.org/10.1111/1467-6494.00144

Miller, J. D., Sharpe, B. M., & Lynam, D. R. (2023). Controversies Surrounding the Dark Triad: A Call to Action. In P. K. Jonason (Ed.), *Shining light on the dark side of personality: Measurement properties and theoretical advances* (pp. 285–293). Hogrefe.

Miller, J. D., Vize, C. E., Crowe, M. L., & Lynam, D. R. (2019). A Critical Appraisal of the Dark-Triad Literature and Suggestions for Moving Forward. *Current Directions in Psychological Science*, *28*(4), 353–360. https://doi.org/10.1177/0963721419838233

Morey, L. C., Good, E. W., & Hopwood, C. J. (2020). Global personality dysfunction and the relationship of pathological and normal trait domains in the DSM-5 alternative model for personality disorders. *Journal of Personality.* Advance online publication. https://doi.org/10.1111/jopy.12560

Moshagen, M., Bader, M., Zettler, I., & Hilbig, B. E. (2023). The role of impulse and interference control in aversive personality: A comprehensive assessment. *Acta Psychologica*, *239*, 104018. https://doi.org/10.1016/j.actpsy.2023.104018

Moshagen, M., Hilbig, B. E., & Zettler, I. (in press). How and why aversive personality is expressed in political preferences. *Journal of Personality and Social Psychology*.

Moshagen, M., Hilbig, B. E., & Zettler, I. (2018). The dark core of personality. *Psychological Review*, *125*(5), 656–688. https://doi.org/10.1037/rev0000111

Moshagen, M., Zettler, I., & Hilbig, B. E. (2020). Measuring the Dark Core of Personality. *Psychological Assessment*, *32*(2), 182–196. https://doi.org/10.1037/pas0000778

Moshagen, M., Zettler, I., Horsten, L. K., & Hilbig, B. E. (2020). Agreeableness and the common core of dark traits are functionally different constructs. *Journal of Research in Personality*, *87*, 103986. https://doi.org/10.1016/j.jrp.2020.103986

Muris, P., Merckelbach, H., Otgaar, H., & Meijer, E. (2017). The Malevolent Side of Human Nature. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *12*(2), 183–204. https://doi.org/10.1177/1745691616666070

Murphy, R. O., Ackermann, K. A., & Handgraaf, M. (2011). Measuring Social Value Orientation. *SSRN Electronic Journal.* Advance online publication. https://doi.org/10.2139/ssrn.1804189

O'Connor, B. P. (2011). Social allergens. In L. M. Horowitz & S. Strack (Eds.), *Handbook of interpersonal psychology: Theory, research, assessment, and therapeutic interventions* (pp. 269–280). John Wiley & Sons, Inc.

Paulhus, D. L. (2014). Toward a Taxonomy of Dark Personalities. *Current Directions in Psychological Science*, *23*(6), 421–426. https://doi.org/10.1177/0963721414547737

Paulhus, D. L. (2023). Controversies in Dark Trait Research: A Call for Reason. In P. K. Jonason (Ed.), *Shining light on the dark side of personality: Measurement properties and theoretical advances* (pp. 295–306). Hogrefe.

Paulhus, D. L., & Dutton, D. G. (2016). Everyday sadism. In *The dark side of personality: Science and practice in social, personality, and clinical psychology* (pp. 109–120). American Psychological Association. https://doi.org/10.1037/14854-006

Paulhus, D. L., & Klaiber, P. (2020). HEXACO, Dark Personalities, and Brunswik Symmetry. *European Journal of Personality*, *34*, 541–542.

Paulhus, D. L., & Williams, K. M. (2002). The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, *36*(6), 556–563. https://doi.org/10.1016/S0092-6566(02)00505-6

Pincus, A. L., Ansell, E. B., Pimentel, C. A., Cain, N. M., Wright, A. G. C., & Levy, K. N. (2009). Initial construction and validation of the Pathological Narcissism Inventory. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, *21*(3), 365–379. https://doi.org/10.1037/a0016530

Plieninger, H. (2017). Mountain or Molehill? A Simulation Study on the Impact of Response Styles. *Educational and Psychological Measurement*, *77*(1), 32–53. https://doi.org/10.1177/0013164416636655

Prentice, M., Jayawickreme, E., Hawkins, A., Hartley, A., Furr, R. M., & Fleeson, W. (2019). Morality as a Basic Psychological Need. *Social Psychological and Personality Science*, *10*(4), 449–460. https://doi.org/10.1177/1948550618772011

Raskin, R. N., & Hall, C. S. (1979). A narcissistic personality inventory. *Psychological Reports*, *45*(2), 590. https://doi.org/10.2466/pr0.1979.45.2.590

Ray, J. J., & Ray, J. A. (1982). Some apparent advantages of subclinical psychopathy. *The Journal of Social Psychology*, *117*(First Half), 135–142. https://doi.org/10.1080/00224545.1982.9713415

Rief, W., Glombiewski, J. A., Gollwitzer, M., Schubö, A., Schwarting, R., & Thorwart, A. (2015). Expectancies as core features of mental disorders. *Current Opinion in Psychiatry*, *28*(5), 378–385. https://doi.org/10.1097/YCO.0000000000000184

Ro, E., Nuzum, H., & Clark, L. A. (2017). Antagonism trait facets and comprehensive psychosocial disability: Comparing information across self, informant, and interviewer reports. *Journal of Abnormal Psychology*, *126*(7), 890–897. https://doi.org/10.1037/abn0000298

Rose, L., Lynam, D. R., & Miller, J. D. (in press). Measuring the “Dark” Triad: Comparing the Five-Factor Model Antagonistic Triad Measure to other commonly used self-report instruments. *Assessment*.

Rose, L., Miller, J. D., & Lynam, D. R. (2023). Validation of the Five-Factor Model Antagonistic Triad Measure. *Assessment*, *30*(3), 782–797. https://doi.org/10.1177/10731911211068083

Rose, L., Sleep, C. E., Lynam, D. R., & Miller, J. D. (2022). Welcome to the Jangle: Comparing the Empirical Profiles of the "Dark" Factor and Antagonism. *Assessment*, 10731911221124847. https://doi.org/10.1177/10731911221124847

Rudloff, J. P., Hutmacher, F., & Appel, M. (2022). Beliefs about the nature of knowledge shape responses to the pandemic: Epistemic beliefs, the Dark Factor of Personality, and COVID-19-related conspiracy ideation and behavior. *Journal of Personality.* Advance online publication. https://doi.org/10.1111/jopy.12706

Rudloff, J. P., Hutmacher, F., & Appel, M. (2023). Post-truth epistemic beliefs rooted in the Dark Factor of Personality are associated with higher COVID-19 vaccination refusal. *Scientific Reports*, *13*(1), 4254. https://doi.org/10.1038/s41598-023-31079-9

Russell, T. D., Doan, C. M., & King, A. R. (2017). Sexually violent women: The PID-5, everyday sadism, and adversarial sexual attitudes predict female sexual aggression and coercion against male victims. *Personality and Individual Differences*, *111*, 242–249. https://doi.org/10.1016/j.paid.2017.02.019

Russell, T. D., & King, A. R. (2017). Mean girls: PID-5 personality traits and everyday sadism predict hostile femininity. *Personality and Individual Differences*, *104*, 252–257. https://doi.org/10.1016/j.paid.2016.08.020

Samuel, D. B., & Widiger, T. A. (2008). A meta-analytic review of the relationships between the five-factor model and DSM-IV-TR personality disorders: A facet level analysis. *Clinical Psychology Review*, *28*(8), 1326–1342. https://doi.org/10.1016/j.cpr.2008.07.002

Saucier, G. (2002). Orthogonal Markers for Orthogonal Factors: The Case of the Big Five. *Journal of Research in Personality*, *36*(1), 1–31. https://doi.org/10.1006/jrpe.2001.2335

Saulsman, L. M., & Page, A. C. (2004). The five-factor model and personality disorder empirical literature: A meta-analytic review. *Clinical Psychology Review*, *23*(8), 1055–1085. https://doi.org/10.1016/j.cpr.2002.09.001

Scholz, D. D., Hilbig, B. E., Thielmann, I., Moshagen, M., & Zettler, I. (2022). Beyond (low) Agreeableness: Towards a more comprehensive understanding of antagonistic psychopathology. *Journal of Personality.* Advance online publication. https://doi.org/10.1111/jopy.12708

Scholz, D. D., Kraus, J., & Miller, L. (in press). Measuring the Propensity to Trust in Automated Technology: Examining Similarities to Dispositional Trust in Other Humans and Validation of the PTT-A Scale. *International Journal of Human–Computer Interaction*.

Scholz, D. D., Thielmann, I., & Hilbig, B. E. (2023). Down to the core: The role of the common core of dark traits for aversive relationship behaviors. *Personality and Individual Differences*, *213*, 112263. https://doi.org/10.1016/j.paid.2023.112263

Scholz, D. D., Zimmermann, J., Moshagen, M., Zettler, I., & Hilbig, B. E. (in press). Theoretical and empirical integration of ‘dark’ traits and socially aversive personality psychopathology. *Journal of Personality Disorders*.

Schreiber, A., & Marcus, B. (2020). The place of the "Dark Triad" in general models of personality: Some meta-analytic clarification. *Psychological Bulletin*, *146*(11), 1021–1041. https://doi.org/10.1037/bul0000299

Schrödter, R., Laborde, S., & Allen, M. S. (2021). The dark core of personality and sexism in sport. *Personality and Individual Differences*, *183*, 111119. https://doi.org/10.1016/j.paid.2021.111119

Segrin, C., & Dillard, J. P. (1992). The Interactional Theory of Depression: A Meta-Analysis of the Research Literature. *Journal of Social and Clinical Psychology*, *11*(1), 43–70. https://doi.org/10.1521/jscp.1992.11.1.43

Shalvi, S., Gino, F., Barkan, R., & Ayal, S. (2015). Self-Serving Justifications. *Current Directions in Psychological Science*, *24*(2), 125–130. https://doi.org/10.1177/0963721414553264

Sharp, C. (2022). Fulfilling the promise of the LPF: Comment on Morey et al. (2022). *Personality Disorders: Theory, Research, and Treatment*, *13*(4), 316–320. https://doi.org/10.1037/per0000567

Sharp, C., & Miller, J. D. (2022). Ten-year retrospective on the DSM-5 alternative model of personality disorder: Seeing the forest for the trees. *Personality Disorders: Theory, Research, and Treatment*, *13*(4), 301–304. https://doi.org/10.1037/per0000595

Sharp, C., & Wall, K. (2021). Dsm-5 Level of Personality Functioning: Refocusing Personality Disorder on What It Means to Be Human. *Annual Review of Clinical Psychology*, *17*, 313–337. https://doi.org/10.1146/annurev-clinpsy-081219-105402

Sleep, C. E., Crowe, M. L., Carter, N. T., Lynam, D. R., & Miller, J. D. (2020). Uncovering the structure of antagonism. *Personality Disorders: Theory, Research, and Treatment.* Advance online publication. https://doi.org/10.1037/per0000416

Somma, A., Borroni, S., Sellbom, M., Markon, K. E., Krueger, R. F., & Fossati, A. (2020). Assessing dark triad dimensions from the perspective of moral disengagement and DSM-5 alternative model of personality disorder traits. *Personality Disorders: Theory, Research, and Treatment*, *11*(2), 100–107. https://doi.org/10.1037/per0000388

Spearman, C. (1927). The Measurement of Intelligence. *Nature*, *120*(3025), 577–578. https://doi.org/10.1038/120577a0

Tappin, B. M., & McKay, R. T. (2017). The Illusion of Moral Superiority. *Social Psychological and Personality Science*, *8*(6), 623–631. https://doi.org/10.1177/1948550616673878

Thielmann, I., & Hilbig, B. E. (2019). Nomological consistency: A comprehensive test of the equivalence of different trait indicators for the same constructs. *Journal of Personality*, *87*(3), 715–730. https://doi.org/10.1111/jopy.12428

Thielmann, I., & Hilbig, B. E. (2023). Generalized Dispositional Distrust as the Common Core of Populism and Conspiracy Mentality. *Political Psychology*, Article pops.12886. Advance online publication. https://doi.org/10.1111/pops.12886

Thielmann, I., Moshagen, M., Hilbig, B. E., & Zettler, I. (2021). On the Comparability of Basic Personality Models: Meta-Analytic Correspondence, Scope, and Orthogonality of the Big Five and HEXACO Dimensions. *European Journal of Personality*, 089020702110267. https://doi.org/10.1177/08902070211026793

Thomaes, S., Brummelman, E., Miller, J. D., & Lilienfeld, S. O. (2017). The dark personality and psychopathology: Toward a brighter future. *Journal of Abnormal Psychology*, *126*(7), 835–842. https://doi.org/10.1037/abn0000305

Tiwari, S., Moshagen, M., Hilbig, B. E., & Zettler, I. (2021). The Dark Factor of Personality and Risk-Taking. *International Journal of Environmental Research and Public Health*, *18*(16). https://doi.org/10.3390/ijerph18168400

Trull, T. J., & Durrett, C. A. (2005). Categorical and dimensional models of personality disorder. *Annual Review of Clinical Psychology*, *1*, 355–380. https://doi.org/10.1146/annurev.clinpsy.1.102803.144009

Tyrer, P., Reed, G. M., & Crawford, M. J. (2015). Classification, assessment, prevalence, and effect of personality disorder. *The Lancet*, *385*(9969), 717–726. https://doi.org/10.1016/S0140-6736(14)61995-4

Vize, C. E., Collison, K. L., & Lynam, D. R. (2020). The Importance of Antagonism: Explaining Similarities and Differences in Psychopathy and Narcissism's Relations With Aggression and Externalizing Outcomes. *Journal of Personality Disorders*, *34*(6), 842–854. https://doi.org/10.1521/pedi_2020_34_342

Vize, C. E., Collison, K. L., Miller, J. D., & Lynam, D. R. (2020). The "core" of the dark triad: A test of competing hypotheses. *Personality Disorders: Theory, Research, and Treatment*, *11*(2), 91–99. https://doi.org/10.1037/per0000386

Vize, C. E., & Lynam, D. R. (2021). On the importance of the assessment and conceptualization of Agreeableness: A commentary on "Agreeableness and the common core of dark traits are functionally different constructs". *Journal of Research in Personality*, *90.* https://doi.org/10.1016/j.jrp.2020.104059

Vize, C. E., Lynam, D. R., Collison, K. L., & Miller, J. D. (2018). Differences among dark triad components: A meta-analytic investigation. *Personality Disorders: Theory, Research, and Treatment*, *9*(2), 101–111. https://doi.org/10.1037/per0000222

Weekers, L. C., Hutsebaut, J., & Kamphuis, J. H. (2019). The Level of Personality Functioning Scale-Brief Form 2.0: Update of a brief instrument for assessing level of personality functioning. *Personality and Mental Health*, 3–14. https://doi.org/10.1002/pmh.1434

Widiger, T. A., & Samuel, D. B. (2005). Diagnostic categories or dimensions? A question for the Diagnostic And Statistical Manual Of Mental Disorders--fifth edition. *Journal of Abnormal Psychology*, *114*(4), 494–504. https://doi.org/10.1037/0021-843X.114.4.494

Widiger, T. A., & Simonsen, E. M. D. (2005). Alternative Dimensional Models of Personality Disorder: Finding a Common Ground. *Journal of Personality Disorders*, *19*(2), 110–130. https://doi.org/10.1521/pedi.19.2.110.62628

Widiger, T. A., Trull, T. J., Clarkin, J. F., Sanderson, C., & Costa, P. T [Paul T.]. (2002). A description of the DSM-IV personality disorders with 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. 89–99). American Psychological Association. https://doi.org/10.1037/10423-006

Wiggins, J. S., & Pincus, A. L. (1989). Conceptions of personality disorders and dimensions of personality. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, *1*(4), 305–316. https://doi.org/10.1037/1040-3590.1.4.305

World Health Organization. (1992). *International Classification of Diseases, 10th Revision.* World Health Organization.

World Health Organization. (2023). *International statistical classification of diseases and related health problems*. (11th ed.). https://icd.who.int/

Wright, A. G. C., & Hopwood, C. J. (2022). Integrating and distinguishing personality and psychopathology. *Journal of Personality*, *90*(1), 5–19. https://doi.org/10.1111/jopy.12671

Wright, A. G. C., Ringwald, W. R., Hopwood, C. J., & Pincus, A. L. (2022). *It's Time to Replace the Personality Disorders with the Interpersonal Disorders*. Center for Open Science. https://doi.org/10.31234/osf.io/7syvf

Zeigler-Hill, V., & Marcus, D. K. (Eds.). (2016a). *Dark side of personality - science and practice in social, personality, and clinical psychology*. American Psychological Association.

Zeigler-Hill, V., & Marcus, D. K. (2016b). Introduction: A bright future for dark personality features? In V. Zeigler-Hill & D. K. Marcus (Eds.), *Dark side of personality - science and practice in social, personality, and clinical psychology* (pp. 3–22). American Psychological Association. https://doi.org/10.1037/14854-001

Zimmermann, J., Kerber, A., Rek, K., Hopwood, C. J., & Krueger, R. F. (2019). A Brief but Comprehensive Review of Research on the Alternative DSM-5 Model for Personality Disorders. *Current Psychiatry Reports*, *21*(9), 21–92. https://doi.org/10.1007/s11920-019-1079-z

Danksagung

Wie für viele Promovenden war die Zeit als Doktorand für mich nicht selten eine wahre Achterbahnfahrt (der Gefühle). Wie bei jeder guten Achterbahn, gab es Momente der Euphorie, aber auch Momente des Zögerns und Hinterfragen, warum man diese Fahrt eigentlich angetreten ist. Auf der einen Seite hat man die emotionalen Höhepunkte, so z.B., wenn man eine Theorie oder ein Konstrukt (glaubt) endlich verstanden zu haben oder wenn ein Paper zur Publikation akzeptiert wird, auf der andere Seite die Tiefpunkte, z.B. wenn man merkt, dass man einen Fehler in der Datenanalyse hatte oder wenig konstruktives Feedback im Zuge eines Review Prozesses bekommt. Auch wenn mir manchmal nach der ein oder anderen Fahrt ein bisschen schwindelig war, freue ich mich, dass ich verschiedenste Fahrten ausprobieren konnte. Ich bin zudem dankbar, dass ich diese Fahrten auch noch in guter Gesellschaft angetreten bin.

Als erstes gilt mein Dank hierbei meinen Betreuer, Benjamin Hilbig, der mit ähnlicher Abenteuerlust immer ganz vorne im ersten Wagen der Achterbahn mit dabei war. Tatsächlich hätte ich mir dabei keinen besseren Mitfahrer (bzw. Betreuer und Chef) wünschen können. Benni, für dein immer ehrliches und konstruktives Feedback, schnelle Erreichbarkeit, die Freiheiten in meiner Arbeitsweise, deine pragmatischen Lösungsvorschläge und das Wissen, dass das Wohl deiner Mitarbeiter und Doktoranden immer an erste Stelle kommt bin ich dir äußerst dankbar und bist mir in diesen Dingen auf jeden Fall ein großes Vorbild geworden. Danke, dass du mich auch auf die etwas unsicher wirkenden, aber doch irgendwo spaßigen Fahrgeschäften hingewiesen, dann aber auch belgeitet und unterstützt hast!

Als nächstes möchte ich bei meinen weiteren Ko-Autoren bedanken (Morten Moshagen, Isabel Thielmann, Ingo Zettler und Johannes Zimmermann), die bei einigen der Fahrten direkt im Wagen hinter uns mit dabei waren. Morten, Isa, Ingo und Johannes, ich danke euch sehr für den Ideenaustausch, die konstruktiven Rückmeldungen und das auch ihr euch von den vielleicht manchmal eher ungewöhnlichen Fahrten nicht habt abschrecken lassen!

Ein ebenso treuer Begleiter war mir Johannes Kraus, der mich überhaupt erst ermutigt und, zu vielen Teilen, befähigt hat, es überhaupt zu wagen, diese Achterbahnfahrt anzutreten. Johannes, ich bin sehr glücklich und dankbar, dass ich immer auf deine Unterstützung zählen kann und dass wir auch hier wieder die ein oder andere Fahrt zusammen angetreten sind.

Auch möchte ich meinen Kollegen der Abteilung Experimentelle Psychologie und Persönlichkeitsforschung, sowie meinen SMiPster Kollegen meinen Dank aussprechen. Auch wenn wir unterschiedliche Achterbahnen ausprobiert haben, tat es sehr gut, sich über die doch ähnlichen Erfahrungen auszutauschen! Insbesondere danke ich Luisa Horsten, an die ich mich immer mit meinen Fragen zu verschiedenen Themen rund um die Promotion wenden konnte. In diesem Zuge möchte ich auch dem ganzen Graduiertenkolleg *Statistical Modeling in Psychology* (SMiP) danken, für die hervorragenden Weiterbildungsangebote und Austauschmöglichkeiten und natürlich auch der Deutschen Forschungsgemeinschaft (DFG), die den ganzen Ausflug finanziert hat (#GRK 2277)

Zuletzt, aber nicht weniger wichtig, gilt mein Dank natürlich meiner Familie, meinen Freunden und meiner Freundin, die zwar alle nicht bei der Achterbahnfahrt selber dabei sein konnten, aber zumindest von außen zu geschaut haben und mir geholfen haben, wenn ich mich doch mal übernommen hatte und mir schwindlig wurde. Zudem ist es ja bekanntlich alles spaßiger, wenn einem jemand zuschaut und zuwinkt. Mam und Pap, ich kann euch nicht genug danken für all die Liebe, Möglichkeiten und Rückhalt, den ihr mir auf den verschiedensten Phasen meines Lebens immer gegeben habt! Joana, das gleiche gilt für dich in der kurzen Zeit in der wir uns nun schon kennen.

Acknowledgments

As for many doctoral candidates, my time as a doctoral student was often a real (emotional) rollercoaster. As with any good rollercoaster, there were moments of euphoria, but also moments of hesitation and questioning why you went on this ride. On the one hand, you have the emotional highlights, such as when you (think you) finally understand a theory or a construct or when a paper is accepted for publication, on the other hand the low points, e.g., when you realize that you made a mistake in the data analysis or when you receive little constructive feedback during a review process. Even though I sometimes felt a little dizzy after a ride I am happy that I was able to try out a wide variety of rides. I am also grateful that I took these rides in good company.

First of all, I would like to thank my supervisor, Benjamin Hilbig, who, with a similar spirit of adventure, was always with me right at the front of the first car of the roller coaster. In fact, I couldn't have wished for a better fellow passenger (or supervisor and boss). Benni, I am extremely grateful to you for your always honest and constructive feedback, quick availability, the freedom in my way of working, your pragmatic suggestions for solutions and the knowledge that the well-being of your employees and doctoral students always comes first. You have definitely become a great role model for me in these matters. Thank you for pointing out the rather unsafe, yet somewhat enjoyable, rides to me, but then also for guiding and supporting me there!

Next, I'd like to thank my other co-authors (Morten Moshagen, Isabel Thielmann, Ingo Zettler and Johannes Zimmermann) who were right in the car behind us on some of the rides. Morten, Isa, Ingo and Johannes, thank you very much for the exchange of ideas, the constructive feedback and for not letting yourselves be put off by what were perhaps sometimes rather unusual trips!

Johannes Kraus was an equally loyal companion, who encouraged me in the first place and, in many ways, enabled me to dare to take this rollercoaster ride. Johannes, I am very happy and grateful that I can always count on your support and that we have once again embarked on some rides together.

I would also like to thank my colleagues in the Experimental Psychology and Personality Lab and my SMiPster colleagues. Even though we tried out different rollercoasters, it was great to share our experiences! In particular, I would like to thank Luisa Horsten, to whom I could always turn with my questions on various topics relating to the doctorate. Along these lines, I would also like to thank the entire Research Training Group Statistical Modeling in Psychology (SMiP) for the excellent training and exchange opportunities and, of course, the German Research Foundation (DFG), who funded the whole trip (#GRK 2277).

Last but not least, I would of course like to thank my family, my friends and my girlfriend, who, although they couldn't be on the rollercoaster ride themselves, at least watched from the outside and helped me when I got dizzy. Besides, as we all know, everything is more fun when someone is watching and waving at you. Mom and Dad, I can't thank you enough for all the love, opportunities and support you have always given me at different stages of my life! Joana, the same goes for you in the short time we have known each other.

Declaration of Originality

I hereby declare that I have written the present dissertation independently, without assistance from external parties, and without use of other resources than those indicated. The ideas taken directly or indirectly from external sources (including electronic sources) are duly acknowledged in the text. This dissertation has not been previously submitted by me (in full or in part) for grading at this or any other academic or governmental institution. CRediT authorship contribution statements for the jointly authored manuscripts are presented on the following pages.

February 20, 2024 David Scholz

Manuscript: Scholz, D. D., Hilbig, B. E., Thielmann, I., Moshagen, M., & Zettler, I. (2022). Beyond (low) Agreeableness: Towards a more comprehensive understanding of antagonistic psychopathology. *Journal of Personality, 90*(6), 956–970. https://doi.org/10.1111/jopy.12708

CRediT Role	Author 1: David Scholz	Author 2: Benjamin Hilbig	Author 3: Isabel Thielmann	Author 4: Morten Moshagen	Author 5: Ingo Zettler
1. Conceptualization	☒ equal	☒ equal	☒ supporting	☒ supporting	☒ supporting
2. Data curation	☒ lead	☐ --	☒ supporting	☐ --	☐ --
3. Formal analysis	☒ lead	☒ supporting	☐ --	☒ supporting	☐ --
4. Funding acquisition	☐ --	☒ --	☐ --	☐ --	☐ --
5. Investigation	☐ --	☒ equal	☒ equal	☐ --	☐ --
6. Methodology	☒ equal	☒ equal	☒ supporting	☒ supporting	☒ supporting
7. Project administration	☒ equal	☒ equal	☒ equal	☐ --	☐ --
8. Resources	☐ --	☒ --	☐ --	☐ --	☐ --
9. Software	☐ --	☐ --	☐ --	☐ --	☐ --
10. Supervision	☐ --	☒ --	☐ --	☐ --	☐ --
11. Validation	☐ --	☐ --	☐ --	☐ --	☐ --
12. Visualization	☒ lead	☒ supporting	☐ --	☐ --	☐ --
13. Writing – original draft	☒ lead	☒ supporting	☐ --	☐ --	☐ --
14. Writing – review & editing	☒ lead	☒ supporting	☒ supporting	☒ supporting	☒ supporting

Manuscript: Scholz, D. D., Zimmermann, J., Moshagen, M., Zettler, I., & Hilbig, B. E. (in press). Theoretical and empirical integration of 'dark' traits and socially aversive personality psychopathology. *Journal of Personality Disorders*.

CRediT Role	Author 1: David Scholz	Author 2: Johannes Zimmermann	Author 3: Morten Moshagen	Author 4: Ingo Zettler	Author 5: Benjamin Hilbig
1. Conceptualization	☒ equal	☒ equal	☒ supporting	☒ supporting	☒ equal
2. Data curation	☒ lead	☐ --	☐ --	☐ --	☒ supporting
3. Formal analysis	☒ lead	☐ --	☒ supporting	☐ --	☒ supporting
4. Funding acquisition	☐ --	☒ equal	☐ --	☐ --	☒ equal
5. Investigation	☒ equal	☐ --	☐ --	☐ --	☒ equal
6. Methodology	☒ equal	☒ equal	☒ supporting	☒ supporting	☒ equal
7. Project administration	☒ lead	☐ --	☐ --	☐ --	☒ supporting
8. Resources	☒ equal	☒ equal	☒ equal	☒ equal	☒ equal
9. Software	☐ --	☐ --	☐ --	☐ --	☐ --
10. Supervision	☐ --	☐ --	☐ --	☐ --	☒ --
11. Validation	☐ --	☐ --	☐ --	☐ --	☐ --
12. Visualization	☒ lead	☒ supporting	☐ --	☐ --	☒ supporting
13. Writing – original draft	☒ lead	☒ supporting	☐ --	☐ --	☒ supporting
14. Writing – review & editing	☒ lead	☒ supporting	☒ supporting	☒ supporting	☒ supporting

Manuscript: Scholz, D. D. & Hilbig, B. E. (2024). Disentangling the shared and unique aspects of "clinical" and "subclinical" socially aversive traits relevant for (interpersonal) personality dysfunctioning. Manuscript submitted for publication.

CRediT Role	Author 1: David Scholz	Author 2: Benjamin Hilbig
1. Conceptualization	☒ equal	☒ equal
2. Data curation	☒ --	☐ --
3. Formal analysis	☒ lead	☒ supporting
4. Funding acquisition	☐ --	☒ --
5. Investigation	☒ lead	☒ supporting
6. Methodology	☒ equal	☒ equal
7. Project administration	☒ lead	☒ supporting
8. Resources	☒ equal	☒ equal
9. Software	☐ --	☐ --
10. Supervision	☐ --	☒ --
11. Validation	☐ --	☐ --
12. Visualization	☒ lead	☒ supporting
13. Writing – original draft	☒ lead	☒ supporting
14. Writing – review & editing	☒ lead	☒ supporting

Curriculum Vitae

Personal Details

Name:	David D. Scholz
E-Mail:	david.scholz@rptu.de
Research Gate:	https://www.researchgate.net/profile/David_Scholz5

Education

12/2022 – 07/2023	Qualification as a Systemic Consultant (BTB)
10/2017 – 09/2019	Psychology (M. Sc.) at Ulm University
10/2014 – 09/2017	Psychology (B. Sc.) at Ulm University
09/2004 – 06/2013	"Abitur"-certificate at the Peutinger-Gymnasium Augsburg

Scientific Activity

07/2022 – 09/2022:	Lab visit at the University of Copenhagen
Since 10/2020	Ph.D. Student at the Research Training Group "Statistical Modeling in Psychology" (Supervisor: Benjamin E. Hilbig)
Since 08/2020:	Research & teaching assistant at the Experimental Psychology & Personality Lab at the RPTU Kaiserslautern-Landau
10/2018 – 12/2018	Visiting Student at the Tokyo Institute of Technology supported by the PROMOS Scholarship Programme
04/2019 – 08/2019	Statistic tutor for the lecture "Evaluation" by Prof. Moshagen
07/2016 – 09/2019	Student assistant at the department of Human Factors, University of Ulm
05/2015 – 09/2019	Student assistant at the department of Clinical and Health Psychology, University of Ulm

Award, grants, and scholarships

2023	Open Science Award (Gold) of the Open Science Commission, Department of Psychology, RPTU Kaiserslautern-Landau
2022	Open Science Award (Gold) of the Open Science Commission, Department of Psychology, RPTU Kaiserslautern-Landau
2018	PROMOS Scholarship for a master thesis, DAAD (German Academic Exchange Service), University of Ulm

Teaching

Winter 2023/24	Basics of web-based programming (seminar; level: M.Sc.)
Winter 2023/24	Maladaptive traits as flavored manifestations of their aversive core - Part II" (lab-tutorial & methods course; level: B.Sc.)
Summer 2023:	Maladaptive traits as flavored manifestations of their aversive core - Part I" (lab-tutorial & methods course; level: B.Sc.)
Winter 2022/23:	Justifying beliefs for "dark" personality tendencies - Part II" (lab-tutorial & methods course; level: B.Sc.)
Summer 2022:	Justifying beliefs for "dark" personality tendencies - Part I" (lab-tutorial & methods course; level: B.Sc.)
Summer 2021:	Cognitive psychology (seminar; level: B.Sc., online)

Publications

Peer-Reviewed Papers:

Scholz, D. D.*, Kraus, J.*, & Miller, L. (in press). Measuring the Propensity to Trust in Automated Technology: Examining Similarities to Dispositional Trust in Other Humans and Validation of the PTT-A Scale. *International Journal of Human–Computer Interaction*. [*shared first authorship]

Scholz, D.D., Zimmermann, J, Moshagen, M., Zettler, I., & Hilbig, B. E. (in press). Theoretical and empirical integration of 'dark' traits and socially aversive personality psychopathology. *Journal of Personality Disorders.*

Scholz, D. D., Bader, M., Betsch, C., Böhm, R., Lilleholt, L., Sprengholz, P., & Zettler, I. (2023). The moderating role of trust in pandemic-relevant institutions on the relation between pandemic fatigue and vaccination intentions. *Journal of health psychology*, 13591053231201038. https://doi.org/10.1177/13591053231201038

Horsten, L. K., Thielmann, I., Moshagen, M., Zettler, I., **Scholz, D.**, & Hilbig, B. E. (2023). Testing the equivalence of the aversive core of personality and a blend of agreeableness(-related) items. *Journal of personality.* Vorab-Onlinepublikation. https://doi.org/10.1111/jopy.12830

Scholz, D. D., Thielmann, I., & Hilbig, B. E. (2023). Down to the core: The role of the common core of dark traits for aversive relationship behaviors. *Personality and Individual Differences*, *213*, 112263. https://doi.org/10.1016/j.paid.2023.112263

Kraus, J., Miller, L., Klumpp, M., Babel, F., **Scholz, D.**, Merger, J., & Baumann, M. (2023). On the Role of Beliefs and Trust for the Intention to Use Service Robots: An Integrated Trustworthiness Beliefs Model for Robot Acceptance. *International Journal of Social Robotics.* Vorab-Onlinepublikation. https://doi.org/10.1007/s12369-022-00952-4

Scholz, D. D., Hilbig, B. E., Thielmann, I., Moshagen, M., & Zettler, I. (2022). Beyond (low) Agreeableness: Towards a more comprehensive understanding of antagonistic psychopathology. *Journal of personality.* Vorab-Onlinepublikation. https://doi.org/10.1111/jopy.12708

Kraus, J., **Scholz, D.**, & Baumann, M. (2021). What's Driving Me? Exploration and Validation of a Hierarchical Personality Model for Trust in Automated Driving. *Human factors*, *63*(6), 1076–1105. https://doi.org/10.1177/0018720820922653

Kraus, J., **Scholz, D.,** Stiegemeier, D., & Baumann, M. (2020). The More You Know: Trust Dynamics and Calibration in Highly Automated Driving and the Effects of Take-Overs, System Malfunction, and System Transparency. *Human factors*, *62*(5), 718–736. https://doi.org/10.1177/0018720819853686

Kraus, J., **Scholz, D.**, Messner, E.-M., Messner, M., & Baumann, M. (2019). Scared to Trust? - Predicting Trust in Highly Automated Driving by Depressiveness, Negative Self-Evaluations and State Anxiety. *Frontiers in Psychology*, *10*, 2917. https://doi.org/10.3389/fpsyg.2019.02917

Conference Contributions:

Scholz, D., Thielmann, I., & Hilbig, B. E. (2023). *Besinnung auf Gemeinsamkeiten statt Unterschiede: Die Rolle des gemeinsamen Kerns der dark tetrad für aversives Verhalten in Beziehungen*. Biennial Conference of the German Psychological Society - Personality Psychology and Psychological Diagnostics (DPPD) Section, 25-27 September 2023 in Salzburg, Austria

Scholz, D., Kraus, J., & Miller, L. (2022). *Comparing the structure of Propensity to Trust in Human-Human and Human-Technology interaction*. Tagung experimentell arbeitender Psycholog:innen (Conference of Experimental Psychologists; TEAP), 20-23 March 2022 in Cologne, Germany

Scholz, D., Hilbig, B. E., Thielmann, I., Moshagen, M., & Zettler, I. (2021). *The Differential Roles of Agreeableness, Honesty Humility and the Dark Factor of Personality for Antagonistic Psychopathology*. Biennial Conference of the German Psychological Society - Personality Psychology and Psychological Diagnostics (DPPD) Section, 12-15 September 2021 in Ulm, Germany

Posters:

Scholz, D. *The dark factor of personality as a common basis for aversive maladaptive traits.* SMiP Fall Retreat, 25-26 November 2022.

Scholz, D. *Beyond (low) Agreeableness: A closer view on antagonistic psychopathology.* Online SMiP-IOPS Conference, 08-09 June 2021.

Ad-hoc reviews

- Computers in Human Behavior
- European Journal of Personality
- Personality and Individual Differences
- Personality and Mental Health

Appendix: Copies of Articles

Received: 24 November 2021 | Revised: 25 January 2022 | Accepted: 13 February 2022

DOI: 10.1111/jopy.12708

ORIGINAL ARTICLE

WILEY

Beyond (low) Agreeableness: Toward a more comprehensive understanding of antagonistic psychopathology

David D. Scholz[1] | **Benjamin E. Hilbig**[1] | **Isabel Thielmann**[1,2] | **Morten Moshagen**[3] | **Ingo Zettler**[4,5]

[1]Department of Psychology, University of Koblenz-Landau, Landau, Germany

[2]Department of Criminology, Max Planck Institute for the Study of Crime, Security and Law, Freiburg im Breisgau, Germany

[3]Institute of Psychology and Education, University of Ulm, Ulm, Germany

[4]Department of Psychology, University of Copenhagen, Copenhagen, Denmark

[5]Copenhagen Center for Social Data Science (SODAS), University of Copenhagen, Copenhagen, Denmark

Correspondence
David D. Scholz, Cognitive Psychology, University of Koblenz-Landau, Campus Landau, Fortstraße 7, Landau 76829, Germany.
Email: scholz@uni-landau.de

Funding information
Deutsche Forschungsgemeinschaft

Abstract

In clinical psychopathology research, up to seven traits have been suggested as instances of antagonistic psychopathology. Those antagonistic traits, in turn, are commonly viewed as reflections of low Agreeableness as per the Big Five (BF-AG). However, specific theoretical differences between antagonistic traits suggest that other broad, basic dimensions beyond BF-AG ought to provide further points of correspondence. Specifically, whereas primarily affective antagonistic traits are closely aligned with BF-AG, primarily behavioral antagonistic traits are better aligned with Honesty-Humility (HH) from the HEXACO model and primarily cognitive antagonistic traits are better aligned with the common core of aversive traits (the Dark Factor of Personality, D). Indeed, from a theoretical perspective, D seems to be the only candidate sufficiently covering all aspects of antagonistic traits (affect, behavior, and cognition) to a comparable extent and thus affording a balanced representation of antagonistic psychopathology. We critically test these conjectures in a large and heterogeneous online sample ($N = 3{,}396$), investigating the overlap between antagonistic traits and basic personality via structural equation modeling. Results show that BF-AG, HH, and D each yield particularly strong ties to one group of antagonistic traits (affective, behavioral, and cognitive, respectively), while D offers the most balanced representation of all (groups of) antagonistic traits.

KEYWORDS

Agreeableness, Antagonism, Dark Factor of Personality, Honesty-Humility, psychopathology

1 | INTRODUCTION

Recent years have seen notable shifts within clinical psychology toward the constructs and frameworks more traditionally studied in personality science. Specifically, there has been a shift from categorical to dimensional models of (personality) disorders (Hopwood et al., 2018; Kotov et al., 2017; Widiger et al., 2019). The underlying idea is that clinically relevant tendencies represent maladaptive variants or levels of more general traits, such as those defined in models of basic personality structure, especially the Five-Factor Model (Samuel & Widiger, 2008).

In the realm of socially and/or ethically aversive psychopathology, seven clinically relevant traits have been

suggested in the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association, 2013), all of which are currently used for personality disorder (PD) diagnoses according to the DSM-5 alternative model: attention-seeking, callousness, deceitfulness, grandiosity, hostility, manipulativeness, and suspiciousness. Despite some ongoing discussions on which of these traits actually define the higher-order domain of Antagonism in the sense of a hierarchical model (see e.g., Al-Dajani et al., 2016; Kotov et al., 2017; Sleep et al., 2020), there is consensus that the most prominent aspect shared by these seven traits is Antagonism and they thereby "put the individual at odds with other people" (American Psychiatric Association, 2013). Following the terminology used in psychopathology, we term these seven "antagonistic traits", that is, traits that share the aspect of antagonistic psychopathology (see Widiger & Simonsen, 2005). These traits and their definitions are presented in Table 1 along with example items from the most prominent corresponding measure, the Personality Inventory for DSM-5 (PID-5; Krueger et al., 2012).

In line with the view that clinically relevant tendencies represent maladaptive variants of broad and general basic traits, these seven antagonistic traits – through their common aspect of Antagonism – have been suggested to reflect low Agreeableness (AG; see, e.g., American Psychiatric Association, 2013; Krueger et al., 2012; Krueger & Markon, 2014; Sleep et al., 2020; Widiger et al., 2019) as per the Big Five (BF; McCrae & Costa, 2008). In other words, it is assumed that the common feature of these seven antagonistic traits, Antagonism, essentially reflects (extremely) low BF-AG. BF-AG is defined as the "motivation to maintain positive relations with others" (Graziano & Tobin, 2009, p. 46)[1] and by adjectives such as kind, sympathetic, undemanding, and warm versus harsh, unsympathetic, demanding, and cold (Saucier, 2002). Thus, it indeed appears antithetical to the Antagonism aspect shared by all seven antagonistic traits.

Following this reasoning, the seven antagonistic traits should all be strongly reflected in (low) BF-AG. However, on closer theoretical inspection, it appears that, whereas some of the antagonistic traits are more clearly aligned with BF-AG, other represent features that are not (as comprehensively) covered by BF-AG and indeed systematically so. Therefore, they may be better understood as reflections of broad personality traits other than BF-AG.

In essence, we argue that BF-AG cannot account for all antagonistic traits *alike*. Instead, we will theoretically derive that BF-AG should primarily reflect those antagonistic traits that are focused on affect, whereas it is less strongly aligned with traits that are focused on behavior and (social) cognition. These latter aspects, in turn, are key to other broad personality traits, namely Honesty-Humility from the HEXACO model (Ashton & Lee, 2007) and the Dark Factor of Personality (Moshagen et al., 2018). The

TABLE 1 Antagonistic traits with definitions and example items, sorted by their primary aspect

Primary aspect	Trait	Definition DSM-5 (American Psychiatric Association, 2013)	Example item PID-5
Affect	Callousness	Lack of concern for the feelings or problems of others; lack of guilt or remorse about the negative or harmful effects of one's actions on others	I don't care about other peoples' problems.
	Hostility	Persistent or frequent angry feelings; anger or irritability in response to minor slights and insults; mean, nasty, or vengeful behavior	I have a very short temper.
Behavior	Manipulativeness	Use of subterfuge to influence or control others; use of seduction, charm, glibness, or ingratiation to achieve one's ends	Sweet-talking others helps me get what I want.
	Deceitfulness	Dishonesty and fraudulence; misrepresentation of self; embellishment or fabrication when relating events	I'll stretch the truth if it's to my advantage.
	Attention-Seeking	Engaging in behavior designed to attract notice and to make oneself the focus of others' attention and admiration	I like to draw attention to myself.
Cognition	Grandiosity	Believing that one is superior to others and deserves special treatment; self-centeredness; feelings of entitlement; condescension toward others	I'm better than almost everyone else.
	Suspiciousness	Expectations of – and sensitivity to – signs of interpersonal ill-intent or harm; doubts about loyalty and fidelity of others; feelings of being mistreated, used, and/or persecuted by others	Plenty of people are out to get me.

goal of the current work is to test this proposition empirically so as to avail a more thorough understanding of the (maladaptive) dispositional basis of antagonistic psychopathology.

1.1 | Theoretical differences between antagonistic traits

Personality traits, maladaptive or not, by most common definitions encompass typical tendencies related to affect, behavior, and/or cognitions to some degree (Funder, 2001). Nonetheless, traits can encompass one or more of these three aspects to different degrees. For example, BF Neuroticism primarily encompasses affective aspects (anxious, touchy, nervous versus unemotional, relaxed; Goldberg, 1992). By comparison, BF Openness subsumes far more – though not exclusively – cognitive aspects (creative, bright, artistic versus unimaginative, imperceptive; Goldberg, 1992). In other words, any trait can have a specific theoretical focus on affect, behavior, or cognition.

Likewise, antagonistic traits – despite their shared prominent aspect of being socially aversive – also arguably differ in how strongly they reflect affect, behavior, or cognition. For example, manipulativeness, defined as the use of subterfuge or ingratiation to achieve one's ends, primarily describes a *behavior* associated with an antagonistic disposition, hardly referring to any affect or cognitions accompanying such behavior. By comparison, grandiosity, defined as the belief of superiority, primarily refers to the *cognition* aspect of traits. Specifically, it constitutes a justification for antagonistic behavior, namely, one's perceived higher value and deservingness (Calvete, 2008; Calvete & Orue, 2013).

Due to the different focus of antagonistic traits on affect, behavior, and cognition, we will refer to traits focusing primarily on the same aspect as a *group of traits* in what follows. More specifically, we will refer to callousness and hostility as primarily affective antagonistic traits, to manipulativeness, deceitfulness, and attention-seeking as primarily behavioral antagonistic traits, and to grandiosity and suspiciousness as primarily cognitive antagonistic traits. Note that these labels are not meant to imply that any of the antagonistic traits is purely or exclusively affective, behavioral, or cognitive - let alone that traits can be grouped into three empirical factors. Indeed, a trait with one particular focus may nonetheless correlate more strongly with traits involving a different focus than with traits involving the same. For example, although grandiosity and suspiciousness have a cognitive focus, grandiosity usually correlates much stronger with deceitfulness than with suspiciousness (Bo et al., 2016; Crego et al., 2015; Quilty et al., 2013). This is highly plausible because the belief that one is superior to others and deserves special treatment will often facilitate dishonest behavior, but has no similarly obvious connection with how strongly one assumes ill intent of others, i.e., trait suspiciousness. Nevertheless, the antagonistic traits systematically differ in which of the three aspects is primary in their definition and operationalization.

1.2 | Antagonistic traits focusing on affect: The role of Agreeableness

Given that antagonistic traits differ in these primary defining aspects (i.e., their focus on affect, behavior, or cognition), one might question whether low BF-AG serves as an *equally* suitable counterpart for all these traits. BF-AG has a strong focus on affect, as is apparent from adjectives defining the BF-AG domain. As, for example, noted by Tobin et al. (2000) the "best Agreeableness markers are emotion terms like 'kind', 'considerate', 'empathic', and 'tender-minded'" (pp. 656–657). Consequently, the primarily affective antagonistic traits most closely align with BF-AG on a theoretical level. Specifically, callousness (defined as "lack of concern for the feelings or problems of others"; American Psychiatric Association, 2013) resembles adjectives like warm, empathic, and tender-minded. Likewise, hostility (defined as "persistent or frequent angry feelings"; American Psychiatric Association, 2013) relates closely to adjectives such as kind or harsh. Supporting this reasoning, the observed correlations of BF-AG with these primarily affective antagonistic traits are particularly strong ($r \leq -0.58$, Quilty et al., 2013; Watson et al., 2013).

By comparison, traits related to the behavioral manifestation of an antagonistic disposition, e.g., exploitative (manipulativeness), dishonest (deceitfulness), and admiration-seeking behavior (attention-seeking), are more distal to BF-AG (Ashton & Lee, 2005; Lee & Ashton, 2005). Those aspects are arguably more aligned with other broad trait dimensions beyond BF-AG as we will reason in what follows.

1.3 | Antagonistic traits focusing on behavior: The role of Honesty-Humility

A broad personality trait well-suited to particularly capture behavioral aspects of socially aversive psychopathology, namely exploitation, dishonesty, and admiration seeking, is Honesty-Humility (HH) from the HEXACO model (Ashton & Lee, 2007; Ashton et al., 2004). It is defined as "the tendency to be fair and genuine in dealing with others in the sense of cooperating with others

even when one might exploit them without suffering retaliation" (Ashton & Lee, 2007, p. 156) and defined by adjectives such as sincere, fair-minded, unassuming, and modest versus hypocritical, sly, greedy, and boastful (Ashton & Lee, 2007).

Thus, by its very definition, (non-)exploitation is central to HH which renders it distinct from BF-AG (Ashton & Lee, 2020). Indeed, HH typically outperforms BF-AG in the prediction of exploitative and dishonest behaviors (Ashton & Lee, 2008, 2020; Heck et al., 2018; Thielmann et al., 2020). Likewise, greed and actively searching for social attention or status is a defining feature of (low) HH (Ashton et al., 2014), whereas these aspects are largely absent in BF-AG (Ashton & Lee, 2005). Therefore, HH constitutes a prime candidate for traits emphasizing the behavioral expressions of a socially aversive disposition. For the primarily affective antagonistic traits, by contrast, HH bears little conceptual overlap; indeed, affective aspects are beyond the scope of HH in the HEXACO model and assigned to the Emotionality and (HEXACO-)Agreeableness dimensions (see Ashton & Lee, 2005).

Although HH conceptually encompasses aspects beyond the realm of BF-AG (the primarily behavioral antagonistic traits of manipulativeness, deceitfulness and attention-seeking), and vice versa (the primarily affective antagonistic traits hostility and callousness), *both* involve only limited theoretical focus on the more cognitive aspects of traits relating to beliefs about others (suspiciousness) and oneself (grandiosity). These tendencies, in turn, are a defining aspect of the Dark Factor of Personality (D), the common core of so-called dark traits (Moshagen et al., 2018), i.e., socially and/or ethically aversive traits.

1.4 | Antagonistic traits focusing on cognition: The role of D

D is conceptualized as the underlying disposition from which all aversive traits arise as specific, flavored manifestations. In other words, it is the common core of all aversive traits, defined as "the general tendency to maximize one's individual utility – disregarding, accepting, or malevolently provoking disutility for others – *accompanied by beliefs that serve as justifications*" (Moshagen et al., 2018, p. 657, emphasis added). As per this definition, D places strong focus on the beliefs and attitudes which serve to justify one's behavior, including grandiosity and suspiciousness. By comparison, BF-AG only peripherally reflects modesty (vs. grandiosity) and trust (vs. suspiciousness), and HH only captures beliefs of superiority or entitlement (grandiosity), without any direct reference to trust versus suspiciousness. Correspondingly, it has been shown that D is strongly associated with several (classes of) justifying beliefs and predicts incremental variance over and above BF-AG and HH (and even the entire HEXACO space) in distrust and hierarchy-related beliefs in particular (Hilbig, Moshagen, Thielmann, et al., 2022; Horsten et al., 2021; Moshagen et al., 2018, 2020). Moreover, D predicted incremental variance beyond the HEXACO space in instances of aversive psychopathology that are primarily related to beliefs, especially narcissism (for which believing in one's superiority is integral) and paranoia (for which suspicion toward others is integral; Hilbig, Thielmann, et al., 2021). Therefore, D is arguably a prime candidate (per se and as compared to BF-AG and HH) to account for the primarily cognitive antagonistic traits, that is, grandiosity and suspiciousness.

1.5 | A balanced view of antagonistic psychopathology

Overall, both BF-AG and HH correspond primarily to one of the three groups of antagonistic traits. Specifically, BF-AG mostly represents the primarily affective antagonistic traits with notably less theoretical overlap with behavioral traits both in absolute terms and *in comparison* to HH and D. In turn, HH mostly represents the primarily behavioral antagonistic traits, with practically no theoretical reference to affective aspects, both in absolute terms and *in comparison* to BF-AG and D. Moreover, both BF-AG and HH show only limited theoretical overlap with the primarily cognitive antagonistic traits both in absolute terms and *in comparison* to D.

Thus, in contrast to BF-AG and HH, D does not primarily represent (only) one group of antagonistic traits, but actually yields comparable theoretical overlap with all three.[2] By definition, D encompasses cognition (justifying beliefs), behavior (serving the aim of utility maximization), *and* affective aspects (lack of concern for the disutility in others). By implication, D offers a relatively balanced representation of antagonistic traits overall – and more so than BF-AG or HH – involving correspondence with traits from all three groups to a relatively similar extent. This claim is supported by findings indicating that D predicts incremental variance in regard to dishonest and cheating behavior over and above BF-AG and incremental variance in outcomes related to aggression and insensitivity over and above HH (Moshagen et al., 2018). Stated simply, D should essentially cover affect, behavior, and cognition comparably – whereas the reasoning above suggests that BF-AG and HH should not (each corresponds more closely with a particular subset of antagonistic traits).

1.6 | Hypotheses

Given the reasoning above, our **first set of hypotheses (H1)** maintains that each broad trait "candidate" (BF-AG, HH, D) will be substantially related to one group of antagonistic traits overall, and also with the respective traits of that group individually. Specifically, we expect BF-AG to be substantially correlated with primarily affective antagonistic traits (**H1a**), whereas HH will correlate substantially with primarily behavioral antagonistic traits (**H1b**) and D will correlate substantially with primarily cognitive antagonistic traits (**H1c**). Following common standards (Cohen, 1992), we consider a large effect size of approximately $r = 0.50$ (and thus $R^2 = 0.25$) or higher to be substantial.

Second, it follows from the above that each candidate (BF-AG, HH, D) ought to predict their respective trait group and the individual traits (as per H1) better than the respective other two candidates, which is our **second set of hypotheses (H2)**. Specifically, *out of all candidates,* BF-AG should relate most strongly to primarily affective antagonistic traits (**H2a**), whereas HH should relate most strongly to the behavioral antagonistic traits (**H2b**), and D most strongly to the cognitive antagonistic traits (**H2c**).

Third, as argued above, D theoretically represents all groups of traits (affective, behavioral, and cognitive) to a comparable extent and indeed more so than BF-AG and HH. Thus, our **third set of** hypotheses (**H3**) predicts that D entails *unique* shared variance with primarily behavioral antagonistic traits over and above BF-AG (**H3a**) and with primarily affective antagonistic traits over and above HH (**H3b**). Finally, D should predict unique variance in primarily cognitive antagonistic traits over both BF-AG and HH, respectively (**H3c**). The hypotheses were not pre-registered.

2 | METHOD

2.1 | Participants

Data for all analyses were collected as part of the Prosocial Personality Project, which contained six waves of data collection for the base study. The data used here come from the first two waves. Participants were recruited via a professionally managed online panel in Germany to take part in the first wave (T1) and re-invited 41 days later on average to wave 2 (T2). A detailed documentation of the project as well as the a priori specified exclusion criteria and other publications using (other subsets of) data from the project are available on the Open Science Framework (OSF; https://osf.io/c9kma/). Applying the specified exclusion criteria to the data used herein led to a final sample of $N = 3{,}396$ for the current analyses (1,694 female, 1,696 male, 6 diverse; aged 18–74 years, $M = 42.0$, $SD = 12.8$). Participants were financially remunerated (time-based) by the panel provider. The present study was conducted in full accordance with the Ethical Guidelines of the American Psychological Association (APA). The study was run based on approval from the local ethics committee.

2.2 | Measures

All measures and items used are provided in the documentation of the Prosocial Personality Project (see participants). At T1, scales for D, HH, and BF-AG were administered. To measure D, we used the German version of the 70-item set (D-70; Moshagen et al., 2020). To measure HH, we relied on the German version (Moshagen et al., 2014) of the 60-item HEXACO Personality Inventory-Revised (HEXACO-60; Ashton & Lee, 2009) which measures each of the six HEXACO dimensions (including HH) with 10 items. BF-AG was measured by AG scales from four well-established and widely used BF inventories, namely the NEO Five Factor Inventory (NEO-FFI; 12 items; Costa & McCrae, 1992; McCrae & Costa, 2004; German translation by Borkenau & Ostendorf, 2008), the Big Five Aspects Scales (BFAS; 20 items; DeYoung et al., 2007; German translation by Mussel & Paelecke, 2018), the International Personality Item Pool (IPIP) Big Five scales (8 items[3]; Goldberg, 1992; German translation by Treiber et al., 2013), and the Big Five Inventory 2 (BFI-2; 12 items; Soto & John, 2017; German translation by Danner et al., 2019). Thus, BF-AG was indicated by 50 items in total. To account for the imbalance in the number of items used to measure BF-AG and D in comparison to HH, additional analyses with fewer indicators for BF-AG and D were conducted (for details see the modeling section). All responses at T1 were given on five-point Likert-type scales ranging from "strongly disagree" to "strongly agree". The order of the scales at T1 was randomized between participants.

At T2, the antagonistic traits were measured using the 100-item version of the PID-5 (Maples et al., 2015) with the German translation from Zimmermann et al. (2014). Thus, each antagonistic trait was measured using four items. The PID-5 is the most widely used inventory for measuring maladaptive traits according to the alternative model for personality disorders of the DSM-5 (Krueger & Hobbs, 2020; Watters & Bagby, 2018). For the PID-5, responses were given on a four-point Likert-type scale ranging from "does not apply at all" to "totally applies".

2.3 | Modeling

All models were identified using reference indicators and raw item scores were used for model estimation. For BF-AG and D, a bi-factor modeling approach was chosen (e.g. Reise, 2012; for additional analyses with each AG scale as a single latent factor, please see Supplement 3 on the OSF). Bi-factor models decompose the shared variance of the manifest indicators between (typically) one general factor and multiple orthogonal specific factors or facets, respectively. The general factor extracts the shared variance across all observed indicators, whereas the specific factors capture the remaining shared variances between certain items. Although for both BF-AG and D a higher-order model can also be appropriate, higher-order models require more assumptions than bi-factor models (Bader & Moshagen, 2021).

In the case of BF-AG (see Figure 1), all 50 items of the four AG scales served as indicators (for similar approaches modeling a general AG-factor, see, e.g., Crowe et al., 2018; Moshagen, Zettler, Horsten, et al., 2020). Hence, the general factor extracted from the bi-factor model represents the broadest instance of BF-AG possible as compared to any single scale. The specific factors were orthogonal to the general factor and to each other so as to account for the residualized variance between the items of each BF-AG scale. Thus, the specific factors in this model capture the shared variance unique to items of a specific BF scale. In turn, given that residualized specific factors are generally difficult to interpret, especially in the presence of a strong general factor (Sellbom & Tellegen, 2019), we did not analyze them any further.

To control for the disbalance in the number of items used to measure BF-AG as compared to HH, a reduced set of items was also used to model BF-AG in a separate analysis. Specifically, the 15 items with the highest loading on the general BF-AG factor in the bi-factor model were selected (see Supplement 1 on the OSF for details). These items were then specified to load on a single BF-AG factor.

For D, we chose the bi-factor model as suggested by Bader et al. (2021). Here, D constitutes the general factor subsuming most of the variance of all items. The residualized covariances of the items are further subsumed by five specific factors (called themes), which are orthogonal to each other and to the general factor. Again, as residualized specific factors are generally difficult to interpret in the presence of a strong general factor such as D, we did not consider them any further. As for BF-AG, an additional analysis with a reduced set of 16 items for D, i.e., the D16 (Moshagen, Zettler, & Hilbig, 2020), was conducted, to control for the disbalance in the amount of items used to measure D and HH. As with the reduced BF-AG factor, the D16 was modeled as a single factor.

Finally, HH was modeled as a single factor, with all ten HH items loading on that factor. The antagonistic traits were modeled as single factors, too, with their four respective items as indicators.

2.4 | Analysis

All analyses were conducted using the lavaan package (version 0.6-6; Rosseel et al., 2021) in R (version 4.0.2; R Core Team, 2021) relying on the MLM estimator, which uses robust standard errors and a Satorra-Bentler scaled test statistic to address multivariate non-normal distributed data. As the chi-square statistic is uninformative in large samples with high statistical power,[4] we considered

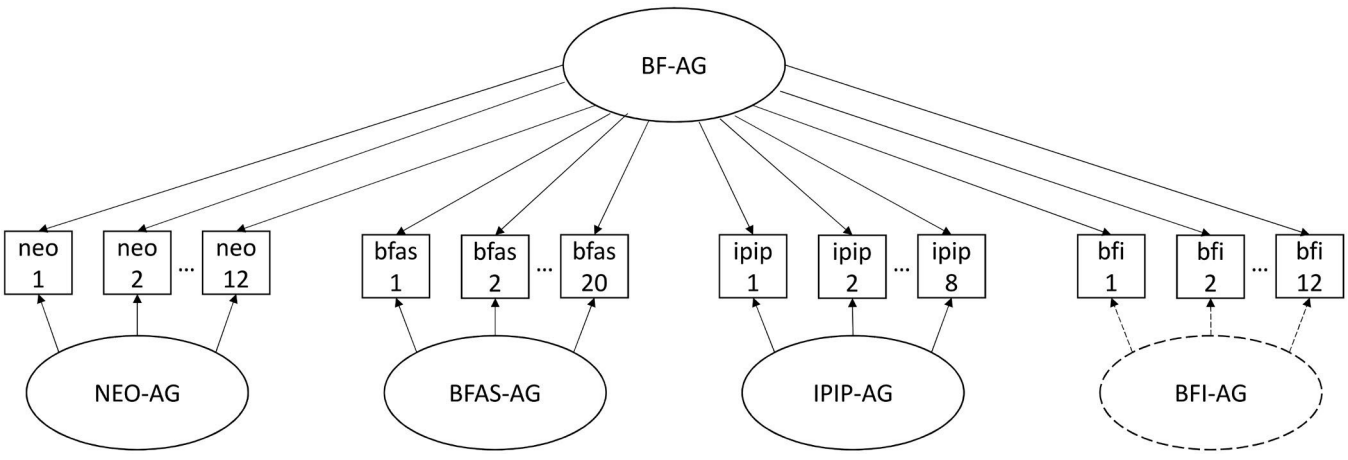


FIGURE 1 Structure of the bi-factor model for Big Five Agreeableness (BF-AG). Here, BF-AG represents the general factor affecting all items of the four Big Five inventories. The scales are modeled as specific factors capturing shared variance among their items that is not accounted for by the general BF-AG factor. The dashed lines for the BFI-AG specific factor indicate that it was omitted from all further analyses because its variance was close to zero.

SRMR and RMSEA as primary indicators of model fit, as these are sensitive to different types of model misspecifications (Moshagen & Auerswald, 2018).

Additionally, to evaluate the proportion of variance that is captured by the general factor in contrast to the specific factors in each of the two bi-factor models (BF-AG and D), we considered the explained common variance (*ECV*; Berge & Sočan, 2004). The *ECV* varies between 0 and 1, with 0 indicating that no variance can be explained by the general factor and 1 indicating that all explained variance in the items is due to the general factor. Hence, higher values indicate a strong general factor (Rodriguez et al., 2016).

To test our hypotheses, we estimated latent regression models. First, to test H1 and H2, each of the antagonistic traits was regressed on the three candidates (i.e., BF-AG, HH, and D) separately to obtain zero-order effects (R^2). Next, to test H3, each antagonistic trait was regressed on the three candidates as predictors in a single model, assessing the unique effect of each candidate over and above the other two. Differences in explained variance are descriptively indicated by ΔR^2. Classical significance tests were supplemented by normalized Evidence Ratios (*ER*; Wagenmakers & Farrell, 2004; Wu et al., 2020) to inform the hypotheses of H3. An *ER* indicates the likelihood given the data that the more restricted model is superior to a less restricted comparison model and ranges from 0 to 1. For example, $ER = 0.90$ indicates that the less restricted model is 9 times ($0.90/(1–0.90) = 9$) more likely than the more restricted model given the data. To determine the *p*-value and *ER* of the incremental effects, we compared a model with BF-AG and D (H3a), a model with HH and D (H3b), and a model with BF-AG, HH, and D (H3c) as predictors, all freely estimating the effect of the three predictors, against a respective comparison model fixing one regression coefficient to zero. These analyses were not pre-registered.

3 | RESULTS

Latent correlations and internal consistencies of all traits are reported in Table 2. The data and analysis scripts are available on the OSF, along with the results of the supplementary analyses (https://osf.io/c9kma/). The supplementary analyses using the reduced (15- and 16-) item sets for BF-AG and D led to identical conclusions regarding all hypotheses.

The bi-factor model for BF-AG yielded satisfactory fit to the data, $\chi^2\,(1234) = 18{,}490$, $p < 0.001$, $SRMR = 0.070$, $RMSEA = 0.063$, 90% CI [0.062, 0.064]. All items loaded on the general BF-AG factor ($Mdn\ \lambda = 0.52$); thus, reliability of the general BF-AG factor was high ($\omega = 0.94$). Moreover, the ECV indicated that 81% of the common variance among all BF-AG indicators was accounted for by the general BF-AG factor. As the BFI-2 specific factor in the bi-factor model had practically no variance, it was omitted in all further analyses, which did not affect model fit in any meaningful way ($\Delta RMSEA = 0.001$, $\Delta SRMR = 0.001$). Moreover, the general BF-AG factor strongly subsumed the common explained variance of the indicators from all remaining specific factors ($ECV_{NEO} = 0.76$, $ECV_{IPIP} = 0.76$, $ECV_{BFAS} = 0.75$).

Similarly, for D, the bi-factor structure resulted in a good fit to the data, $\chi^2\,(2275) = 14{,}242$, $p < 0.001$, $SRMR = 0.046$, $RMSEA = 0.039$, 90% CI [0.039, 0.040]. All items loaded on the general factor ($Mdn\ \lambda = 0.46$). Thus, the observed reliability of D was also high ($\omega = 0.90$). Correspondingly, as indicated by the *ECV*, D explained 74% of common variances among all its indicators and was strongly reflected in all five themes ($0.62 \leq ECV \leq 0.79$). Finally, the single factor for HH yielded an acceptable fit to the data, $\chi^2\,(35) = 1{,}724$, $p < 0.001$, $SRMR = 0.088$, $RMSEA = 0.119$, 90% CI [0.115, 0.124] as did modeling the antagonistic traits as correlated single factors, $\chi^2\,(329) = 3{,}011$, $p < 0.001$, $SRMR = 0.062$, $RMSEA = 0.049$, 90% CI [0.048, 0.050].

Table 3 shows the zero-order effects of all antagonistic traits regressed on each of the candidates (BF-AG, HH, and D). First, as hypothesized (H1a), BF-AG was substantially related to both of the primarily affective antagonistic traits, that is, callousness and hostility, individually and on average. Second, HH was substantially related to the behavioral trait group, thus supporting H1b. Specifically, HH correlated substantially with two of the three behavioral traits, that is, manipulativeness and deceitfulness. A weaker (but still moderate) correlation was found for attention-seeking. Finally, D correlated substantially with grandiosity and suspiciousness individually and thereby with the primarily cognitive antagonistic traits on average, supporting H1c.

As per H2, we further expected each of the candidates to more strongly relate to their respective trait group than the other two candidates. In line with expectations, BF-AG explained most variance in the primarily affective antagonistic traits out of all candidates, both overall and in callousness and hostility individually, supporting H2a. In turn, HH explained more variance than BF-AG or D in all three primarily behavioral antagonistic traits, that is, manipulativeness, deceitfulness, and attention-seeking, therefore supporting H2b. Finally, in line with H2c, D related to the two primarily cognitive antagonistic traits, that is, grandiosity and suspiciousness, more strongly than BF-AG or HH.

In reference to H3, Table 4 shows the unique variance of D over and above BF-AG and HH, respectively, for each antagonistic trait (group). As hypothesized (H3a),

TABLE 2 Latent correlations and internal consistencies (Cronbach's alpha, in bold on the diagonal)

	Man	Dec	Cal	Gran	Att	Hos	Sus	BF-AG	HH	D
Man	**0.69**									
Dec	0.97 [0.95, 0.99]	**0.81**								
Cal	0.68 [0.63, 0.72]	0.71 [0.68, 0.75]	**0.83**							
Gran	0.70 [0.67, 0.74]	0.70 [0.66, 0.73]	0.63 [0.59, 0.67]	**0.79**						
Att	0.51 [0.47, 0.55]	0.46 [0.42, 0.50]	0.22 [0.17, 0.26]	0.56 [0.52, 0.60]	**0.83**					
Hos	0.38 [0.33, 0.42]	0.47 [0.43, 0.51]	0.46 [0.43, 0.50]	0.35 [0.31, 0.40]	0.17 [0.12, 0.21]	**0.80**				
Sus	0.39 [0.35, 0.44]	0.52 [0.48, 0.56]	0.53 [0.48, 0.57]	0.51 [0.46, 0.55]	0.21 [0.17, 0.26]	0.64 [0.60, 0.67]	**0.70**			
BF-AG	−0.50 [−0.54, −0.46]	−0.54 [−0.57, −0.51]	−0.67 [−0.70, −0.64]	−0.46 [−0.50, −0.42]	−0.14 [−0.18, −0.10]	−0.48 [−0.52, −0.45]	−0.39 [−0.43, −0.34]	**0.94**		
HH	−0.58 [−0.62, −0.54]	−0.65 [−0.68, −0.62]	−0.47 [−0.50, −0.43]	−0.52 [−0.56, −0.49]	−0.35 [−0.39, −0.31]	−0.35 [−0.39, −0.31]	−0.32 [−0.37, −0.27]	0.55 [0.52, 0.58]	**0.75**	
D	0.52 [0.48, 0.55]	0.61 [0.57, 0.64]	0.64 [0.61, 0.67]	0.61 [0.58, 0.64]	0.25 [0.21, 0.29]	0.33 [0.29, 0.37]	0.46 [0.42, 0.50]	−0.70 [−0.73, −0.67]	−0.73 [−0.75, −0.70]	**0.95**

Note: Square brackets indicate a 95% confidence interval.

Abbreviations: Att, attention-seeking; BF-AG, Big Five Agreeableness; Cal, callousness; D, Dark Factor of Personality; Dec, deceitfulness; Gran, grandiosity; HH, Honesty-Humility; Hos, hostility; Man, manipulativeness; Sus, suspiciousness

TABLE 3 Explained variance (R^2) in each antagonistic trait (group) by BF-AG, HH, and D

Outcome	BF-AG (%)	HH (%)	D (%)
Primarily affective antagonistic traits (mean)	**35**	17	26
Callousness	**45**	22	41
Hostility	**24**	12	11
Primarily behavioral antagonistic traits (mean)	19	**29**	23
Manipulativeness	25	**33**	26
Deceitfulness	29	**42**	37
Attention-seeking	2	**12**	6
Primarily cognitive antagonistic traits (mean)	18	19	**29**
Grandiosity	21	27	**36**
Suspiciousness	15	10	**22**

Note: Hypothesized (and highest) explained variance in each trait (group) is printed in bold.

Abbreviations: BF-AG, Big Five Agreeableness; D, Dark Factor of Personality; HH, Honesty-Humility.

TABLE 4 Incremental variance (ΔR^2) in each antagonistic trait (group)

Outcome	D > BF-AG (%)	D > HH (%)
Primarily affective antagonistic traits (mean)	3	**10**[a]
Callousness	6	**19**[a]
Hostility	<1	**1**[a]
Primarily behavioral antagonistic traits (mean)	**7**[b]	2
Manipulativeness	**5**[b]	2
Deceitfulness	**11**[b]	4
Attention-Seeking	**4**[b]	<1
Primarily cognitive antagonistic traits (mean)	**12**[c]	**12**[c]
Grandiosity	**16**[c]	**11**[c]
Suspiciousness	**8**[c]	**12**[c]

Note: Hypothesized incremental variances are printed in bold. All effects ≥1% are significant ($p < 0.001$) and have an *ER* of ≥ 0.998.

Abbreviations: BF-AG, Big Five Agreeableness; D, Dark Factor of Personality; HH, Honesty-Humility.

[a]Effects related to Hypothesis 3a.

[b]Effects related to Hypothesis 3b.

[c]Effects related to Hypothesis 3c.

D predicted considerably more variance in primarily behavioral antagonistic traits over and above BF-AG (overall 7%), especially in deceitfulness. Moreover, in line with H3b, D predicted 10% incremental variance in the primarily affective antagonistic traits over and above HH, though almost exclusively in the callousness trait. Finally, as expected (H3c), D explained 12% incremental variance in the primarily cognitive antagonistic traits over and above both BF-AG and HH, respectively.

4 | DISCUSSION

In clinical psychopathology research, seven traits have been suggested to share the common aspect of Antagonism which, in turn, has repeatedly been equated with (extremely) low Agreeableness (AG) as per the Big Five (BF) approach. However, other broad traits, particularly Honesty-Humility (HH) from the HEXACO model and the Dark Factor of Personality (D), are reasonable candidates to reflect antagonistic traits that actually show only limited theoretical overlap with BF-AG. The present work set out to critically test the role of these broad, basic personality dimensions beyond BF-AG in accounting for the seven antagonistic traits reflecting primarily affective versus behavioral versus cognitive aspects.

As theoretically derived, affective traits (callousness and hostility) were best accounted for by BF-AG, whereas primarily behavioral antagonistic traits (manipulativeness, deceitfulness, and attention-seeking) were more

strongly related to HH. In turn, primarily cognitive traits (grandiosity and suspiciousness) were best covered by D. Moreover, D was the only broad trait that explained a similar proportion of variance in affective, behavioral, and cognitive antagonistic traits alike. Arguably, D thus offers the most balanced representation of the full spectrum of antagonistic traits.

As an aside, none of the three candidates (BF-AG, HH, D) were exclusively related to one group of antagonistic traits (specifically those hypothesized in H1), but to others as well. This is to be expected, as neither of the candidates, nor the antagonistic traits, are exclusively affective, behavioral, or cognitive in content. By definition, all (broad) traits cover these aspects to some degree (Funder, 2001). Nonetheless, results clearly confirm that BF-AG, HH, and D reflect (groups of) antagonistic traits to varying degrees and indeed in line with their theoretical definitions.

Moreover, the present findings align with prior research. The high correlations of BF-AG with especially the primarily affective antagonistic traits has been noted before (Quilty et al., 2013; Watson et al., 2013). Also, the stronger relation of the primarily behavioral antagonistic traits, and thus behavior related to exploitation of others, with HH over BF-AG is in agreement with studies revealing HH as a stronger predictor of prosociality and (non-) exploitation than BF-AG (for a recent meta-analysis, see Thielmann et al., 2020). Furthermore, the present results indicating D to be most strongly related to the cognitive traits aligns with recent findings showing that D is a superior predictor (over and above BF-AG or HH) of justifying beliefs for socially aversive behavior (Hilbig, Moshagen, Thielmann, et al., 2022; Horsten et al., 2021; Moshagen, Zettler, Horsten, et al., 2020). Finally, the incremental effects of D over and above BF-AG in deceitfulness and manipulation (i.e., dishonesty and cheating) and the incremental effects of D over and above HH in hostility and callousness (i.e., aggression and insensitivity) are in line with prior findings (Moshagen et al., 2018).

Another noteworthy finding is that specifically those antagonistic traits which show relatively strong empirical overlap with all other antagonistic traits – that is manipulation, deceitfulness, callousness, and grandiosity – are also best explained by HH, AG, and D overall. By comparison, antagonistic traits involving less overlap with the other antagonistic traits overall and thus those with more specific variance – that is, suspiciousness, hostility, and attention-seeking – are accounted for less well by AG, HH, and D. This finding is generally aligned with the theoretical reasoning behind D that a common cause determines the *communalities* of aversive traits and that D will thus more strongly account for variance in exactly those aversive traits that entail more overlap with other aversive traits (Zettler et al., 2021).

As a clear exception, attention-seeking was the only antagonistic trait not *substantially* related to any of the three broad candidate traits considered here. In fact, attention-seeking only showed weak to moderate correlations with the candidates overall. As such, whereas it seems reasonable that such behavior can become *maladaptive* at some level, it seems questionable whether seeking attention is *antagonistic* per se. In fact, to gain others admiration and attention, one may be just as likely to show socially acceptable behavior rather than engaging in behavior which is at odds with others' interests. In line with this reasoning, the most recent conceptualization of a higher-order antagonistic psychopathology domain does not include attention-seeking as a lower level trait (Sleep et al., 2020). However, the conceptual decision whether a trait ought to be considered antagonistic is actually challenging in some cases. One notable problem in this respect is that Antagonism, as a domain, is defined so vaguely ("odds with others"). As a consequence, no clear-cut criteria exist by which one may determine whether a trait qualifies as antagonistic. For example, trait eccentricity can arguably place an individual at odds with others, e.g., when unusual experiences and beliefs are perceived as dangerous. However, eccentricity is not commonly considered antagonistic. Vice versa, attention-seeking sometimesis considered antagonistic, even though it need not, per se, places an individual at odds with others. Therefore, future work should investigate which maladaptive traits overlap conceptually – i.e., with respect to a specific definition of what makes a trait, say, antagonistic – rather than grouping traits by mere empirical overlap. Nevertheless, the correlation of attention-seeking and HH was still moderate in size and the highest among all candidates, thus supporting the prediction that primarily behavioral antagonistic traits are best reflected in HH.

Comparatively speaking, D yielded the most balanced coverage of explained variance in the antagonistic traits and, more importantly, added incremental predictive variance in those (mostly cognitive) trait aspects missed by BF-AG or HH. Crucially, this effect was not attributable to the larger number of items used to measure D as compared to HH or BF-AG, as the supplementary analyses using a reduced item set for D showed. Interestingly, there was rarely a case in which BF-AG explained meaningful variance over and above D in behavioral or cognitive traits, or in which HH explained meaningful variance over and above D for affective or cognitive traits (see Supplement 3 on the OSF for details). Overall, then, D not only explained more variance over and above BF-AG in behavioral and cognitive traits, and over and above HH in affective and cognitive traits, but D did also not miss out on relevant trait content of antagonistic traits that would best be covered by BF-AG or HH, respectively.[5]

4.1 | Implications

Taken together, the present findings imply that equating Antagonism exclusively with low BF-AG, as has been suggested at times (Kotov et al., 2017; Sleep et al., 2020; Widiger et al., 2019; Wright et al., 2012), is an oversimplification that comes with a primarily affective and much less behavioral or cognitive understanding of Antagonism. Similarly, equating Antagonism exclusively with low HH – although not as explicitly suggested so far – would imply a mostly behavioral understanding of Antagonism. Our findings thus echo the position that – in addition to BF-AG or HH – antagonistic traits involve features related to other trait dimensions. Indeed, Antagonism as identified by Widiger and Simonsen (2005) and conceptualized in the PID-5 (Krueger et al., 2012; Krueger & Markon, 2014) was never explicitly designed to be the low pole of only *one* basic trait dimension in isolation. Instead, to adequately represent antagonistic traits (and their shared aspect of Antagonism), a blend of multiple basic personality dimensions might be necessary. This aligns with our finding that D offered the most balanced representation of the seven antagonistic traits under scrutiny: In terms of basic personality models, D is best understood as a blend of several dimensions (BF: AG, Conscientiousness and to a smaller extent Neuroticism and Extraversion; HEXACO: HH, AG, Conscientiousness, and to a smaller extent Emotionality) and cannot be equated or reduced to any one dimension, not even those it is most strongly associated with, that is, BF-AG or HH (Horsten et al., 2021; Moshagen et al., 2018; Moshagen, Zettler, Horsten, et al., 2020).

Importantly, despite the promise of D to account for individual variation in antagonistic traits, the present findings also do not warrant the conclusion that D can be equated with Antagonism. Shared variance of 26% between D and the antagonistic traits (on average) is too low to consider the two constructs equivalent. In fact, aspects covered by D and antagonistic traits differ in relevant ways. First, D encompasses sadistic and spiteful tendencies, for which no antagonistic trait (or maladaptive trait in general in the DSM-5) has been suggested (Hilbig, Thielmann, et al., 2021). Second, tendencies described by D, although aversive by definition, are not necessarily maladaptive or strictly psychopathological by definition. Instead, aversive traits "do place individuals at risk for maladjustment, including psychopathology" (Thomaes et al., 2017, p. 836). Third, as argued before, behavior aimed at attention-seeking is partially incompatible with the disutility aspect of D because gaining others' attention and admiration may arguably involve positive behavior toward others, rather than causing them harm. In turn, as discussed before, one might question whether attention-seeking ought to be considered antagonistic per se, with more recent approaches to the structure of psychopathological traits actually excluding attention-seeking from the Antagonism domain (see Sleep et al., 2020).

4.2 | Limitations and directions for future research

Some limitations of the current research need to be acknowledged. First, reliance on the PID-5 inherently means measuring endorsement of highly undesirable, low-base-rate statements, all of which are scored in the direction of greater personality pathology (see Supplement 4 on the OSF). Thus, a substantial proportion of PID-5 scale variance will be attributable to trait-unrelated individual differences in the tendency to endorse such items (Ashton et al., 2017; Furnham, 1986). Moreover, we relied on the 100-item version of the PID-5 rather than using the full (220-item) inventory (Krueger et al., 2012), thus potentially limiting the breadth of trait content covered. However, the 100-item version shows almost the same reliability and validity on the facet level (Maples et al., 2015), suggesting at the very least that the current findings ought to replicate when using more comprehensive measures of antagonistic traits. Second and similarly, HH was measured with considerably fewer items than BF-AG and D and, thus, potentially less comprehensively. Importantly, however, the additional analysis with reduced item sets for BF-AG and D essentially yielded the same conclusions, thus ruling out that the comparison was generally biased against HH. Third, although the sample was large and diverse, we relied on a community rather than a clinical sample. Although the very notion underlying the shift toward dimensional models of psychopathology strictly implies that antagonistic traits are not restricted to some narrow clinical populations, the most extreme ends of the spectrum were likely underrepresented in our sample. Finally, one may argue that different operationalizations of BF-AG might have produced other findings. However, a particular strength of the present study is the reliance on a broad representation of BF-AG measured by multiple established AG scales and modeled as their common core. Thus, we maintain that our findings shed light on the role of a particularly broad AG dimension that is not limited to one specific operationalization of BF-AG. To our knowledge, this is the first study to do so.

However, we would also like to point out again that we implemented a broad operationalization of BF-AG that subsumes several particularly common conceptualizations and operationalizations. As a consequence, it does not perfectly conform to any one specific conceptualization and operationalization (which tend to differ, e.g., Hilbig et al., 2016; Thielmann et al., 2021) nor is

it equivalent to cross-dimension variants such as the construct suggested by Crowe et al. (2018), which comprises aspects of HH and the interstitial Altruism facet of the HEXACO model (see also Vize & Lynam, 2021; Vize et al., 2021). The latter, although also labeled Agreeableness, is actually incompatible with AG as a largely orthogonal factor within the BF or Five-Factor Model (Hilbig et al., 2021). In any case, we explicitly limit our conclusions to a construct that is meant to be one of five largely orthogonal factors (within the BF) and represented by the common variance of corresponding scales.

4.3 | Conclusion

As essentially intended by shifting toward a dimensional understanding of personality psychopathology, antagonistic traits are reflected in broad traits describing socially aversive behavior to a notable degree, thus supporting the general view that pathological traits can be understood as extreme levels of dimensions within the basic personality space. However, simply equating antagonistic traits with any one *single* basic dimension such as BF-AG is an oversimplification. Indeed, even D – which can be understood as a blend of dimensions such as BF-AG and HH (and others) – is not equivalent to Antagonism although it does offer the most comprehensive and balanced basic trait representation of all antagonistic traits so far.

ACKNOWLEDGMENTS

This research was supported by a grant from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – GRK 2277 "Statistical Modeling in Psychology". Open access funding enabled and organized by ProjektDEAL.

ETHICS APPROVAL

The present study was conducted in full accordance with the Ethical Guidelines of the American Psychological Association (APA). The study was run based on approval from the local ethics committee.

AUTHOR CONTRIBUTIONS

All authors contributed to and jointly wrote the present manuscirpt. Writing was headed by David D. Scholz and Benjamin E. Hilbig. Benjamin E. Hilbig and Isabel Thielmann collected the data. David D. Scholz analyzed the data together with Morten Moshagen and Benjamin E. Hilbig.

ORCID

David D. Scholz https://orcid.org/0000-0003-1491-5247
Benjamin E. Hilbig https://orcid.org/0000-0002-1470-1882
Isabel Thielmann https://orcid.org/0000-0002-9071-5709
Morten Moshagen https://orcid.org/0000-0002-2929-7288
Ingo Zettler https://orcid.org/0000-0001-6140-7160

ENDNOTES

[1] Note that definitions of BF-AG vary to some degree (Graziano & Tobin, 2017). For the current work, BF-AG is conceptualized as a largely orthogonal factor within the BF (Goldberg, 1992) and operationally defined by what is commonly measured by widely used BF-AG scales, e.g. the International Personality Item Pool Big Five scales (Goldberg, 1992); the NEO Five Factor Inventory (Costa & McCrae, 1992; McCrae & Costa, 2004); the Big Five Aspects Scales (DeYoung et al., 2007); and the Big Five Inventory-2 (Soto & John, 2017).

[2] The only antagonistic trait that is arguably beyond the scope of D is attention-seeking: Behavior shown to garner others' attention and admiration is not necessarily aversive and therefore beyond the definition of D.

[3] The IPIP-50 actually comprises 10 items for BF-AG. However, two of these items are identical to the BFAS-AG scale. Thus, we omitted these two items from the IPIP-AG scale.

[4] For example, a post-hoc power analysis for a global hypothesis test as described by Jobst et al. (2021) for the bi-factor model of D (df =2275) or BF-AG (df =1234) revealed an extremely high power of greater than 99.9% for detecting even negligible model misspecifications, i.e., *RMSEA* of 0.010, at an α-level of 0.05.

[5] Note that D correlates as strongly as BF-AG and HH with *all* other PID-5 traits, with median r (D) = 0.25, r (BF-AG) = −0.27, r (HH) = −0.22. Thus, the relatively high correlations of D with the antagonistic traits cannot be simply due to a similar framing of items measuring D and psychopathological traits in general.

REFERENCES

Al-Dajani, N., Gralnick, T. M., Bagby, R. M. (2016). A Psychometric review of the personality inventory for DSM–5 (PID–5): Current status and future directions. *Journal of Personality Assessment*, *98*(1), 62–81. https://doi.org/10.1080/00223891.2015.1107572

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

Ashton, M. C., de Vries, R. E., & Lee, K. (2017). Trait variance and response style variance in the scales of the personality inventory for DSM-5 (PID-5). *Journal of Personality Assessment*, *99*(2), 192–203. https://doi.org/10.1080/00223891.2016.1208210

Ashton, M. C., & Lee, K. (2005). Honesty-humility, the big five, and the five-factor model. *Journal of Personality*, *73*(5), 1321–1353. https://doi.org/10.1111/j.1467-6494.2005.00351.x

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. (2008). The prediction of Honesty–Humility-related criteria by the HEXACO and Five-Factor

Models of personality. *Journal of Research in Personality*, *42*(5), 1216–1228. https://doi.org/10.1016/j.jrp.2008.03.006

Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment*, *91*(4), 340–345. https://doi.org/10.1080/00223890902935878

Ashton, M. C., & Lee, K. (2020). Objections to the HEXACO model of personality structure—And why those objections fail. *European Journal of Personality*, *34*(4), 492–510. https://doi.org/10.1002/per.2242

Ashton, M. C., Lee, K., & de Vries, R. E. (2014). The HEXACO Honesty-Humility, Agreeableness, and Emotionality factors. *Personality and Social Psychology Review*, *18*(2), 139–152. https://doi.org/10.1177/1088868314523838

Ashton, M. C., Lee, K., Perugini, M., Szarota, P., de Vries, R. E., Di Blas, L., Boies, K., & 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

Bader, M., Hartung, J., Hilbig, B. E., Zettler, I., Moshagen, M., & Wilhelm, O. (2021). Themes of the dark core of personality. *Psychological Assessment*, *33*(6), 511–525.

Bader, M., & Moshagen, M. (2021). No probifactor model fit index bias, but a propensity toward selecting the best model. *Journal of Abnormal Psychology*. https://doi.org/10.1037/abn0000685

Bo, S., Bach, B., Mortensen, E. L., & Simonsen, E. (2016). Reliability and hierarchical structure of DSM-5 pathological traits in a Danish mixed sample. *Journal of Personality Disorders*, *30*(1), 112–129. https://doi.org/10.1521/pedi_2015_29_187

Borkenau, P., & Ostendorf, F. (2008). *NEO-Fünf-Faktoren Inventar nach Costa und McCrae (NEO-FFI): Manual (2., neu normierte und vollständig überarbeitete Auflage)*. Hogrefe.

Calvete, E. (2008). Justification of violence and grandiosity schemas as predictors of antisocial behavior in adolescents. *Journal of Abnormal Child Psychology*, *36*(7), 1083–1095. https://doi.org/10.1007/s10802-008-9229-5

Calvete, E., & Orue, I. (2013). Cognitive mechanisms of the transmission of violence: Exploring gender differences among adolescents exposed to family violence. *Journal of Family Violence*, *28*(1), 73–84. https://doi.org/10.1007/s10896-012-9472-y

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

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

Crego, C., Gore, W. L., Rojas, S. L., & Widiger, T. A. (2015). The discriminant (and convergent) validity of the Personality Inventory for DSM-5. *Personality Disorders: Theory, Research, and Treatment*, *6*(4), 321–335. https://doi.org/10.1037/per0000118

Crowe, M. L., Lynam, D. R., & Miller, J. D. (2018). Uncovering the structure of agreeableness from self-report measures. *Journal of Personality*, *86*(5), 771–787. https://doi.org/10.1111/jopy.12358

Danner, D., Rammstedt, B., Bluemke, M., Lechner, C., Berres, S., Knopf, T., Soto, C. J., & John, O. P. (2019). Das Big Five Inventar 2. *Diagnostica*, *65*(3), 121–132. https://doi.org/10.1026/0012-1924/a000218

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

Funder, D. C. (2001). Personality. *Annual Review of Psychology*, *52*, 197–221. https://doi.org/10.1146/annurev.psych.52.1.197

Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences*, *7*(3), 385–400. https://doi.org/10.1016/0191-8869(86)90014-0

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, *4*, 26–42. https://doi.org/10.1037/1040-3590.4.1.26

Graziano, W. G., & Tobin, R. M. (2009). Agreeableness. In M. R. Leary & R. H. Hoyle (Eds.), *Handbook of individual differences in social behavior* (pp. 46–61). The Guilford Press.

Graziano, W. G., & Tobin, R. M. (2017). Agreeableness and the five factor model. In T. A. Widiger (Ed.), *Oxford handbook of the five factor model* (pp. 105–132). Oxford University Press.

Heck, D., Thielmann, I., Moshagen, M., & Hilbig, B. E. (2018). Who lies? A large-scale reanalysis linking basic personality traits to unethical decision making. *Judgment and Decision Making*, *13*(4), 356–371.

Hilbig, B. E., Moshagen, M., Horsten, L. K., & Zettler, I. (2021). Agreeableness is dead. Long live Agreeableness? Reply to Vize and Lynam. *Journal of Research in Personality*, *91*, 104074.

Hilbig, B. E., Moshagen, M., Thielmann, I., & Zettler, I. (2022). *Making rights from wrongs: The crucial role of beliefs and justifications for the expression of aversive personality*. Unpublished manuscript.

Hilbig, B. E., Moshagen, M., & Zettler, I. (2016). Prediction consistency: A test of the equivalence assumption across different indicators of the same construct. *European Journal of Personality*, *30*(6), 637–647. https://doi.org/10.1002/per.2085

Hilbig, B. E., Thielmann, I., Klein, S. A., Moshagen, M., & Zettler, I. (2021). The dark core of personality and socially aversive psychopathology. *Journal of Personality*, *89*(2), 216–227. https://doi.org/10.1111/jopy.12577

Hopwood, C. J., Kotov, R., Krueger, R. F., Watson, D., Widiger, T. A., Althoff, R. R., Ansell, E. B., Bach, B. O., Michael Bagby, R., Blais, M. A., Bornovalova, M. A., Chmielewski, M., Cicero, D. C., Conway, C., De Clercq, B., De Fruyt, F., Docherty, A. R., Eaton, N. R., Edens, J. F., … Zimmermann, J. (2018). The time has come for dimensional personality disorder diagnosis. *Personality and Mental Health*, *12*(1), 82–86. https://doi.org/10.1002/pmh.1408

Horsten, L. K., Moshagen, M., Zettler, I., & Hilbig, B. E. (2021). Theoretical and empirical dissociations between the Dark Factor of Personality and low Honesty-Humility. *Journal of Research in Personality*, *95*, 104154. https://doi.org/10.1016/j.jrp.2021.104154

Jobst, L., Bader, M., & Moshagen, M. (2021). A tutorial on assessing statistical power and determining sample size for structural equation models. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000423

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., … Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to

traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454–477. https://doi.org/10.1037/abn0000258

Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2012). Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychological Medicine*, *42*(9), 1879–1890. https://doi.org/10.1017/S0033291711002674

Krueger, R. F., & Hobbs, K. A. (2020). An overview of the DSM-5 alternative model of personality disorders. *Psychopathology*, *53*(3–4), 126–132. https://doi.org/10.1159/000508538

Krueger, R. F., & Markon, K. E. (2014). The role of the DSM-5 personality trait model in moving toward a quantitative and empirically based approach to classifying personality and psychopathology. *Annual Review of Clinical Psychology*, *10*, 477–501. https://doi.org/10.1146/annurev-clinpsy-032813-153732

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

Maples, J. L., Carter, N. T., Few, L. R., Crego, C., Gore, W. L., Samuel, D. B., Williamson, R. L., Lynam, D. R., Widiger, T. A., Markon, K. E., Krueger, R. F., & Miller, J. D. (2015). Testing whether the DSM-5 personality disorder trait model can be measured with a reduced set of items: An item response theory investigation of the Personality Inventory for DSM-5. *Psychological Assessment*, *27*(4), 1195–1210. https://doi.org/10.1037/pas0000120

McCrae, R. R., & Costa, P. T. (2004). A contemplated revision of the NEO Five-Factor Inventory. *Personality and Individual Differences*, *36*(3), 587–596. https://doi.org/10.1016/S0191-8869(03)00118-1

McCrae, R. R., & Costa, P. T. (2008). The five-factor theory of personality. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research*. The Guilford Press.

Moshagen, M., & Auerswald, M. (2018). On congruence and incongruence of measures of fit in structural equation modeling. *Psychological Methods*, *23*(2), 318–336. https://doi.org/10.1037/met0000122

Moshagen, M., Hilbig, B. E., & Zettler, I. (2014). Faktorenstruktur, psychometrische Eigenschaften und Messinvarianz der deutschsprachigen Version des 60-Item HEXACO Persönlichkeitsinventars. *Diagnostica*, *60*(2), 86–97. https://doi.org/10.1026/0012-1924/a000112

Moshagen, M., Hilbig, B. E., & Zettler, I. (2018). The dark core of personality. *Psychological Review*, *125*(5), 656–688. https://doi.org/10.1037/rev0000111

Moshagen, M., Zettler, I., & Hilbig, B. E. (2020). Measuring the dark core of personality. *Psychological Assessment*, *32*(2), 182–196. https://doi.org/10.1037/pas0000778

Moshagen, M., Zettler, I., Horsten, L. K., & Hilbig, B. E. (2020). Agreeableness and the common core of dark traits are functionally different constructs. *Journal of Research in Personality*, *87*, 103986. https://doi.org/10.1016/j.jrp.2020.103986

Mussel, P., & Paelecke, M. (2018). BFAS-G – Big Five Aspect Scales – German. *PsyArXiv*. https://doi.org/10.23668/psycharchives.2341

Quilty, L. C., Ayearst, L., Chmielewski, M., Pollock, B. G., & Bagby, R. M. (2013). The psychometric properties of the personality inventory for DSM-5 in an APA DSM-5 field trial sample. *Assessment*, *20*(3), 362–369. https://doi.org/10.1177/1073191113486183

R Core Team. (2021). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing https://www.r-project.org/

Reise, S. P. (2012). Invited paper: The rediscovery of Bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667–696. https://doi.org/10.1080/00273171.2012.715555

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*(2), 137–150. https://doi.org/10.1037/met0000045

Rosseel, Y., Jorgensen, T. D., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, E., Hallquist, M., Rhemtulla, M., Katsikatsou, M., Barendse, M., Scharf, F., & Du, H. (2021). *Package 'lavaan'* [Computer software]. https://cran.r-project.org/web/packages/lavaan/lavaan.pdf

Samuel, D. B., & Widiger, T. A. (2008). A meta-analytic review of the relationships between the five-factor model and DSM-IV-TR personality disorders: A facet level analysis. *Clinical Psychology Review*, *28*(8), 1326–1342. https://doi.org/10.1016/j.cpr.2008.07.002

Saucier, G. (2002). Orthogonal markers for orthogonal factors: The case of the Big Five. *Journal of Research in Personality*, *36*(1), 1–31. https://doi.org/10.1006/jrpe.2001.2335

Sellbom, M., & Tellegen, A. (2019). Factor analysis in psychological assessment research: Common pitfalls and recommendations. *Psychological Assessment*, *31*(12), 1428–1441. https://doi.org/10.1037/pas0000623

Sleep, C. E., Crowe, M. L., Carter, N. T., Lynam, D. R., & Miller, J. D. (2020). Uncovering the structure of antagonism. *Personality Disorders: Theory, Research, and Treatment*, *12*(4), 300–311. https://doi.org/10.1037/per0000416

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

ten Berge, J. M., & Sočan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika*, *69*(4), 613–625. https://doi.org/10.1007/BF02289858

Thielmann, I., Moshagen, M., Hilbig, B., & Zettler, I. (2021). On the comparability of basic personality models: Meta-analytic correspondence, scope, and orthogonality of the Big Five and HEXACO dimensions. *European Journal of Personality*. Advance online publication. https://doi.org/10.1177/08902070211026793

Thielmann, I., Spadaro, G., & Balliet, D. (2020). Personality and prosocial behavior: A theoretical framework and meta-analysis. *Psychological Bulletin*, *146*(1), 30–90. https://doi.org/10.1037/bul0000217

Thomaes, S., Brummelman, E., Miller, J. D., & Lilienfeld, S. O. (2017). The dark personality and psychopathology: Toward a brighter future. *Journal of Abnormal Psychology*, *126*(7), 835–842. https://doi.org/10.1037/abn0000305

Tobin, R. M., Graziano, W. G., Vanman, E. J., & Tassinary, L. G. (2000). Personality, emotional experience, and efforts to control emotions. *Journal of Personality and Social Psychology*, *79*(4), 656–669. https://doi.org/10.1037//0022-3514.79.4.656

Treiber, L., Thunsdorff, C., Schmitt, M. & Schreiber, W. H. (2013). Testing the German 300-Item-IPIP-Scale. The translation and

convergent validation of the 300-Item-IPIP-scale with its well-known counterpart. Poster presented at the World Conference on Personality.

Vize, C. E., & Lynam, D. R. (2021). On the importance of the assessment and conceptualization of Agreeableness: A commentary on "Agreeableness and the common core of dark traits are functionally different constructs". *Journal of Research in Personality*, *90*, 104059. https://doi.org/10.1016/j.jrp.2020.104059

Vize, C. E., Miller, J. D., & Lynam, D. R. (2021). Examining the conceptual and empirical distinctiveness of Agreeableness and "dark" personality items. *Journal of Personality*, *89*(3), 594–612. https://doi.org/10.1111/jopy.12601

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*(1), 192–196. https://doi.org/10.3758/BF03206482

Watson, D., Stasik, S. M., Ro, E., & Clark, L. A. (2013). Integrating normal and pathological personality: Relating the DSM-5 trait-dimensional model to general traits of personality. *Assessment*, *20*(3), 312–326. https://doi.org/10.1177/1073191113485810

Watters, C. A., & Bagby, R. M. (2018). A meta-analysis of the five-factor internal structure of the Personality Inventory for DSM-5. *Psychological Assessment*, *30*(9), 1255–1260. https://doi.org/10.1037/pas0000605

Widiger, T. A., Sellbom, M., Chmielewski, M., Clark, L. A., DeYoung, C. G., Kotov, R., Krueger, R. F., Lynam, D. R., Miller, J. D., Mullins-Sweatt, S., Samuel, D. B., South, S. C., Tackett, J. L., Thomas, K. M., Watson, D., & Wright, A. G. C. (2019). Personality in a hierarchical model of psychopathology. *Clinical Psychological Science*, *7*(1), 77–92. https://doi.org/10.1177/2167702618797105

Widiger, T. A., & Simonsen, E. M. D. (2005). Alternative dimensional models of personality disorder: Finding a common ground. *Journal of Personality Disorders*, *19*(2), 110–130. https://doi.org/10.1521/pedi.19.2.110.62628

Wright, A. G. C., Thomas, K. M., Hopwood, C. J., Markon, K. E., Pincus, A. L., & Krueger, R. F. (2012). The hierarchical structure of DSM-5 pathological personality traits. *Journal of Abnormal Psychology*, *121*(4), 951–957. https://doi.org/10.1037/a0027669

Wu, H., Fai Cheung, S., & On Leung, S. (2020). Simple use of BIC to assess model selection uncertainty: An illustration using mediation and moderation models. *Multivariate Behavioral Research*, *55*(1), 1–16. https://doi.org/10.1080/00273171.2019.1574546

Zettler, I., Moshagen, M., & Hilbig, B. E. (2021). Stability and change: The dark factor of personality shapes dark traits. *Social Psychological and Personality Science*, *12*(6), 974–983. https://doi.org/10.1177/1948550620953288

Zimmermann, J., Altenstein, D., Krieger, T., Holtforth, M. G., Pretsch, J., Alexopoulos, J., Spitzer, C., Benecke, C., Krueger, R. F., Markon, K. E., & Leising, D. (2014). The structure and correlates of self-reported DSM-5 maladaptive personality traits: Findings from two German-speaking samples. *Journal of Personality Disorders*, *28*(4), 518–540. https://doi.org/10.1521/pedi_2014_28_130

How to cite this article: Scholz, D. D., Hilbig, B. E., Thielmann, I., Moshagen, M., & Zettler, I. (2022). Beyond (low) Agreeableness: Toward a more comprehensive understanding of antagonistic psychopathology. *Journal of Personality*, 00, 1–15. https://doi.org/10.1111/jopy.12708

This work is scheduled to appear in

Journal of Personality Disorders

Theoretical and empirical integration of 'dark' traits and socially aversive personality psychopathology

David D. Scholz[1], Johannes Zimmermann[2], Morten Moshagen[3], Ingo Zettler[4,5], and Benjamin E. Hilbig[1]

[1] Department of Psychology, University of Kaiserslautern-Landau (RPTU), Germany

[2] Department of Psychology, University of Kassel, Germany

[3] Institute of Psychology and Education, University of Ulm, Germany

[4] Department of Psychology, University of Copenhagen, Denmark

[5] Copenhagen Center for Social Data Science (SODAS), University of Copenhagen, Denmark

Author Note

David D. Scholz https://orcid.org/0000-0003-1491-5247

Johannes Zimmermann https://orcid.org/0000-0001-6975-2356

Morten Moshagen https://orcid.org/0000-0002-2929-7288

Ingo Zettler https://orcid.org/0000-0001-6140-7160

Benjamin E. Hilbig https://orcid.org/0000-0002-1470-1882

Data used for the pilot study stem from the *Prosocial Personality Project* (PPP) and are, in part, a reanalysis of data used by both Scholz et al. (2022) and Hilbig et al. (2023). For details on that as well as access to all materials, data, and analysis scripts, please see the files accompanying this manuscript on the Open Science Framework (OSF; https://osf.io/uc38x/?view_only=16fe0ecd308941858b6553d1b6d83770).

The authors declare that there were no conflicts of interest with respect to the authorship or the publication of this article. This research was supported by a grant from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - GRK 2277 "Statistical Modeling in Psychology".

Correspondence concerning this article should be addressed to David D. Scholz, Department of Psychology, RPTU Kaiserslautern-Landau, Campus Landau, K Building, 1st floor, Fortstraße 7, D-76829 Landau, Germany. Email: david.scholz@rptu.de

Abstract

Conceptual work integrating constructs from mainstream personality research (especially so-called “dark” traits) and clinical psychopathology research has been limited. Herein, we propose *all* socially and/or ethically aversive traits as “flavored” manifestation of the D-Factor of Personality (D). We argue that the D-framework provides the commonality of all aversive traits, including the aversive traits from the Alternative DSM-5 Model for Personality Disorders (AMPD), a more thorough theoretical foundation. Moreover, D covers aspects which are not captured by any of the aversive AMPD traits directly (e.g., Greed), thus offering indications for possible expansions to the AMPD. We tested our predictions in two online studies (N = 1,781 and N = 2,006) using quota-representative samples of the German population regarding age and gender. Twelve aversive traits from mainstream personality research and eight aversive AMPD traits were assed together with consequential behavior in an economic game. Analyses via structural equation modelling overall confirmed predictions.

Keywords: Alternative DSM-5 Model for Personality Disorders, Antagonism, Dark Factor of Personality, Maladaptive Traits, Psychopathology, Dark Traits, Dark Triad

Theoretical and empirical integration of 'dark' traits and socially aversive personality psychopathology.

Socially and/or ethically aversive tendencies constitute a notable challenge for society. Whereas some individuals with such tendencies may seek clinical treatment or even end up in confinement, others seem to find socially acceptable ways of acting on these tendencies and may even flourish in everyday life (Paulhus, 2014) although still causing harm for others and society at large. In mainstream personality research, such tendencies are often referred to as "dark" traits[1], most prominently the *Dark Triad* (i.e., Narcissism, Psychopathy, and Machiavellianism; Paulhus & Williams, 2002), although many more socially and/or ethically aversive traits have been suggested, e.g., Greed (Marcus & Zeigler-Hill, 2015), Moral Disengagement (Moore et al., 2012), or Sadism (Paulhus, 2014).

In parallel and somewhat independently from mainstream personality research, dimensional trait models for personality disorders have been established in the field of clinical personality psychopathology research, as exemplified by the *Alternative Model for Personality Disorders* (AMPD) in Section III of the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5; American Psychiatric Association, 2013). The AMPD covers a fixed set of 25 so-called maladaptive trait facets. To taxonomically describe co-occurrences of these traits in a parsimonious manner, they are assigned to five higher-order dimensions (Krueger & Hobbs, 2020; Zimmermann et al., 2019). With respect to socially and/or ethically aversive tendencies, the traits subsumed under the *Antagonism* domain in the AMPD are of prime relevance: Attention Seeking, Callousness, Deceitfulness, Grandiosity, Hostility[2], and Manipulativeness (American Psychiatric Association, 2013;

[1] Note that the term „dark" is somewhat vague and, some have argued, potentially stigmatizing (Thomaes et al. 2017). In essence, "dark" traits are socially and/or ethically aversive traits (Thomaes et al. 2017; Moshagen et al., 2018). In the following, we will maintain the term "dark" when it is important to unambiguously link our reasoning or findings to previous research such as the *Dark Triad*. Otherwise, we will use the term *aversive traits from mainstream personality research* or *aversive MPR traits* for short.

[2] In the AMPD, Hostility is conceptualized as an interstitial facet, assigned to both Antagonism and Negative Affectivity. Recent meta-analytic research suggests that it primarily loads on Antagonism (Watter et al. 2018; Clark & Watson, 2022).

Watters & Bagby, 2018). Similar notions can be found in the eleventh edition of the International Classification of Diseases (ICD-11; World Health Organization, 2023) with its *Dissociality* domain, or the Hierarchical Taxonomy of Psychopathology (HiTOP; Kotov et al., 2017) with the *Antagonistic Externalizing* spectrum.

Whereas some socially and/or ethically aversive traits from mainstream personality research (herein referred to as *aversive MPR traits* for short) were originally adopted from clinical or forensic psychopathology research, e.g., Narcissism or Psychopathy (Furnham et al., 2013), there is only sporadic empirical work connecting the multitude of aversive MPR and aversive traits from clinical models of personality structure, let alone much theoretical work providing a unifying account of why and how such traits may be related (Hilbig et al., 2020; Miller & Campbell, 2008; Thomaes et al., 2017). Clearly, it is highly unsatisfactory that two subdisciplines within psychology sharing the goal of explaining and predicting socially and/or ethically aversive behavior each rely on largely distinct sets of multiple constructs with limited empirical and still less theoretical integration.

Thus, we herein argue that and test whether the socially and/or ethically aversive traits from each field – mainstream personality research (commonly referred to as "dark" traits) and clinical personality psychopathology research (herein exemplified through the AMPD and thus referred to as *aversive AMPD traits*) – can be integrated within a common theoretical framework, specifying the commonalities and differences of *all* aversive traits. Based on this common framework, we further show that potential "blank spots" of aversiveness within the AMPD (Clark & Watson, 2022; Hilbig et al., 2020) – i.e., cases of aversive MPR traits for which no corresponding trait has been specified within the AMPD – are covered at least indirectly.

The need for a common framework for all aversive traits

Existing evidence supports the idea that aversive traits from mainstream personality and clinical psychopathology research share a common basis. Most prominently, there is notable evidence for a substantial association between Antagonism (as per the AMPD) and (low) Agreeableness (Crowe et al., 2018; Lynam & Miller, 2019; Sleep et al., 2020), or low Honesty-

Humility as per the HEXACO-model (Ashton et al., 2012). Moreover, it has been shown that different conceptualizations of Antagonism, each assigning a different set of traits (and thus facets) to the higher-order Antagonism domain, account for the shared variance of the Dark Triad components (Rose et al., 2023; Sleep et al., 2017; Somma et al., 2020; Vize, Collison & Lynam, 2020; Vize, Collison, Miller & Lynam, 2020). As such, there is diverse evidence for notable overlap between aversive MPR and AMPD traits.

Whereas such findings imply that the quest for a common theoretical basis is worthwhile, two broad limitations have remained so far: First, from a theoretical and taxonomical point of view, it is unsatisfactory that different conceptualizations of Antagonism co-exist without theoretical integration, each assigning different sets of traits to the domain. As a consequence, it remains vague what exactly constitutes the higher-order domain Antagonism, and, as such, the shared aspect of the traits assigned to it. Indeed, different plausible suggestions beyond the AMPD have been put forward, e.g., Attention Seeking not being part of Antagonism (Sleep et al., 2020; Widiger, 2011), or Suspiciousness (Negative Affectivity in the AMPD) being part of Antagonism (Krueger et al., 2021; Sleep et al., 2020; Widiger & Simonsen, 2005). Similarly, interstitial traits in the AMPD – e.g., Hostility (which is interstitial between Antagonism and Negative Affectivity) – essentially leave it to idiosyncrasies where they are located taxonomically (Clark & Watson, 2022; Watters & Bagby, 2018), thus leaving the theoretical nature of Antagonism volatile and dependent on such idiosyncrasies. Consequently, theoretical integration with other research areas, e.g., aversive MPR traits, is close to impossible.

Second, extant knowledge on the empirical overlap between aversive MPR and AMPD traits has been limited to the Dark Triad (e.g., Grigoras & Wille, 2017; Sleep et al., 2017; Vize, Collison, Miller & Lynam, 2020; Rose et al., in press), occasionally plus one additional trait at most (Russell & King, 2017; Somma et al., 2020). Thus, the majority of aversive MPR traits have never been considered – let alone in combination. This is particularly problematic because the Dark Triad actually provides a relatively poor approximation of the common core shared by all aversive MPR

traits (Horsten, Moshagen et al., 2023). As such, the known overlap between the Dark Triad and Antagonism and/or low Agreeableness is, at best, necessary but definitely not sufficient for the conclusion that all aversive MPR and AMPD traits share a common core.

To remedy these limitations and achieve theoretical integration of aversive MPR and AMPD traits, a common framework making testable assumptions about the (aversive) core shared by *any* aversive trait is needed – along with corresponding empirical tests including a broad array of aversive traits (from mainstream personality research and the AMPD alike). In the remainder of this manuscript, we offer both.

A comprehensive theoretical framework for *all* aversive traits

In mainstream personality research, it is now widely accepted that aversive traits – especially the Dark Triad components (i.e., Narcissism, Psychopathy, Machiavellianism) but also other (such as Entitlement, Greed, Sadism, or Spitefulness) – share a single common *aversive* core (Kowalski et al., 2021; Muris et al., 2017; Schreiber & Marcus, 2020; Vize, Collison, Miller & Lynam, 2020). A comprehensive definition of this common core, which Moshagen et al. (2018) labeled the *D-Factor of Personality* (D)[3], is "the general tendency to maximize one's individual utility – disregarding, accepting, or malevolently provoking disutility for others – , accompanied by beliefs that serve as justifications" (p. 657). Note that, within this definition, (dis)utility is meant in a broad sense – not limited to direct material utility, but also including psychological utilities such as pleasure. Note also the emphasis on ignoring, accepting, or even seeking disutility in others, which discriminates socially/ethically aversive traits from other socially undesirable tendencies (such as Depressivity).

The substantive definition of D and its suitability to describe the shared variance of aversive MPR traits has been studied extensively. The proposed primary features of the common core —

[3]The label "D-Factor" stems from its development within the research tradition coined as "dark" personality. As stated above, we acknowledge that in the context of clinical psychology research, the term "dark" may be stigmatizing. Nonetheless, for the sake of clarity and consistency, we use the label "D-Factor" or D for short.

indicating a) *diverse* forms of utility maximization at others' expense and b) being accompanied by a multitude of justifying belief—are directly supported by a large body of research. For example, D is a strong predictor of a variety of aversive outcomes, referring to utility maximization while either disregarding, accepting, or provoking disutility in others, both self-reported (e.g., violence, vandalism, internet trolling) and consequential behavior (e.g., dishonest, sadistic, or anti-social behavior) with specific aversive traits explaining little to no incremental variance over and above D (e.g., Bader, Hartung, et al., 2021; Bader, Horsten, et al., 2021; Hilbig et al., 2023; Moshagen et al., 2018; Moshagen, Zettler, & Hilbig, 2020; Scholz et al., 2023; Schrödter et al., 2021). Moreover, studies have shown that the common core is strongly associated with multiple beliefs that can be – and indeed are – used to justify aversive behavior (Hilbig, Moshagen, et al., 2022; Horsten, Thielmann, et al., 2023; Rudloff et al., 2022, 2023; Scholz et al., 2022; Thielmann & Hilbig, 2023; Moshagen et al., in press). In fact, for many such beliefs (e.g. Competitive Jungle Social World View, Authoritarianism, or Normlessness) D predicts incremental variance over and above common operationalizations of FFM-Agreeableness and the entire HEXACO-space (Hilbig, Moshagen, et al., 2022).

Furthermore, it has been shown that features of, or related to, Disinhibition, such as Impulsivity or Risk Taking, are beyond the scope of D, as studies indicate only weak correlations among self-reported Disinhibition and little to no association with behavioral measures (Moshagen et al., 2023; Tiwari et al., 2021). Correspondingly, the common aversive core has also been dissociated from other constructs subsuming Disinhibition that were previously suggested as representations of a common core among (some) aversive traits, e.g., a *Fast Life History Strategy* (Horsten et al., 2022).

Crucially for the present work and independent of the substantive definition or nomological net of the common core, Moshagen et al. (2018) conceptualized D within a theoretical framework, allowing for testable assumptions. Specifically, the common core is conceptualized as the disposition underlying *all* socially/ethically aversive traits. In turn, all aversive traits are understood as “flavored

manifestations" of D (Bader et al., 2022; Moshagen et al., 2018), a term borrowed from the ability domain and specifically the two-factor theory of intelligence (Carroll, 2009; Spearman, 1927). "Flavored" means that (a) each aversive trait may emphasize a certain aspect of the D-definition (e.g., provoking disutility is very prominent for Sadism, whereas justifying beliefs are particularly prominent in Narcissism) and (b) each aversive trait might entail further aspects beyond D (e.g., Disinhibition in Psychopathy; Bader et al., 2022). Importantly, any aspect that is socially and/or ethically *aversive* must be due to D. In other words, D represents the "aversive essence" (Hilbig et al., 2023) of all aversive traits[4].

Viewed through this theoretical lens, certain AMPDs traits may also be seen as a specific manifestation of D: To the extent that an AMPD trait involves aversive aspects in the sense of the D-definition (utility maximization at others' expense accompanied by justifying beliefs) it can be assumed to be a flavored manifestation of D, just like any other aversive trait in the "subclinical" realm (e.g., Dark Triad, Sadism, Greed). Thereby, the D-Framework serves as a unifying explanation for the shared variance of *all* socially and/or ethically aversive traits, no matter which research tradition they happen to stem from.

Of note, the advantage of uniting all aversive traits within one conceptual framework is largely independent of recent discussions regarding the substantive overlap of D and (low) Agreeableness (see Hilbig et al., 2021; Horsten, 2023; Horsten, Thielmann et al., 2023; Moshagen, Zettler, Horsten & Hilbig, 2020; Rose et al., 2022; Vize, Collison, Miller & Lynam, 2020; Vize & Lynam, 2021; Vize et al., 2021) and such debates are also out of the scope of the current research. We will discuss our findings in light of this debate, but for now, focus on the crucial point that the D-Framework yields the strictly testable prediction(s) that aversive AMPD traits can be seen as

[4] This does not strictly rule out that some manifestations correlate even after controlling for D. If a certain manifestation comprises similar non-aversive flavors – e.g. Psychopathy and Irresponsibility share Disinhibition – they must correlate. However, the crucial point is that the only feature shared by *all* aversive traits is D.

manifestations of one single aversive common core across all aversive traits (labeled *D* within this framework and defined as utility maximization at others' expense accompanied by justifying beliefs).

D as an underlying disposition of aversive AMPD traits

Fulfilling the most basic necessary condition for theoretical integration within the D-framework, a number of traits subsumed under different AMPD domains are substantively related to the D-Definition (for the definition of each aversive AMPD trait together with example items see Table 1). Prime instances exemplifying the feature of utility maximization are AMPD traits describing that one is willing to use others to achieve one's goals without feeling remorse, i.e., Manipulativeness, Deceitfulness, Callousness (Antagonism in the AMPD), but also Irresponsibility (Disinhibition in the AMPD), which implies a failure to honor one's commitments and obligations (see also its example item in Table 1), thereby potentially maximizing one's own utility at the expense of others under certain conditions. Moreover, utility can come in many forms. In particular, regulating negative emotions such as anger by harming or insulting others is also a possible manifestation of D (see Moshagen et al., 2018) which is reflected in Hostility. Also, Grandiosity and Suspiciousness (the latter mapped onto Negative Affectivity and Detachment in the AMPD) substantively overlap with the D definition, as they refer to the justifying belief aspect of D (especially the belief that others are exploitative is a prime justification for aversive behavior, see Hilbig, Moshagen, et al., 2022). By contrast, Attention Seeking, although empirically related to other traits subsumed under Antagonism as per the AMPD (e.g., Wright & Simms, 2014), shows hardly any substantive overlap with the definition of D and would thus not be considered a manifestation of D. Supporting these implications, previous research has indicated that traits typically considered relevant for antagonistic psychopathology (with the exception of Attention Seeking) are strongly associated with D (Scholz et al., 2022) and that D can serve as a common cause explanation for certain aversive (i.e., narcissistic, antisocial, and paranoid) personality disorders (Hilbig et al., 2020).

Moreover, a relatively strong prediction unique to the D-framework is that no (potential) manifestation of the common core may account for meaningful variance over and above the

common core in the tendency to maximize one's own utility at others' expense (Hilbig et al., 2023). Since this very tendency is part of the definition of D and since it is assumed that D *fully subsumes* this tendency to the degree it is present within any one trait, no trait should account for this tendency over and above D. The prediction has been shown to hold for as many as of 57 (potentially) aversive traits, none of which predicted meaningful variance over and above D in *Social Value Orientation* (SVO; Murphy et al., 2011), a direct and consequential measure of "the weights people assign to their own and others' outcomes in situations of interdependence" (Balliet et al., 2009, p. 533). In other words, SVO can be seen as a (behavioral) expression of part of the definition of D and thereby the very behavior that any aversive trait may account for *because* it is a manifestation of D.

Manifestation of D not (directly) represented in the AMPD

Another advantage of the D-framework is that the common core is explicitly considered to be a fluid construct, that is, reflected comparably in any sufficiently broad set of aversive trait indicators. This view is again adopted from the two-factor theory of intelligence and Spearman's "indifference of the indicator" principle. In other words, as long as a sufficient number of (items from) different manifestations are measured, one can approximate the same underlying construct (Horsten, Moshagen et al., 2023) and thereby approximate (i.e., predict) any other aversive trait – even those not included in the measurement. Correspondingly, it has been shown that any combination of at least six aversive traits essentially indicates the same D (Horsten, Moshagen et al., 2023) and that D longitudinally predicts aversive traits excluded from the measurement of D (Zettler et al., 2021). For example, D modeled without any Psychopathy items longitudinally predicts Psychopathy as well as Psychopathy predicts itself (Zettler et al., 2021). In the present context this fluid nature has an important implication: Even if certain aversive traits are not directly represented in the AMPD, as has been argued in previous research (Clark & Watson, 2022; Hilbig et al., 2020), considering aversive AMPD traits as manifestations of D implies that, in combination, aversive AMPD traits may approximate these allegedly missing pieces (at least, the aversive part of these).

For example, it has been noted that no AMPD trait directly or specifically refers to a primarily materialistic form of utility maximization (Clark & Watson, 2022), commonly referred to as *Greed* ("a desire to acquire goods, status, or power for the sake of acquisition without much regard for other individuals or the common good"; Marcus & Zeigler-Hill, 2015, p.436) in mainstream personality research. Similarly, the very tendency to construct justifications from diverse beliefs – often referred to as *Moral Disengagement* ("an individual difference in the way that people cognitively process decisions and behavior with ethical import that allows those inclined to morally disengage to behave unethically without feeling distress", Moore et al., 2012, p.2; see also Bandura, 1999, 2011) – is not reflected directly by any of the aversive AMPD traits. Lastly, deriving utility from actively provoking disutility in others – described in mainstream personality research by traits such as *Sadism* ("a person who humiliates others, shows a longstanding pattern of cruel or demeaning behavior to others, or intentionally inflicts physical, sexual, or psychological pain or suffering on others in order to assert power and dominance or for pleasure and enjoyment"; O'Meara et al., 2011, p. 523) or *Spitefulness* ("a preference that would harm another but that would also entail harm to oneself."; Marcus et al., 2014, p. 566) – is also not represented by any AMPD trait. Thus, aversive MPR traits such as Greed, Moral Disengagement, Sadism, and Spitefulness could be considered as missing pieces in the AMPD.

However, as per the D-framework, the extant aversive AMPD traits – in combination – should nonetheless reasonably account for such tendencies: Since the shared variance of extant AMPD should approximate D and since Greed, Moral Disengagement, Sadism, and Spitefulness are also manifestations of D, the latter ought to be accounted for by the former (in combination). In other words, even though the AMPD traits will arguably yield stronger overlap with aversive MPR traits with direct counterparts in the AMPD (such as Narcissism or Psychopathy), it should nonetheless be possible to predict substantial variance in aversive MPR traits outside the scope of the AMPD with the aversive AMPD traits combined. Stated simply, viewed through the lens of the D-

framework, there are no actual missing pieces in the AMPD as long as it includes a sufficiently broad set of D manifestations.

Hypotheses and predictions

As reasoned above, all aversive AMPD traits from the Antagonism domain (except for Attention Seeking) overlap with D substantively and may thus be seen as flavored manifestations of D. In addition, there are also traits typically located outside the Antagonism domain (and/or interstitial with other domains), namely Suspiciousness, Irresponsibility, and Hostility, which likewise substantively overlap with the definition of D; they, too, can be seen as flavored manifestations of D. By implication, the relation of any aversive AMPD trait with the prime behavioral expression of D – the tendency to maximize one's utility at others' expense – should be due to D. Thus, once controlling for D, no AMPD trait should be (negatively) related to SVO, with higher values indicating a lesser tendency to maximize one's own utility at the expense of others (see Hilbig et al., 2023). Taken together, we propose the following two general hypotheses:

Hypothesis 1 (H1): All AMPD traits conceptually overlapping with the theoretical definition of D are flavored manifestations of D. Specifically, this should apply for a) Callousness, b) Deceitfulness, c) Grandiosity, d) Hostility, e) Irresponsibility, f) Manipulativeness, and g) Suspiciousness, but not h) Attention Seeking.

Hypothesis 2 (H2): None of the traits a) to h) will be meaningfully (negatively) related to utility maximization at others' expense (as measured by SVO) after controlling for D.

In addition, as explained above, there are certain manifestations of D not directly covered by the AMPD, namely, Greed, Moral Disengagement, Sadism, and Spitefulness. Nonetheless, the extant aversive AMPD traits, in combination, ought to predict even these manifestations of D. Note, however, that this prediction depends entirely on the prime hypothesis specified above – that aversive AMPD traits *are* manifestations of D, that is, they share a single common core with each other and with all aversive MPR traits – being empirically supported. Therefore, we do not specify a hypothesis for this prediction as it is only sensible if H1 and H2 hold.

To test these hypotheses and predictions, we first conducted a re-analysis of existing data, i.e., the pilot study (*N* = 1,781), and then replicated the findings in a new dataset, i.e., the pre-registered main study (*N* = 2,006).

Transparency and openness

All data, analysis code, and research materials for both studies are available at the Open Science Framework (OSF; https://osf.io/uc38x/?view_only=16fe0ecd308941858b6553d1b6d83770). Data was analyzed using R, version 4.2.0 (R Core Team, 2023) and the package lavaan, version 0.6-11 (Rosseel et al., 2023). The pilot study was not preregistered. The hypotheses and analyses of the main study were preregistered (see https://aspredicted.org/TBM_4X1). Both studies were conducted in full accordance with the Ethical Guidelines of the American Psychological Association (APA) and were run based on approval from the local ethics committee (#LEK-477).

Pilot Study

Method

Sample

Data used for the pilot study stem from the Prosocial Personality Project (PPP) and are, in part, a reanalysis of data used by Scholz et al. (2022) and Hilbig et al. (2023). The PPP contains six consecutive primary waves (and additional follow up waves), covering various aspects of pro- and anti-social personality. For details on the sample characteristics, a-priori defined exclusion criteria, analyses scripts, data files, full documentation of all studies using (part of) the same data sets, and the PPP in general see the OSF (https://osf.io/m2abp/). The overall sample size for the present analyses is *N* = 1,781 (958 male, 821 female, 2 diverse; aged 18–74, *M* = 45.0, *SD* = 12.1 years), except for the analysis including SVO (*N* = 1,604), due to a smaller number of participants conducting the additional point of measurement for SVO (see *Measures & Procedure* for details). Participants had diverse educational backgrounds, with 36.8% holding a general certificate of secondary education (German: Realschule), 27.1% holding a vocational diploma or university-entrance diploma (German: Fachabitur or Abitur), and 35.1% holding a university/college degree. Similarly, the income

of the individual participants was quite heterogenous, with 30.9% earning less than 1,500€, 41.9% earning between 1,500€ and 3,000€, 13.8% earning between 3,000€ and 5,000€, and 2.6% earning more than 5,000€ a month (10.8% preferred not to say). These data stem from Wave 2 (T2; January 2020), Wave 3 (T3; January-February 2020), Wave 5 (T5; March 2020), and the follow up wave 2020-5b (T7b; May 2020) of the PPP.

Measures & Procedure

The following aversive MPR traits were measured either at T3 or T7b (average time lag between T3 and T7b = 3 months) by 118 items in total: Amoralism Crudelia (Knežević, 2003), Amoralism Frustralia (Knežević, 2003), Egoism (Weigel et al., 1999), Greed (Seuntjens et al., 2015), Machiavellianism (Jones & Paulhus, 2014), Moral Disengagement (Moore et al., 2012), Narcissism (Leckelt et al., 2018), Psychological Entitlement (Campbell et al., 2004), Psychopathy (Jones & Paulhus, 2014), Sadism (O'Meara et al., 2011), Self-Centeredness (Arneklev et al., 1993), and Spitefulness (Marcus et al., 2014). These traits were all measured on a five-point type Likert scale. Traits were selected to represent all aspects of D and to maximize consistency with prior research on D (Moshagen et al., 2018; Moshagen, Zettler & Hilbig, 2020) and for their typically well-established psychometric properties, also in their German versions as used in the present study (for details see Zettler et al., 2021). D was modeled as a general factor of these twelve traits (see the Analyses & Inference Criteria for details), given that previous research showed that D is indeed the common core of those twelve traits (Moshagen, Zettler & Hilbig, 2020).

The aversive AMPD traits (Attention Seeking, Deceitfulness, Callousness, Hostility, Irresponsibility, Grandiosity, Manipulativeness, & Suspiciousness) were measured with the Personality Inventory for DSM-5 Short Form (PID-5-SF; Maples et al., 2015; German version: Zimmermann et al., 2014) at T2 on a four-point Likert-type scale. The PID-5-SF assesses each AMPD trait with four items. We chose the PID-5-SF because the PID-5 represents the most widely used instrument for AMPD traits (Krueger & Hobbs, 2020; Zimmermann et al., 2019) and its short form has good psychometric properties (Maples et al., 2015).

The weight one assigns to one's own utility relative to others' utility was measured by the *SVO slider measure* (see Murphy et al., 2011) at T5 which has been shown to be superior to other measures of social preferences regarding psychometric properties and efficiency (Murphy & Ackermann, 2014). Therein, participants allocate points that are worth money between themselves and another individual. Together, the six items (see Figure 1) cover the full range of preferences from prosocial (assigning the same positive weight to one's own and the other's utility) to competitive (assigning the same positive weight to one's own and *negative* weight to the other's utility). For the statistical analyses, we transformed the decisions over all six items into one continuous measure of SVO, the so-called *SVO angle*, as recommended by Murphy et al. (2011) such that greater values will indicate more prosocial tendencies and smaller values more anti-social preferences. In implementing the task, we followed common, well-established standards for implementing economic games (Thielmann et al., 2021): Participants were informed that ten points in the SVO paradigm are worth 0.20€ of additional payoff. It was explained that *after* the experiment, participants would be randomly assigned the role of the sender (i.e., be paid according to their own allocation decision) or the role of the receiver (i.e., be paid according to the allocation decision made by someone else) and one of the six decisions of the sender (*during* the study all participants acted as senders) would be chosen at random to determine their payment (payed out by the panel provider after completion of the study). We exactly followed this protocol and hence no deception was used in this study (Hilbig & Thielmann, 2021; Hilbig, Thielmann & Böhm, 2022). Participants could earn a maximum bonus payment of 2.00€, i.e., more than twice their flat fee for participation at this measurement occasion. These incentives and actual assignment of senders and receivers render the SVO-slider a measure of "actual" (Baumeister et al., 2007), i.e., consequential (Klein & Hilbig, 2019), behavior in the sense that it is *structurally* comparable to certain real-life situations. For example, the Dictator Game (having an endowment and giving some of it to an anonymous other) is structurally comparable to donating blood (Thielmann et al., 2021). Of course, the two are not the same. But they entail the same underlying incentive and decision structure and thus one (the Dictator Game) "models" the other (donating blood). Therefore, the SVO paradigm,

as opposed to any self-report, does not (or at least to a lesser degree) share response tendencies (especially social desirability) with self-reports on aversive traits. For the complete instructions of the SVO slider measures we refer readers to the OSF repository.

Analyses & Inference Criteria

We performed structural equating modeling (SEM) using the R package "lavaan" (Rosseel et al., 2023) employing ML-estimation with robust standard errors and a Satorra-Bentler scaled test statistic. Raw item scores were used as indicators for all analyses. Latent factors were identified by fixing their variance to one. Absolute model fit was evaluated using the *Root Mean Square Error of Approximation (RMSEA)* and the *Standardized Root Mean Square Residual (SRMR)* as these are sensitive to different types of misspecifications (Moshagen & Auerswald, 2018). In light of recent criticism on the *comparative fit index* (CFI), e.g., that it conflates the magnitude of loadings with model fit (Moshagen & Auerswald, 2018; van Laar & Braeken, 2022), we did not consider it for the evaluation of the models but report it for the sake of transparency. Due to the large sample size, the chi-squared goodness of fit statistic was not interpreted for an evaluation of general model fit, as even minor and neglectable deviations in the likelihood ratio test from the saturated model will turn out significant. For example, a post-hoc power analyses for global hypotheses testing (see Jobst et al., in press) revealed a power greater than 99% to detect an even minor and neglectable deviation (*RMSEA* = .01) from a saturated model for the sample size of *N* = 1,781.

For H1, we modeled D via the twelve aversive MPR traits in a bi-factor structure (see also Moshagen et al., 2018). Hence, all items load on D as the general factor (g-factor). In addition, items load on a specific factor for their respective scale (i.e., only items from the same scale load on this factor). Specific factors account for the residualized covariances between items of a specific trait. All factors were specified to be orthogonal to each other, as it is typically done in bi-factor models (Reise, 2012). Thus, the shared variance of all indicators of one aversive trait scale is decomposed into two parts: First, shared variance due to D (i.e., shared with the items from all other aversive trait scales) and second, variance shared only among the indicators of one aversive trait beyond D,

i.e., the specific "unique flavor" of a manifestation (which is not shared with D or any other manifestation). We will refer to this model as the *base model* in what follows (see Figure 2).

Next, we introduced each of the aversive AMPD traits into the base model separately, i.e., the items of each specific trait load on D (as the g-factor) and again also on a specific factor representing the respective AMPD traits (see Figure 3). Thus, for each AMPD trait, there is a separate model. Again, all factors are orthogonal to each other, thus also allowing for decomposing the shared variance of the indicators for each aversive AMPD trait into two independent sources.

For evaluating H1, we considered the *explained common variance* (*ECV*; Berge & Sočan, 2004). The *ECV* indicates the ratio of common variance explained in the items of a particular trait that is captured by the general factor (i.e., D) relative to the total common variance. A value of 0 thus indicates that none of the shared variance between items is due to the g-factor (and, thus, that all shared variance is due to the specific factor), whereas a value of 1 indicates that the entire shared variance is explained by the g-factor. In other words, the *ECV* directly indicates the D-saturation of each trait. Hence, the interpretation is similar to R^2 in regression models. In line with common heuristics regarding the effect size of R^2 according to Cohen (1992), we considered a medium to strong effect, i.e., an *ECV* > 20%, as an indication that the respective trait can be considered as a *sufficiently* aversive manifestation of D (see H1). To further test this statistically, we compared two models for each AMPD trait, one freely estimating the ECV and another with ECV fixed to 0%. These model comparisons were based on corrected chi square differences (Satorra & Bentler, 2010).

Moreover, to test whether any of the aversive AMPD traits is negatively related to SVO over and above D (H2), we predicted SVO by D and each aversive AMPD trait, respectively, in latent regression models: First separately, to obtain zero-order effects, and then combined (D and the respective AMPD trait), to obtain unique effects of each AMPD trait beyond D. Again, as power to detect any effect was very high, only traits with a unique negative beta weight and an increase of R^2 of over 2%, i.e., a non-trivial effect (Cohen, 2013, p. 413), were considered to have accounted for relevant variance beyond D.

Finally, in order to test the coverage of aversive MPR traits by the aversive AMPD traits, we modeled the aversive MPR traits without D as the common core, i.e., as correlated factors with the items of each aversive MPR trait only loading on one respective factor for each scale. Then we regressed the twelve aversive MPR traits onto the eight aversive traits from the AMPD (in a single model). For evaluating the coverage, we compared the explained variance (R^2) in each aversive MPR traits. The reasoning laid out before would indicate that, while all aversive MPR traits should be covered sufficiently, for a) Greed (materialistic form of utility maximization at others' expense), b) Moral Disengagement (constructing justifications for aversive behavior), as well as c) Sadism and Spitefulness (provoking disutility), the aversive AMPD traits exhibit relatively low explained variance compared to the explained variance in other aversive MPR traits. Note that for these analyses, if predictors were highly correlated (i.e., latent correlations between factors of $r > .90$), items of the respective scales were assigned to one single factor to avoid multicollinearity (this was the case for Manipulativeness and Deceitfulness, $r = .97$).

Results

For latent correlations, reliability estimates (ω; McDonald, 1999) of all traits, and the *ECV*s of all aversive MPR traits see Supplementary Table S1 and S2 on the OSF. The base model showed a good fit to the data, χ^2 (6667) = 20,077, *SRMR* = .047, *RMSEA* = .034, 90% CI [.033 - .034], *CFI* = .802. The reliability of D, as the g-factor, was very high (ω = .91). All subsequent models including one of the aversive AMPD traits as an additional facet showed good model fit as well, $Range_{SRMR}$: .047 - .049; $Range_{RMSEA}$: .033 - .034; $Range_{CFI}$ = .798 - .806 (see Supplementary Table S3 on the OSF for details).

Regarding H1, the *ECV*s for all aversive AMPD traits are displayed in Table 2 (column 2). As expected, all aversive AMPD traits, except for Attention Seeking, exhibited an *ECV* larger than 20%, with Hostility and Irresponsibility showing relatively lower *ECV*s of 22% and 23%, respectively. Further in line with H1, Attention Seeking showed the lowest *ECV* (8%) of all aversive AMPD traits. Although this is statistically significant, it is considerably below the threshold set for a *sufficiently*

aversive manifestation of D, i.e., an *ECV* > 20%, which is also not covered by the respective 95%-CI. Furthermore, in line with H2, none of the aversive AMPD traits showed a (significant) negative regression coefficient in a latent regression model predicting SVO once controlling for D, although zero-order relations of the aversive AMPD traits with SVO were all negative (see Table 2, columns 3 and 4). Thus, while all aversive AMPD traits per se predicted utility maximization at others' expense, none did so beyond D. Moreover, the relationship of D and SVO (r = -.30, 95% CI [-.35; -.25], p < .001) remained virtually the same, regardless of which AMPD trait was included in the model (see Table 2, column 5). It should be noted that this is a relatively strong effect, i.e., exceeding the upper end of (meta-analytically established) effect-size estimates on the relation of prosocial behavior (in consequential games) and (self-reported) personality traits of -.18 ≤ r ≤ .26 (see Thielmann et al., 2020), especially considering that the measures were administered at two different measurement occasions.

Regarding the coverage of aversive MPR traits by the aversive AMPD traits, Sadism (R^2 = 29%), Greed (R^2 = 30%), Spitefulness (R^2 = 33%), and Moral Disengagement (R^2 = 35%) were accounted for least well, thus supporting the notion that these aspects might be underrepresented in the AMPD. For comparison, the explained variance for all other aversive MPR traits varied between 39% and 58% (see Supplementary Table S4 on the OSF for details). Overall, however, as implied by confirmation of Hypothesis 1 and 2, the aversive AMPD traits predicted a large proportion of variance in *all* aversive MPR traits, even those allegedly missing in the AMPD

Discussion

In sum, the pilot data largely supports both predictions: Aversive AMPD traits aligned with the definition of D can be seen as specific, flavored manifestation of D. Moreover, aversive MPR traits not explicitly represented in the AMPD could still be explained to a notable extent in absolute terms by aversive AMPD traits (combined), though relatively less well than traits explicitly represented in the AMPD.

However, results are limited in at least three ways. First, the aversive AMPD traits were measured with a short version of the PID-5 (Maples et al., 2015). Thus, measurement may have missed out on the full breadth of aversive AMPD traits, which is potentially problematic as the relevant test statistic (i.e., *ECV*) reflects the relative proportion of a trait covered by D and may also affect conclusions about the breadth of content covered by aversive AMPD traits in aversive MPR traits. This could further be especially relevant for Hostility and Irresponsibility, as the results for these traits were least clear. Second, different traits were measured at different measurement occasions (sometimes with months in between). This is potentially problematic for comparing the unique effects in SVO, as some of the traits were measured temporally closer to SVO than others. Finally, hypotheses and analyses were not pre-registered. To address these limitations, we conducted a preregistered replication study (see https://aspredicted.org/TBM_4X1).

Main Study

Method

Sample & Exclusion criteria

Participants were recruited and financially remunerated through a professional online panel provider. Participants had to be at least be 18 years old, German citizens, and a sufficient level of German to participate in the study. We used quotas during recruiting to ensure an approximately representative distribution of age and gender in the sample. We aimed at a minimum sample size of *N* = 2,000 after exclusion criteria (see below). This sample size was sufficient to detect the previously defined threshold of an ECV greater than 20% with a power of over 99% according to an a-priori local hypothesis testing power analyses (see Jobst et al., in press).

Participants were excluded if they a) did not complete all measures at both measurement occasions, b) failed any of the attention check items embedded in the survey (two items at each measurement occasion), or c) showed very short completion times on any of the scales (i.e., less than 2s per item on average). These criteria were specified a priori and pre-registered. Applying

these criteria resulted in a final sample size of *N* = 2,006 for all analyses (995 female, 1006 male, 5 diverse; aged 18–74, *M* = 47.9, *SD* = 15.1 years).

Procedure

We collected data at two measurement occasions (between May and June 2023) in order to avoid effects of consistent responding as well as fatigue and boredom (average time lag between measurement occasions: seven days). At the first occasion, all aversive MPR traits were measured. At the second occasion, the aversive AMPD traits, as well as the SVO were measured. At both measurement occasions, the order of the traits was randomized between participants. Participants received the same instructions as in the pilot study for all measures.

Measures & Analyses

To maximize consistency, we used the same scales as in the pilot study to measure the 12 aversive MPR traits and again used the 6-item SVO-slider, with the same incentive structure. In contrast to the pilot study, we measured the aversive AMPD traits (Attention Seeking, Deceitfulness, Callousness, Hostility, Irresponsibility, Grandiosity, Manipulativeness, & Suspiciousness) with the respective items from the (full) PID-5 (Krueger et al., 2012; German version: Zimmerman et al., 2014), and thus with 67 items in total. Again, for these traits, answers were given on a four-point Likert-type scale. Analyses were preregistered (see https://aspredicted.org/TBM_4X1)[5] and conducted exactly as in the pilot study using the same statistical approaches. Also, the same criteria were used to evaluate the hypothesis tests.

Results

For latent correlations, reliability estimates (ω; McDonald, 1999) of all traits, and the *ECV*s of all aversive MPR traits see Supplementary Tables S6 and S7 on the OSF. Again, the base model showed a good fit to the data, χ^2 (6667) = 23,114, *SRMR* = .052, *RMSEA* = .035, 90% CI [.035 - .036],

[5] We also preregistered to allow the residuals of the respective items to correlate in case of considerable item overlap (i.e., a raw item correlation of $r > .80$) between the items of any aversive MPR or AMPD trait. However, no such case appeared in the main study (or in the pilot study). Also note that we stated in the pre-registration that the threshold for triviality as per Cohen (2013, p. 413) is 1.8%, whereas it actually is 2%. Crucially, both thresholds lead to the same conclusions regarding all hypotheses.

CFI = .792, and reliability of D, as the g-factor, was very high (ω = .92). All subsequent models including one of the aversive AMPD traits as an additional facet showed good model fit as well, $Range_{SRMR}$: .051 - .055; $Range_{RMSEA}$: .034 - .035; $Range_{CFI}$ = .792 - .806 (see Supplementary Table 8 on the OSF)[6].

Regarding H1, the *ECV*s for all aversive AMPD traits are displayed in Table 3 (column 2). Largely in line with H1, all aversive AMPD traits exhibited an *ECV* larger than 20%. However, this was also the case for Attention Seeking, which contradicts H1h. Furthermore, in line with H2, none of the aversive AMPD traits showed a non-trivial increase of more than 2% in R^2 predicting SVO over and above D (see Table 3, column 6), although all zero-order relations, and some unique effects, of the aversive AMPD traits with SVO were negative (see Table 3, columns 3 and 4). Therefore, as expected, for any manifestation of D, none of the aversive AMPD traits accounted for utility maximization at others' cost beyond D. Again, the relationship between D and SVO (r = -.27, 95% CI [-.32; -.22], p < .001) was stable regardless of which AMPD trait was included in the model (see Table 3, column 5).

Regarding the coverage of aversive MPR traits by the aversive AMPD traits, the AMPD traits accounted least variance in Greed (R^2 = 37%), but considerably more variance in Sadism (R^2 = 42%), Spitefulness (R^2 = 49%), and Moral Disengagement (R^2 = 51%) than in the pilot study. For comparison, the explained variance for all other aversive MPR traits varied between 47% and 67% with a median of 51% (see Supplementary Table S9 for details and Supplementary Table S10 for the ECVs and R^2 of the aversive MPR traits on the OSF). In sum, as implied by the notion that aversive AMPD traits are manifestations of D, substantial variance in *all* aversive MPR traits – those with and without counterpart traits in the AMPD – was accounted for by the aversive AMPD traits in absolute terms (with Greed bringing up the rear in relative terms).

General Discussion

[6] To supplement these analyses, we also estimated a bi-factor model across all 20 aversive traits at once, which also showed good fit to the data, χ^2 (16650) = 50,575, SRMR = .057, RMSEA = .032, CFI = .760.

Describing and understanding socially and/or ethically aversive personality has progressed with relatively little theoretical integration of approaches from psychopathology research – especially the maladaptive trait facets specified in the Alternative Model of Personality Disorders (AMPD) – and mainstream personality research (where these traits are often referred to as "dark" traits). Prior work trying to overcome this divide of aversive tendencies across mainstream personality and clinical research (e.g., Lynam & Miller, 2019; Sleep et al., 2020; Somma et al., 2020) has been limited in two ways: First, there are inconsistencies in clinical research in regard to which traits are mapped onto Antagonism. Thus, it is still unclear (or left to idiosyncrasies) what exactly the shared aspect of such traits, i.e., Antagonism itself, could be. Second, research has almost exclusively focused on the Dark Triad so far (Hilbig et al., 2020; Thomaes et al., 2017) which actually provides a relatively poor approximation of the common core of all aversive MPR traits (Horsten, Moshagen, et al., 2023).

To overcome these limitations, we herein rely on the D-Factor of personality (D; Moshagen et al., 2018), which is conceptualized as the common core of *all* aversive traits within a unifying framework. Specifically, this framework implies that any trait (of either tradition) that is substantively related to the definition of D (i.e., utility maximization at the cost of others, accompanied by justifying beliefs) can be understood as a flavored manifestation of D. A flavored manifestation encompasses (aversive) aspects in line with the D-definition but may comprise aspects beyond this common aversive core, e.g., Psychopathy can be seen as a manifestation of D, flavored by disinhibition (Bader et al., 2022). We investigated these propositions by concurrently considering eight aversive AMPD traits (see Table 1) and twelve aversive traits from mainstream personality research (aversive MPR traits), also testing their nature as flavored manifestations with a behavioral measure of how individuals weigh their own utility relative to others', i.e., the social value orientation (SVO; see Murphy & Ackermann, 2014; Murphy et al., 2011).

Using structural equation modeling, the general proposition that aversive AMPD traits can be seen as a flavored manifestation of D was confirmed both in the pilot study (re-analysis) and the

preregistered main study. The only AMPD trait for which the findings were somewhat inconsistent was Attention Seeking: Whereas the pilot study confirmed the prediction that Attention Seeking is largely unrelated to D and the behavioral manifestation of utility maximization at others' expense, i.e., SVO, the main study suggested that up to 30% of total variance explained in Attention Seeking is due to D. Together with Irresponsibility, this was still the lowest value among all aversive AMPD traits indicating that these traits have a substantial, non-aversive proportion of variance beyond D. Nonetheless, our data cannot rule out that Attention Seeking is also a flavored manifestation of D. This was unexpected on a substantive level and we can only speculate why that is the case. For one, it may be that attention – being a primary reinforcer – simply yields substantial utility; thus, anyone maximizing their utility (someone high in D) may often seek attention, even at others' expense. It may also be that Attention Seeking not only encompasses the *need* for attention but also a sense that one is special and *deserves* the attention, i.e., a sense of entitlement (e.g., exemplified in the item from the PID-5: "I do things so that people just have to admire me."), which would seem compatible with the Narcissistic Entitlement theme subsumed by D (Bader et al., 2022). Clearly, future research will need to address these speculations and thus help to further clarify the extent to which attention seeking is a manifestation of D.

Setting aside the inconsistent findings for Attention Seeking, the results provide notable support for the notion that aversive AMPD traits – like aversive MPR traits – can be seen as flavored manifestations of one common aversive core. As such, D represents the shared aversive essence of all these traits well and its framework serves theoretical integration within a unifying framework.

Moreover, researchers have noted that the AMPD does not fully cover the broader space of aversive tendencies (Clark & Watson, 2022; Hilbig et al., 2020). The D-Framework and the view that aversive AMPD traits are manifestations of D, however, would imply sufficient (albeit indirect) coverage of any aversive trait by the aversive AMPD traits combined. Indeed, all aversive MPR traits were covered well (all $29\% \leq R^2 \leq 67\%$ across both studies, i.e., a large effect) by the AMPD, even for traits which have no clear AMPD counterpart, i.e., Greed, Moral Disengagement, Sadism, and

Spitefulness. Indeed, for the latter three, strong empirical overlap with Antagonism as measured by the PID-5 has been shown before (Russell et al., 2017; Russell & King, 2017; Somma et al., 2020; Zeigler-Hill & Noser, 2018). Thus, in line with the theorized fluid nature of D and prior findings showing that a sufficiently broad set of aversive traits (or items) will always approximate D (Horsten, Moshagen et al., 2023; Moshagen et al., 2018; Zettler et al., 2021) the shared variance of the aversive AMPD traits approximates D, in turn allowing it to account for such tendencies which would seem to be missing in the AMPD.

In the end, different conclusions could be drawn here: On the one hand, it is arguably satisfactory for clinical research to know that the aversive personality space is well covered by the aversive AMPD traits (in combination), albeit indirectly in some cases. As such, calls for including more traits would seem less urgent. On the other hand, clinical practitioners may be interested in a more fine-grained picture of how the aversive core is manifested in each individual. As such, the simple fact that some aversive tendencies – like Greed – cannot be directly measured within AMPD and are clearly not *fully* covered by any of the traits already included might be taken to imply some additions could be worthwhile. Arguably, there is no one correct answer to this trade-off between theoretical parsimony and practical specificity.

The D-framework and Antagonism

As hinted in the introduction, some controversy remains over whether D and Antagonism (and/or low Agreeableness for that matter) should be considered as equivalent. Although this debate is largely about labels and the similarity of nomological nets, neither of which we studied herein, it may be informed by prior and the present evidence that there is a (single) aversive core shared by *all* aversive traits and that this core strongly reflects utility maximization at others' expense, accompanied by justifying beliefs. One could argue that this substantive meaning is subsumed under or implied by Antagonism as per the AMPD (American Psychiatric Association, 2013), which notes that Antagonism includes "both an unawareness of others' needs and feelings and a readiness to use others in the service of self-enhancement" (p. 780) and the "expectation of

special treatment" (p. 780). Whereas this will ultimately depend on idiosyncrasies[7], we must point out that Antagonism has not been conceptualized in a framework that is equivalent to that of D and would have allowed for deriving the predictions tested herein. In other words, the label Antagonism has so far not been used within a framework predicting a) that *any* aversive trait (from MPR or clinical psychology) is a flavored manifestation of a (single aversive) common core, b) that no aversive trait predicts relevant variance beyond the common core in utility maximization at others' costs (i.e., SVO), or c) the aversive AMPD traits in combination will approximate traits not represented within the AMPD.

As such, even if one interprets the substantive definitions of D and AMPD Antagonism as equivalent, concluding that D *is* Antagonism would also entail adopting the (conceptual and statistical) framework of D and imposing it onto Antagonism. So far, Antagonism has not been explicitly defined in terms of a framework from which any of the above strict predictions (a-c) can be derived. The implications of such a proposition should be carefully considered and critically examined in future research and clearly exceed the scope of the present work.

Limitations and directions for future research

First, like most research on aversive personality we mostly relied on self-reports, which may inflate the observed correlations due a common method bias. Nonetheless, we did also include a fully consequential measure of utility maximization (i.e., the SVO slider measure) and the associations between self-reported traits and this measure demonstrate that all these traits are indeed linked to actual utility maximization at others' cost. In turn, no trait accounts for utility

[7] Ultimately, it depends on how much one wants to read into the definition of Antagonism; in our reading the definitions still differ in that D not only involves "unawareness" of others' needs and feeling; instead, it may actually involve behavior for the sake of others' disutility (i.e., Sadism) which necessarily hinges on such awareness: Without awareness of others' disutility, one simply cannot derive pleasure (and thus utility) from it. Moreover, the definition of Antagonism does not explicitly mention *justifying* beliefs; some beliefs, namely Grandiosity (plus Suspiciousness in HiTOP), may be an implicit part of the definition because they are defined as facets of Antagonism, but others are not. The definition of D, by contrast, refers to *any* belief that serves to *justify* aversive behavior and empirical evidence indeed confirms strong links to beliefs such as *Social Dominance Orientation* or *Negative Reciprocity Norm Endorsement*, neither of which are implicitly or explicitly represented in the Antagonism domain of the AMPD (or HiTOP for that matter).

maximization at others' cost once D is accounted for. This pattern cannot simply be due to common method variance alone. Nonetheless, future work might consider observer reports and clinical interviews (e.g., Ro et al., 2017), other and more diverse measures of aversive behavior, or objective life outcomes (such as incarceration; Bonfá-Araujo et al., 2023). Second, although participants could substantially increase their payoff by means of the SVO (doubling their flat fee for the respective measurement occasion in both studies), the maximum incentive of 2€ is limited in absolute terms. Future work may thus consider still larger incentives and/or a measure of how much participants cared about the incentives offered. Third, we relied on the PID-5 as the operationalization of the aversive AMPD traits, as it is the most widely used inventory for the AMPD. Nevertheless, replicating the findings with measures of other trait models, such as the Computerized Adaptive Test of Personality Disorder (CAT-PD; Simms et al., 2011), may be worthwhile, especially if they include substantively distinct (aversive) traits.

Conclusion

In sum, the D-framework offers a unifying theoretical foundation for similarities and differences of socially/ethically aversive traits *across* research areas (i.e., mainstream personality research and clinical psychopathology research) and *within* areas. The latter is particularly apparent when considering the various suggestions of where to locate certain aversive AMPD traits, e.g., Suspiciousness, as well as the interstitially of traits within the AMPD, e.g., Hostility. The D-Framework is a functional theory in that it unifies traits of a certain nature, i.e., all socially/ethically aversive traits, be it constituents of the Dark Triad, other aversive MPR traits, facets of Antagonism, or facets of other higher-order dimensions in the AMPD. As such, the framework is not suggested as a competing model to the AMPD (or HiTOP). Instead, it offers an account of what *all* aversive traits share – independent of their inclusion or location in any such taxonomy and indeed independent of any particular research tradition (mainstream personality or clinical personality psychopathology).

References

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

Arneklev, B. J., Grasmick, H. G., Tittle, C. R., & Bursik, R. J. (1993). Low self-control and imprudent behavior. *Journal of Quantitative Criminology*, *9*(3), 225–247. https://doi.org/10.1007/BF01064461

Ashton, M. C., Lee, K., Vries, R. E. de, Hendrickse, J., & Born, M. (2012). The maladaptive personality traits of the Personality Inventory for DSM-5 (PID-5) in relation to the HEXACO personality factors and schizotypy/dissociation. *Journal of Personality Disorders*, *26*(5), 641–659. https://doi.org/10.1521/pedi.2012.26.5.641

Bader, M., Hartung, J., Hilbig, B. E., Zettler, I., Moshagen, M., & Wilhelm, O. (2021). Themes of the dark core of personality. *Psychological Assessment*.

Bader, M., Hilbig, B. E., Zettler, I., & Moshagen, M. (2022). Rethinking Aversive Personality: Decomposing the Dark Triad Traits into their Common Core and Unique Flavors. *Journal of Personality.* Advance online publication. https://doi.org/10.1111/jopy.12785

Bader, M., Horsten, L. K., Hilbig, B. E., Zettler, I., & Moshagen, M. (2021). Measuring the Dark Core of Personality in German: Psychometric Properties, Measurement Invariance, Predictive Validity, and Self-Other Agreement. *Journal of Personality Assessment*, 1–14. https://doi.org/10.1080/00223891.2021.1984931

Balliet, D., Parks, C., & Joireman, J. (2009). Social Value Orientation and Cooperation in Social Dilemmas: A Meta-Analysis. *Group Processes & Intergroup Relations*, *12*(4), 533–547. https://doi.org/10.1177/1368430209105040

Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review : An Official Journal of the Society for Personality and Social Psychology, Inc*, *3*(3), 193–209. https://doi.org/10.1207/s15327957pspr0303_3

Bandura, A. (2011). *Moral Disengagement*. *The Encyclopedia of Peace Psychology*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470672532.wbepp165

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 : A Journal of the Association for Psychological Science*, *2*(4), 396–403. https://doi.org/10.1111/j.1745-6916.2007.00051.x

Berge, J. M. ten, & Sočan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika*, *69*(4), 613–625.

Bonfá-Araujo, B., Ferreira, L. B., Jesuíno, A. D. S. A., Hauck-Filho, N., & Iglesias, F. (2023). Measuring the dark core: A Brazilian adaptation and comparison between the general population and incarcerated men. *Journal of Criminal Justice*, *89*, 102133. https://doi.org/10.1016/j.jcrimjus.2023.102133

Campbell, W. K., Bonacci, A. M., Shelton, J., Exline, J. J., & Bushman, B. J. (2004). Psychological entitlement: Interpersonal consequences and validation of a self-report measure. *Journal of Personality Assessment*, *83*(1), 29–45. https://doi.org/10.1207/s15327752jpa8301_04

Carroll, J. B. (2009). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.

Clark, L. A., & Watson, D. (2022). The trait model of the DSM-5 alternative model of personality disorder (AMPD): A structural review. *Personality Disorders: Theory, Research, and Treatment*, *13*(4), 328–336. https://doi.org/10.1037/per0000568

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.

Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences*. Elsevier Science.

Crowe, M. L., Lynam, D. R., & Miller, J. D. (2018). Uncovering the structure of agreeableness from self-report measures. *Journal of Personality*, *86*(5), 771–787. https://doi.org/10.1111/jopy.12358

Furnham, A., Richards, S. C., & Paulhus, D. L. (2013). The Dark Triad of Personality: A 10 Year Review. *Social and Personality Psychology Compass*, *7*(3), 199–216. https://doi.org/10.1111/spc3.12018

Grigoras, M., & Wille, B. (2017). Shedding light on the dark side: Associations between the dark triad and the DSM-5 maladaptive trait model. *Personality and Individual Differences*, *104*, 516–521. https://doi.org/10.1016/j.paid.2016.09.016

Hilbig, B. E., Moshagen, M., Horsten, L. K., & Zettler, I. (2021). Agreeableness is dead. Long live Agreeableness? Reply to Vize and Lynam. *Journal of Research in Personality*, *91*, 104074. https://doi.org/10.1016/j.jrp.2021.104074

Hilbig, B. E., Moshagen, M., Thielmann, I., & Zettler, I. (2022). Making rights from wrongs: The crucial role of beliefs and justifications for the expression of aversive personality. *Journal of Experimental Psychology: General.* Advance online publication. https://doi.org/10.1037/xge0001232

Hilbig, B. E., & Thielmann, I. (2021). On the (Mis)Use of Deception in Web-Based Research. *Zeitschrift Für Psychologie*, *229*(4), 225–229. https://doi.org/10.1027/2151-2604/a000466

Hilbig, B. E., Thielmann, I., & Böhm, R. (2022). Bending Our Ethics Code. *European Psychologist*, *27*(1), 62–70. https://doi.org/10.1027/1016-9040/a000431

Hilbig, B. E., Thielmann, I., Klein, S. A., Moshagen, M., & Zettler, I. (2020). The dark core of personality and socially aversive psychopathology. *Journal of Personality.* Advance online publication. https://doi.org/10.1111/jopy.12577

Hilbig, B. E., Thielmann, I., Zettler, I., & Moshagen, M. (2023). The Dispositional Essence of Proactive Social Preferences: The Dark Core of Personality vis-à-vis 58 Traits. *Psychological Science*, *34*(2), 201–220. https://doi.org/10.1177/09567976221116893

Horsten, L. K. (2023). *Manoeuvering through the jangle jungle. What is the common core of aversive personality traits (not)?* [Doctoral dissertation, RPTU Kaiserslautern-Landau]. RIS. https://kluedo.ub.rptu.de/files/7300/dissertation_horsten_2023.pdf

Horsten, L. K., Hilbig, B. E., Thielmann, I., Zettler, I., & Moshagen, M. (2022). Fast, but not so furious. On the distinctiveness of a fast life history strategy and the common core of aversive traits. *Personality Science*, *3*, e6879. https://doi.org/10.5964/ps.6879

Horsten, L. K., Moshagen, M., Thielmann, I., Zettler, I., & Hilbig, B. E. (2023). How many (and which) aversive traits does it take to approximate the common core of aversive personality? [Manuscript submitted for publication].

Horsten, L. K., Thielmann, I., Moshagen, M., Zettler, I., Scholz, D., & Hilbig, B. E. (2023). Testing the equivalence of the aversive core of personality and a blend of agreeableness(-related) items. *Journal of Personality.* Advance online publication. https://doi.org/10.1111/jopy.12830

Jobst, L., Bader, M., & Moshagen, M. (in press). A Tutorial on Assessing Statistical Power and Determining Sample Size for Structural Equation Models. *Psychological Methods*.

Jones, D. N., & Paulhus, D. L. (2014). Introducing the short Dark Triad (SD3): A brief measure of dark personality traits. *Assessment*, *21*(1), 28–41. https://doi.org/10.1177/1073191113514105

Klein, S. A., & Hilbig, B. E. (2019). On the Lack of Real Consequences in Consumer Choice Research. *Experimental Psychology*, *66*(1), 68–76. https://doi.org/10.1027/1618-3169/a000420

Knežević, G. (2003). *Koreni amoralnosti [Roots of amoralism].* Centar za primenjenu psihologiju, IKSI, Institut za psihologiju.

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K. E., . . . Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454–477. https://doi.org/10.1037/abn0000258

Kowalski, C. M., Rogoza, R., Saklofske, D. H., & Schermer, J. A. (2021). Dark triads, tetrads, tents, and cores: Why navigate (research) the jungle of dark personality models without a compass (criterion)? *Acta Psychologica*, *221*, 103455. https://doi.org/10.1016/j.actpsy.2021.103455

Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2012). Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychological Medicine*, *42*(9), 1879–1890. https://doi.org/10.1017/S0033291711002674

Krueger, R. F., & Hobbs, K. A. (2020). An Overview of the DSM-5 Alternative Model of Personality Disorders. *Psychopathology*, *53*(3-4), 126–132. https://doi.org/10.1159/000508538

Krueger, R. F., Hobbs, K. A., Conway, C. C., Dick, D. M., Dretsch, M. N., Eaton, N. R., Forbes, M. K., Forbush, K. T., Keyes, K. M., Latzman, R. D., Michelini, G., Patrick, C. J., Sellbom, M., Slade, T., South, S. C., Sunderland, M., Tackett, J., Waldman, I., Waszczuk, M. A., . . . Kotov, R. (2021). Validity and utility of Hierarchical Taxonomy of Psychopathology (HiTOP): Ii. Externalizing superspectrum. *World Psychiatry : Official Journal of the World Psychiatric Association (WPA)*, *20*(2), 171–193. https://doi.org/10.1002/wps.20844

Leckelt, M., Wetzel, E., Gerlach, T. M., Ackerman, R. A., Miller, J. D., Chopik, W. J., Penke, L., Geukes, K., Küfner, A. C. P., Hutteman, R., Richter, D., Renner, K.-H., Allroggen, M., Brecheen, C., Campbell, W. K., Grossmann, I., & Back, M. D. (2018). Validation of the Narcissistic Admiration and Rivalry Questionnaire Short Scale (NARQ-S) in convenience and representative samples. *Psychological Assessment*, *30*(1), 86–96. https://doi.org/10.1037/pas0000433

Lynam, D. R., & Miller, J. D. (2019). The basic trait of Antagonism: An unfortunately underappreciated construct. *Journal of Research in Personality*, *81*, 118–126. https://doi.org/10.1016/j.jrp.2019.05.012

Maples, J. L., Carter, N. T., Few, L. R., Crego, C., Gore, W. L., Samuel, D. B., Williamson, R. L., Lynam, D. R., Widiger, T. A., Markon, K. E., Krueger, R. F., & Miller, J. D. (2015). Testing whether the DSM-5 personality disorder trait model can be measured with a reduced set of items: An item response theory investigation of the Personality Inventory for DSM-5. *Psychological Assessment*, *27*(4), 1195–1210. https://doi.org/10.1037/pas0000120

Marcus, D. K., & Zeigler-Hill, V. (2015). A Big Tent of Dark Personality Traits. *Social and Personality Psychology Compass*, *9*(8), 434–446. https://doi.org/10.1111/spc3.12185

Marcus, D. K., Zeigler-Hill, V., Mercer, S. H., & Norris, A. L. (2014). The psychology of spite and the measurement of spitefulness. *Psychological Assessment*, *26*(2), 563–574. https://doi.org/10.1037/a0036039

McDonald, R. P. (1999). *Test theory: A unified treatment*. L. Erlbaum Associates. https://www.taylorfrancis.com/books/mono/10.4324/9781410601087/test-theory-roderick-mcdonald https://doi.org/10.4324/9781410601087

Miller, J. D., & Campbell, W. K. (2008). Comparing clinical and social-personality conceptualizations of narcissism. *Journal of Personality*, *76*(3), 449–476. https://doi.org/10.1111/j.1467-6494.2008.00492.x

Moore, C., Detert, J. R., Klebe Treviño, L., Baker, V. L., & Mayer, D. M. (2012). Why employees do bad things: Moral disengagement and unethical organizational behavior. *Personnel Psychology*, *65*(1), 1–48. https://doi.org/10.1111/j.1744-6570.2011.01237.x

Moshagen, M., & Auerswald, M. (2018). On congruence and incongruence of measures of fit in structural equation modeling. *Psychological Methods*, *23*(2), 318–336. https://doi.org/10.1037/met0000122

Moshagen, M., Bader, M., Zettler, I., & Hilbig, B. E. (2023). The role of impulse and interference control in aversive personality: A comprehensive assessment. *Acta Psychologica*, *239*, 104018. https://doi.org/10.1016/j.actpsy.2023.104018

Moshagen, M., Hilbig, B. E., & Zettler, I. (in press). How and why aversive personality is expressed in political preferences. *Journal of Personality and Social Psychology*.

Moshagen, M., Hilbig, B. E., & Zettler, I. (2018). The dark core of personality. *Psychological Review*, *125*(5), 656–688. https://doi.org/10.1037/rev0000111

Moshagen, M., Zettler, I., & Hilbig, B. E. (2020). Measuring the Dark Core of Personality. *Psychological Assessment*, *32*(2), 182–196. https://doi.org/10.1037/pas0000778

Moshagen, M., Zettler, I., Horsten, L. K., & Hilbig, B. E. (2020). Agreeableness and the common core of dark traits are functionally different constructs. *Journal of Research in Personality*, *87*, 103986. https://doi.org/10.1016/j.jrp.2020.103986

Muris, P., Merckelbach, H., Otgaar, H., & Meijer, E. (2017). The Malevolent Side of Human Nature. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *12*(2), 183–204. https://doi.org/10.1177/1745691616666070

Murphy, R. O., & Ackermann, K. A. (2014). Social value orientation: Theoretical and measurement issues in the study of social preferences. *Personality and Social Psychology Review : An Official Journal of the Society for Personality and Social Psychology, Inc*, *18*(1), 13–41. https://doi.org/10.1177/1088868313501745

Murphy, R. O., Ackermann, K. A., & Handgraaf, M. (2011). Measuring Social Value Orientation. *SSRN Electronic Journal.* Advance online publication. https://doi.org/10.2139/ssrn.1804189

O'Meara, A., Davies, J., & Hammond, S. (2011). The psychometric properties and utility of the Short Sadistic Impulse Scale (SSIS). *Psychological Assessment*, *23*(2), 523–531. https://doi.org/10.1037/a0022400

Paulhus, D. L. (2014). Toward a Taxonomy of Dark Personalities. *Current Directions in Psychological Science*, *23*(6), 421–426. https://doi.org/10.1177/0963721414547737

Paulhus, D. L., & Williams, K. M. (2002). The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, *36*(6), 556–563. https://doi.org/10.1016/S0092-6566(02)00505-6

R Core Team. (2023). *R: A language and environment for statistical computing [*Computer software]. R Foundation for Statistical Computing. https://www.r-project.org/

Reise, S. P. (2012). Invited Paper: The Rediscovery of Bifactor Measurement Models. *Multivariate Behavioral Research*, *47*(5), 667–696. https://doi.org/10.1080/00273171.2012.715555

Ro, E., Nuzum, H., & Clark, L. A. (2017). Antagonism trait facets and comprehensive psychosocial disability: Comparing information across self, informant, and interviewer reports. *Journal of Abnormal Psychology*, *126*(7), 890–897. https://doi.org/10.1037/abn0000298

Rose, L., Lynam, D. R., & Miller, J. D. (in press). Measuring the “Dark” Triad: Comparing the Five-Factor Model Antagonistic Triad Measure to other commonly used self-report instruments. *Assessment*.

Rose, L., Miller, J. D., & Lynam, D. R. (2023). Validation of the Five-Factor Model Antagonistic Triad Measure. *Assessment*, *30*(3), 782–797. https://doi.org/10.1177/10731911211068083

Rose, L., Sleep, C. E., Lynam, D. R., & Miller, J. D. (2022). Welcome to the Jangle: Comparing the Empirical Profiles of the "Dark" Factor and Antagonism. *Assessment*, 10731911221124847. https://doi.org/10.1177/10731911221124847

Rosseel, Y., Jorgensen, T. D., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, E., Hallquist, M., Rhemtulla, M., Katsikatsou, M., Barendse, M., Scharf, F., & Du, H. (2023). *Package ‘lavaan’* [Computer software]. https://cran.r-project.org/web/packages/lavaan/lavaan.pdf

Rudloff, J. P., Hutmacher, F., & Appel, M. (2022). Beliefs about the nature of knowledge shape responses to the pandemic: Epistemic beliefs, the Dark Factor of Personality, and COVID-19-related conspiracy ideation and behavior. *Journal of Personality.* Advance online publication. https://doi.org/10.1111/jopy.12706

Rudloff, J. P., Hutmacher, F., & Appel, M. (2023). Post-truth epistemic beliefs rooted in the Dark Factor of Personality are associated with higher COVID-19 vaccination refusal. *Scientific Reports*, *13*(1), 4254. https://doi.org/10.1038/s41598-023-31079-9

Russell, T. D., Doan, C. M., & King, A. R. (2017). Sexually violent women: The PID-5, everyday sadism, and adversarial sexual attitudes predict female sexual aggression and coercion against male victims. *Personality and Individual Differences*, *111*, 242–249. https://doi.org/10.1016/j.paid.2017.02.019

Russell, T. D., & King, A. R. (2017). Mean girls: PID-5 personality traits and everyday sadism predict hostile femininity. *Personality and Individual Differences*, *104*, 252–257. https://doi.org/10.1016/j.paid.2016.08.020

Satorra, A., & Bentler, P. M. (2010). Ensuring Positiveness of the Scaled Difference Chi-square Test Statistic. *Psychometrika*, *75*(2), 243–248. https://doi.org/10.1007/s11336-009-9135-y

Scholz, D. D., Hilbig, B. E., Thielmann, I., Moshagen, M., & Zettler, I. (2022). Beyond (low) Agreeableness: Towards a more comprehensive understanding of antagonistic psychopathology. *Journal of Personality.* Advance online publication. https://doi.org/10.1111/jopy.12708

Scholz, D. D., Thielmann, I., & Hilbig, B. E. (2023). Down to the core: The role of the common core of dark traits for aversive relationship behaviors. *Personality and Individual Differences*, *213*, 112263. https://doi.org/10.1016/j.paid.2023.112263

Schreiber, A., & Marcus, B. (2020). The place of the "Dark Triad" in general models of personality: Some meta-analytic clarification. *Psychological Bulletin*, *146*(11), 1021–1041. https://doi.org/10.1037/bul0000299

Schrödter, R., Laborde, S., & Allen, M. S. (2021). The dark core of personality and sexism in sport. *Personality and Individual Differences*, *183*, 111119. https://doi.org/10.1016/j.paid.2021.111119

Seuntjens, T. G., Zeelenberg, M., van de Ven, N., & Breugelmans, S. M. (2015). Dispositional greed. *Journal of Personality and Social Psychology*, *108*(6), 917–933. https://doi.org/10.1037/pspp0000031

Simms, L. J., Goldberg, L. R., Roberts, J. E., Watson, D., Welte, J., & Rotterman, J. H. (2011). Computerized adaptive assessment of personality disorder: Introducing the CAT-PD project. *Journal of Personality Assessment*, *93*(4), 380–389. https://doi.org/10.1080/00223891.2011.577475

Sleep, C. E., Crowe, M. L., Carter, N. T., Lynam, D. R., & Miller, J. D. (2020). Uncovering the structure of antagonism. *Personality Disorders: Theory, Research, and Treatment.* Advance online publication. https://doi.org/10.1037/per0000416

Sleep, C. E., Lynam, D. R., Hyatt, C. S., & Miller, J. D. (2017). Perils of partialing redux: The case of the Dark Triad. *Journal of Abnormal Psychology*, *126*(7), 939–950. https://doi.org/10.1037/abn0000278

Somma, A., Borroni, S., Sellbom, M., Markon, K. E., Krueger, R. F., & Fossati, A. (2020). Assessing dark triad dimensions from the perspective of moral disengagement and DSM-5 alternative model of personality disorder traits. *Personality Disorders: Theory, Research, and Treatment*, *11*(2), 100–107. https://doi.org/10.1037/per0000388

Spearman, C. (1927). The Measurement of Intelligence. *Nature*, *120*(3025), 577–578. https://doi.org/10.1038/120577a0

Thielmann, I., Böhm, R., Ott, M., & Hilbig, B. E. (2021). Economic Games: An Introduction and Guide for Research. *Collabra: Psychology*, *7*(1), Article 19004. https://doi.org/10.1525/collabra.19004

Thielmann, I., & Hilbig, B. E. (2023). Generalized Dispositional Distrust as the Common Core of Populism and Conspiracy Mentality. *Political Psychology*, Article pops.12886. Advance online publication. https://doi.org/10.1111/pops.12886

Thomaes, S., Brummelman, E., Miller, J. D., & Lilienfeld, S. O. (2017). The dark personality and psychopathology: Toward a brighter future. *Journal of Abnormal Psychology*, *126*(7), 835–842. https://doi.org/10.1037/abn0000305

Tiwari, S., Moshagen, M., Hilbig, B. E., & Zettler, I. (2021). The Dark Factor of Personality and Risk-Taking. *International Journal of Environmental Research and Public Health*, *18*(16). https://doi.org/10.3390/ijerph18168400

van Laar, S., & Braeken, J. (2022). Caught off Base: A Note on the Interpretation of Incremental Fit Indices. *Structural Equation Modeling : A Multidisciplinary Journal*, 1–9. https://doi.org/10.1080/10705511.2022.2050730

Vize, C. E., Collison, K. L., & Lynam, D. R. (2020). The Importance of Antagonism: Explaining Similarities and Differences in Psychopathy and Narcissism's Relations With Aggression and Externalizing Outcomes. *Journal of Personality Disorders*, *34*(6), 842–854. https://doi.org/10.1521/pedi_2020_34_342

Vize, C. E., Collison, K. L., Miller, J. D., & Lynam, D. R. (2020). The "core" of the dark triad: A test of competing hypotheses. *Personality Disorders: Theory, Research, and Treatment*, *11*(2), 91–99. https://doi.org/10.1037/per0000386

Vize, C. E., & Lynam, D. R. (2021). On the importance of the assessment and conceptualization of Agreeableness: A commentary on "Agreeableness and the common core of dark traits are functionally different constructs". *Journal of Research in Personality*, *90.* https://doi.org/10.1016/j.jrp.2020.104059

Vize, C. E., Miller, J. D., & Lynam, D. R. (2021). Examining the conceptual and empirical distinctiveness of Agreeableness and "dark" personality items. *Journal of Personality.* Advance online publication. https://doi.org/10.1111/jopy.12601

Watters, C. A., & Bagby, R. M. (2018). A meta-analysis of the five-factor internal structure of the Personality Inventory for DSM-5. *Psychological Assessment*, *30*(9), 1255–1260. https://doi.org/10.1037/pas0000605

Weigel, R. H., Hessing, D. J., & Elffers, H. (1999). Egoism: Concept, measurement and implications for deviance. *Psychology, Crime & Law*, *5*(4), 349–378. https://doi.org/10.1080/10683169908401777

Widiger, T. A. (2011). The DSM-5 dimensional model of personality disorder: Rationale and empirical support. *Journal of Personality Disorders*, *25*(2), 222–234. https://doi.org/10.1521/pedi.2011.25.2.222

Widiger, T. A., & Simonsen, E. M. D. (2005). Alternative Dimensional Models of Personality Disorder: Finding a Common Ground. *Journal of Personality Disorders*, *19*(2), 110–130. https://doi.org/10.1521/pedi.19.2.110.62628

World Health Organization. (2023). *International statistical classification of diseases and related health problems*. (11th ed.). https://icd.who.int/

Wright, A. G. C., & Simms, L. J. (2014). On the structure of personality disorder traits: Conjoint analyses of the CAT-PD, PID-5, and NEO-PI-3 trait models. *Personality Disorders: Theory, Research, and Treatment*, *5*(1), 43–54. https://doi.org/10.1037/per0000037

Zeigler-Hill, V., & Noser, A. E. (2018). Characterizing Spitefulness in Terms of the DSM-5 Model of Pathological Personality Traits. *Current Psychology*, *37*(1), 14–20. https://doi.org/10.1007/s12144-016-9484-5

Zettler, I., Moshagen, M., & Hilbig, B. E. (2021). Stability and Change: The Dark Factor of Personality Shapes Dark Traits. *Social Psychological and Personality Science*, *12*(6), 974–983. https://doi.org/10.1177/1948550620953288

Zimmermann, J., Altenstein, D., Krieger, T., Holtforth, M. G., Pretsch, J., Alexopoulos, J., Spitzer, C., Benecke, C., Krueger, R. F., Markon, K. E., & Leising, D. (2014). The structure and correlates of self-reported DSM-5 maladaptive personality traits: Findings from two German-speaking samples. *Journal of Personality Disorders*, 518–540. https://doi.org/10.1521/pedi_2014_28_130

Zimmermann, J., Kerber, A., Rek, K., Hopwood, C. J., & Krueger, R. F. (2019). A Brief but Comprehensive Review of Research on the Alternative DSM-5 Model for Personality Disorders. *Current Psychiatry Reports*, *21*(9), 21–92. https://doi.org/10.1007/s11920-019-1079-z

Table 1

Definitions and example items for the aversive maladaptive trait facets of the AMPD

AMPD Trait Facet	Definition (American Psychiatric Association, 2013)	PID-5 example item (Krueger et al., 2012)
Attention Seeking	Engaging in behavior designed to attract notice and to make oneself the focus of others' attention and admiration.	I like to draw attention to myself.
Callousness	Lack of concern for the feelings or problems of others; lack of guilt or remorse about the negative or harmful effects of one's actions on others.	I don't care about other peoples' problems.
Deceitfulness	Dishonesty and fraudulence; misrepresentation of self; embellishment or fabrication when relating events	I use people to get what I want.
Grandiosity	Believing that one is superior to others and deserves special treatment; self-centeredness; feelings of entitlement; condescension toward others.	I'm better than almost everyone else.
Hostility	Persistent or frequent angry feelings; anger or irritability in response to minor slights and insults; mean, nasty, or vengeful behavior.	I am usually pretty hostile.
Irresponsibility	Disregard for—and failure to honor—financial and other obligations or commitments; lack of respect for—and lack of follow through on—agreements and promises; carelessness with others' property.	I've skipped town to avoid responsibilities.
Manipulativeness	Use of subterfuge to influence or control others; use of seduction, charm, glibness, or ingratiation to achieve one's ends.	Sweet-talking others helps me get what I want.
Suspiciousness	Expectations of—and sensitivity to—signs of interpersonal ill intent or harm; doubts about loyalty and fidelity of others; feelings of being mistreated, used, and/or persecuted by others.	Plenty of people are out to get me.

Table 2

ECV of each AMPD trait explained by D, together with zero-order and unique relationships beyond D of each AMPD trait with SVO (and vice versa for D) in the pilot study

AMPD Trait	*ECV*	β	$\beta_{trait > D}$	$\beta_{D > trait}$	$\Delta R^2_{trait > D}$	$\Delta R^2_{D > trait}$
Attention Seeking	8%** [6%; 11%]	-.07* [-.13; -.01]	.01 [-.05; .06]	-.30** [-.36; -.25]	0.0%	8.6%**
Callousness	42%** [38%; 47%]	-.21** [-.27; -.15]	.01 [-.07; .08]	-.30** [-.37; -.23]	0.0%	4.8%**
Deceitfulness	44%** [39%; 49%]	-.21** [-.27; -.15]	.04 [-.05; .12]	-.33** [-.40; -.25]	0.2%	4.7%**
Grandiosity	40%** [34%; 45%]	-.24** [-.30; -.18]	-.04 [-.12; .04]	-.28** [-.35; -.20]	0.0%	3.2%**
Hostility	22%** [18%; 26%]	-.06* [-.12; .01]	.10** [.04; .16]	-.34** [-.40; -.28]	1.0%**	9.5%**
Irresponsibility	23%** [18%; 29%]	-.13** [-.19; -.06]	.12** [.05; .20]	-.36** [-.42; -.30]	1.3%**	8.6%**
Manipulativeness	34%** [29%; 40%]	-.18** [-.24; -.11]	.07 [-.01; .15]	-.34** [-.41; -.27]	0.5%	6.3%**
Suspiciousness	34%** [28%; 39%]	-.18** [-.24; -.12]	.08* [.01; .15]	-.34** [-.41; -.28]	0.5%*	6.2%**

Note. β indicates a standardized regression coefficient (zero-order effect); $\beta_{trait > D}$ indicates the standardized regression coefficient when controlling for D (i.e., unique effect; vice versa for $\beta_{D > trait}$); ΔR^2 indicates the incremental explained variance over and above the other predictor. Square brackets indicate a 95% confidence interval; *p < .05; **p < .01

Table 3

ECV of each AMPD trait explained by D, together with zero-order and unique relationships beyond D of each AMPD trait with SVO (and vice versa for D) in the main study

AMPD Trait	*ECV*	β	$\beta_{trait > D}$	$\beta_{D > trait}$	$\Delta R^2_{trait > D}$	$\Delta R^2_{D > trait}$
Attention Seeking	30%** [25%; 34%]	-.15** [-.20; -.11]	-.01 [-.07; .04]	-.26** [-.32; -.20]	0.0%	4.8%**
Callousness	49%** [44%;54%]	-.25** [-.30; -.20]	-.12** [-.19; -.05]	-.19** [-.25; -.12]	0.5%**	1.5%**
Deceitfulness	48%** [43%; 53%]	-.23** [-.28; -.19]	-.08* [-.15; -.01]	-.21** [-.28; -.14]	0.2%*	2.0%**
Grandiosity	40%** [35%; 44%]	-.27** [-.32; -.22]	-.15** [-.22; -.09]	-.17** [-.24; -.10]	1.1%**	1.3%**
Hostility	42%** [37%; 47%]	-.20** [-.25; -.15]	-.04 [-.11; .02]	-.24** [-.30; -.18]	0.0%	3.3%**
Irresponsibility	28%** [24%; 33%]	-.14** [-.19; -.09]	.02 [-.04; .09]	-.28** [-.34; -.22]	0.1%	5.1%**
Manipulativeness	41%** [37%; 46%]	-.23** [-.28; -.18]	-.05 [-.12; .01]	-.23** [-.30; -.16]	0.1%	2.2%**
Suspiciousness	37%** [33%; 42%]	-.21** [-.26; -.16]	-.04 [-.11; .04]	-.25** [-.31; -.18]	0.0%	2.8%**

Note. β indicates a standardized regression coefficient (zero-order effect); $\beta_{trait > D}$ indicates the standardized regression coefficient when controlling for D (i.e., unique effect; vice versa for $\beta_{D > trait}$); ΔR^2 indicates the incremental explained variance over and above the other predictor. Square brackets indicate a 95% confidence interval; *p < .05; **p < .01

Figure 1

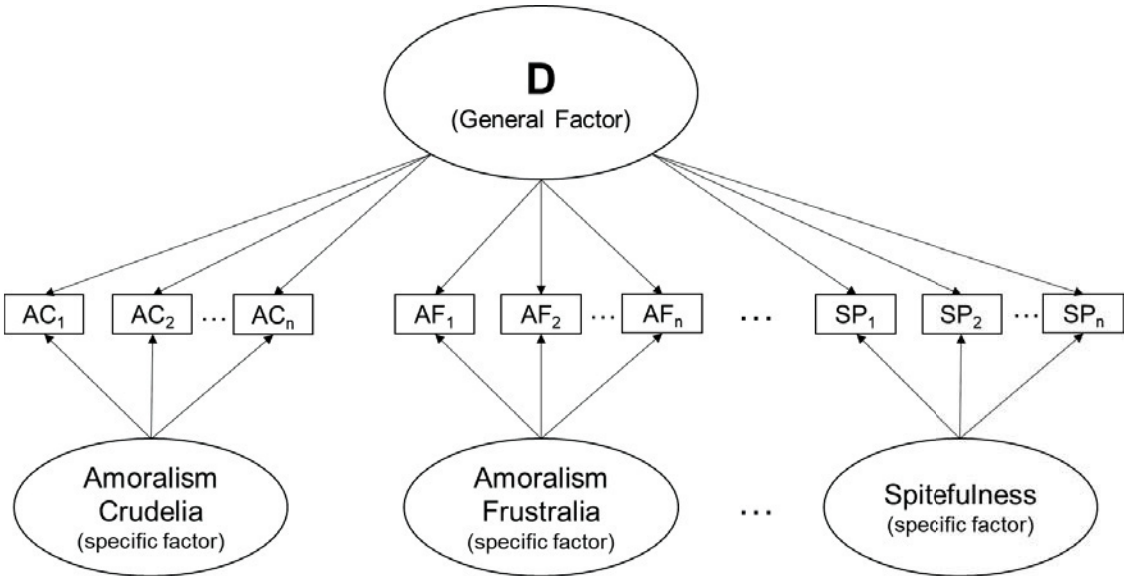
Items of the SVO slider measure

Item 1	self	85	85	85	85	85	85	85	85	85
		○	○	○	○	○	○	○	○	⊗
	other	85	76	68	59	50	41	33	24	15
Item 2	self	85	87	89	91	93	94	96	98	100
		⊗	○	○	○	○	○	○	○	○
	other	15	19	24	28	33	37	41	46	50
Item 3	self	50	54	59	63	68	72	76	81	85
		○	○	○	○	○	○	○	○	⊗
	other	100	98	96	94	93	91	89	87	85
Item 4	self	50	54	59	63	68	72	76	81	85
		○	○	○	○	○	○	○	○	⊗
	other	100	89	79	68	58	47	36	26	15
Item 5	self	100	94	88	81	75	69	63	56	50
		⊗	○	○	○	○	○	○	○	○
	other	50	56	63	69	75	81	88	94	100
Item 6	self	100	98	96	94	93	91	89	87	85
		⊗	○	○	○	○	○	○	○	○
	other	50	54	59	63	68	72	76	81	85

Note. Crosses indicate the choices of a prototypical anti-social player, i.e., assigning positive weights to their own outcome, but negative weights to others' outcomes.

Figure 2

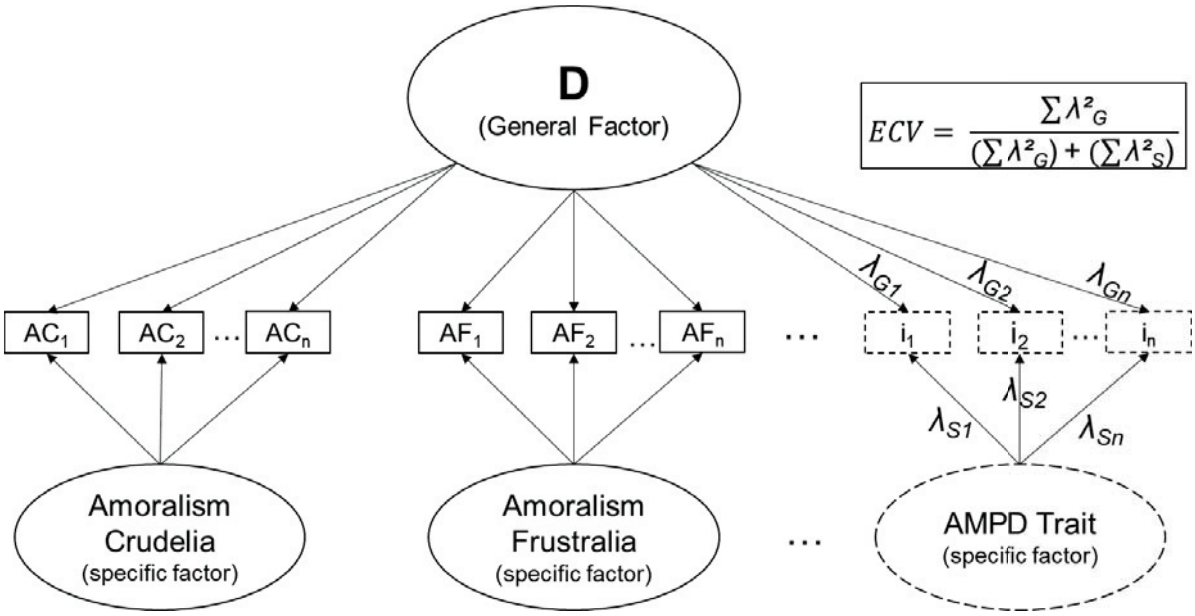
Simplified structure of the bi-factor model for D (i.e., the base model)



Note. Here, D is modeled as the general factor in a bi-factor model of all twelve aversive traits from mainstream personality research with every item loading on D and a respective specific factor. All latent factors were specified orthogonal to each other. Error terms and variances are omitted for the sake of clarity.

Figure 3

Extension of the base model to test Hypothesis 1



Note. The dashed lines indicate that the aversive AMPD trait is not part of the base model. Rather, each aversive AMPD trait was separately introduced into the base model as an additional specific factor for evaluating H1 via the explained common variance (ECV). All latent factors are orthogonal to each other. Error terms and variances are omitted for the sake of clarity.

Disentangling the shared and unique aspects of clinical and subclinical socially aversive traits relevant for (interpersonal) personality dysfunctioning

David D. Scholz[1] and Benjamin E. Hilbig[1]

[1] Department of Psychology, University of Kaiserslautern-Landau (RPTU), Germany

Author Note

David D. Scholz https://orcid.org/0000-0003-1491-5247

Benjamin E. Hilbig https://orcid.org/0000-0002-1470-1882

Data used for Study 1 stem from the Prosocial Personality Project (PPP) and thus may partly overlap with other studies using this data (see https://osf.io/m2abp/ for an overview). Data used in Study 2 stem from the same data set as used by Scholz et al. (in press) and thus show a partial overlap. For details on that as well as access to all materials, data, and analysis scripts, see the files accompanying this manuscript on the Open Science Framework (OSF; https://osf.io/p79bm/?view_only=01e394c7ec8c4165ae7d176d6a794474)

The authors declare that there were no conflicts of interest with respect to the authorship or the publication of this article. This research was supported by a grant from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - GRK 2277 "Statistical Modeling in Psychology".

Correspondence concerning this article should be addressed to David D. Scholz, Department of Psychology, RPTU Kaiserslautern-Landau, Campus Landau, K Building, 1st floor, Fortstraße 7, D-76829 Landau, Germany. Email: david.scholz@rptu.de

Abstract

Typically, socially and/or ethically aversive traits from clinical and broad personality research overlap to a large degree. For the latter however, the association with (Interpersonal) Personality Dysfunctioning (IPD) is understudied. Moreover, it is also unclear to which extent the association of aversive traits with IPD is (mainly) due to their shared vs. unique aspects. We investigate these questions based on a theoretical framework that comprehensively describes the shared variance of all aversive traits. To this end, we concurrently measured 20 aversive traits from clinical and broad personality research together with their common core. Results from five studies (four of them preregistered, total *N*=4,847) revealed that all aversive traits are associated with IPD and that most do so (only) due to their common core. Only three traits offered additional aspects beyond the common core relevant for IPD. The results inform debates about whether to include more traits into the AMPD.

Disentangling the shared and unique aspects of clinical and subclinical socially aversive traits relevant for (interpersonal) personality dysfunctioning

Whereas prosociality is key to interpersonal relationships and societal function, there is notable variation between individuals in prosocial vs. antisocial tendencies. Within clinical psychology, extreme variants of such tendencies are described by dimensional models of personality disorders, as exemplified by the Alternative Model for Personality Disorders (AMPD) in the DSM-5 (Section 3; American Psychiatric Association, 2013) or the ICD-11 (World Health Organization, 2023).

In the AMPD in particular, the *severity* of a personality disorder is assessed by the level of personality (dys)functioning (Criterion A), manifested in problems with the sense of the self and interpersonal problems. In addition, the *style* of a personality disorder (Criterion B) is described by 25 so-called maladaptive trait facets. For these, an exploratively derived higher-order structure has been proposed, mapping these 25 trait facets onto five higher-order domain factors (Krueger et al., 2012). To describe socially and/or ethically aversive individuals, there are a number of aversive AMPD traits (most prominently, although not exclusively, from the Antagonism domain), which – together with Criterion A – may result in the diagnoses of a (socially aversive) personality disorder, e.g., antisocial or narcissistic personality disorder. Although there is an ongoing debate on the distinctiveness of the two criteria (see Sharp & Miller, 2022), there is also broad consensus that Criterion A represents personality impairment, which sets personality disorders apart from other forms of psychopathology (Morey et al., 2020; Morey et al., 2022; Sharp, 2022; Sharp & Wall, 2021; Wright & Ringwald, 2022; Wright et al., 2022); cf. Sleep & Lynam, 2022). Similar notions can also be found in the ICD-11 (Zimmermann et al., 2020). By implication, if one is interested in the relationship of traits with personality disorders, one needs to understand how much they are associated with personality impairment (i.e., Criterion A). Given that socially aversive traits, first and foremost, imply negative consequences for others (e.g., manipulating others to achieve one's end), these traits are especially relevant for *Interpersonal Personality Dysfunctioning* (IPD), which will be the main focus of the current research (rather than self-related dysfunctioning).

However, when it comes to socially and/or ethically aversive tendencies, there is also a large, independent research tradition situated within basic personality research. Therein, many such traits – typically coined as "dark"[1] traits – have been proposed as subclinical expressions of certain tendencies (Paulhus & Williams, 2002), with some of them (e.g., Narcissism or Psychopathy) directly adopted from clinical research (Furnham et al., 2013). These traits are thought of as subclinical in the sense of varying notably within non-clinical populations and correspondingly related to "everyday life outcomes", e.g., work behavior, education, or mating behavior (see Furnham et al., 2013).

The most prominent set of these subclinical aversive traits is the so-called "*Dark" Triad*, i.e., Narcissism, Psychopathy, Machiavellianism (Paulhus & Williams, 2002), but many more such aversive traits have been proposed, e.g., Sadism, Spite, or Greed (Marcus & Zeigler-Hill, 2015; Moshagen et al., 2018). In what follows, we refer to these traits as *aversive subclinical traits*, though the term "subclinical" is only meant to denote their research tradition outside of clinical psychology but *not* that these traits cannot be related to psychopathology.

Despite common roots, these two research traditions (aversive clinical vs. subclinical research) are notably isolated, leaving the role of aversive subclinical traits for personality psychopathology understudied (Hilbig et al., 2020; Miller & Campbell, 2008; Thomaes et al., 2017). This is problematic for several reasons, empirical and conceptual ones: Previous research has indicated notable associations between some (aversive) AMPD and subclinical traits, mainly traits from the "Dark" Triad (Miller et al., 2013; Rose et al., 2023; Scholz et al., 2022; Sleep et al., 2017; Vize et al., 2020; Rose et al., in press). Consequently, aversive subclinical traits – despite their definition as subclinical – appear relevant for psychopathology, at the very least bringing "individuals at risk for maladjustment, including psychopathology" (Thomaes et al., 2017, p.4).

Seeking theoretical integration of these different traditions and accounting for the fact that aversive AMPD traits and subclinical traits are strongly associated – both within each of these groups

[1] The term „dark" has been criticized as vague, potentially stigmatizing, and problematic for other reasons. In essence, "dark" traits are socially and/or ethically aversive traits (Thomaes et al. 2017; Moshagen et al., 2018). In the following, we will only maintain the term "dark", when it is important to link our reasoning the previous research such as the "Dark" Triad. Otherwise, we use the term *aversive subclinical trait(s)*.

and across – Scholz et al. (in press) demonstrated that this overlap can be attributed to a single, broad underlying disposition. Specifically, they based these considerations on the theoretical framework of the D-Factor of Personality (D; Moshagen et al., 2018)[2], defined as the tendency to "the general tendency to maximize one's individual utility – disregarding, accepting, or malevolently provoking disutility for others – , accompanied by beliefs that serve as justifications" (p. 657). D is conceptualized as the common core of *all* socially/ethically aversive traits and Scholz et al. (in press) correspondingly found that all of the investigated aversive traits – eight from the AMPD and twelve subclinical traits – can indeed be understood as manifestations of D. In other words: any trait (from either research tradition) encompassing aversive tendencies can be viewed as a *flavored* manifestation of D which means that each trait encompasses at least some aspect(s) of the definition of D and may also comprise additional (non-aversive) aspects beyond D. For example, Psychopathy is a manifestation of D *flavored* by disinhibition which is its unique component beyond D (Bader et al., 2023).

Although aversive AMPD and aversive subclinical traits can thus be described as manifestations of a single common core[3], it remains unclear whether and to which extent their common core vs. their unique aspects (i.e., their 'flavors' so to speak) contribute to the explanation of personality dysfunctioning. Identifying shared vs. unique variance of aversive traits in accounting for IPD may inform ongoing debates about whether to include more traits in the AMPD (Clark & Watson, 2022). Also, it has been noted that the AMPD traits overlap to a large degree in general, thus raising questions about the added value any one of them may possess (Sleep et al., 2018;

[2] The label "D-Factor" stems from its development within the research tradition coined as "dark" personality research. As stated above, we acknowledge that in the context of clinical psychology research, the term "dark" may be perceived as stigmatizing. For the lack of a more appropriate label in clinical contexts and for the sake of clarity and consistency, we use the label "D-Factor" or D for short.

[3] As noted in Scholz et al. (in press), there is a debate whether D, as the common core of all aversive traits, should be equated with Antagonism (or low Agreeableness) on a *substantive* level. For the current study, however, this question is largely irrelevant, as we concentrate *conceptual* framework of D, defining aversive traits as *flavored manifestations* of a common core (independent of its name or label). Under certain additional assumptions one could translate such a framework to what has been called Antagonism (see Scholz et al., in press), however, this has not been spelled out theoretically and is entirely beyond the scope of the current research.

Ringwald et al., in press). As such, it is all the more crucial to determine the relevant aspects of aversive traits for their relation with IPD as parsimoniously as possible – thus starting with what they all share and (only) then adding any additional aspects if necessary.

Moreover, there is only sporadic work in linking aversive subclinical traits, or D for that matter, to different outcomes of psychopathology, e.g., anxiety (Rose et al., 2023), depressivity (Gómez-Leal et al., 2019), personality disorders (Hilbig et al., 2020), or general well-being (Blasco-Belled et al., 2023). Although findings are compatible with the general notion that aversive subclinical traits place "individuals at risk for maladjustment, including psychopathology" (Thomaes et al., 2017, p. 4), studies are scarce and often limited to small sets of traits. More importantly, such research has not directly addressed the severity of personality psychopathology, i.e., personality dysfunction (Criterion A). This is surprising given that the entire logic of maintaining separate research traditions for subclinical vs. clinical traits clearly hinges on the assumption that the latter are more strongly linked to personality dysfunction than the former (see Jones & Paulhus, 2023; Morey et al., 2020).

Against this background, we investigated how much of the shared vs. unique aspects of every manifestation of D (AMPD or subclinical trait) account for IPD, therefore testing the incremental value (or possible redundancies) across all aversive traits. To this end, we conducted five studies: an explorative pilot study (Study 1) and four preregistered studies (Study 2 and Studies 3a-c). As the role of each individual aversive subclinical trait with IPD, let alone the contribution of any aversive trait over and above their common core has not yet been studied, we first chose an explorative approach in Study 1.

Transparency and openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the manuscript. All data, analysis code, and research materials for all five studies are available at the Open Science Framework (OSF; https://osf.io/p79bm/?view_only=01e394c7ec8c4165ae7d176d6a794474). Data was analyzed using

R, version 4.3.1 (R Core Team, 2024) and the package lavaan, version 0.6-16 (Rosseel et al., 2023). Study 1 was not preregistered. The hypotheses and analyzes of Study 2 and Studies 3a-c were preregistered (see Study 2: https://aspredicted.org/PTN_L91; Study 3a: https://aspredicted.org/KWH_C91; Study 3b: https://aspredicted.org/CKF_FPJ; Study 3c: https://aspredicted.org/2GY_VRD). All studies were conducted in full accordance with the Ethical Guidelines of the American Psychological Association (APA) and were run based on approval from the local ethics committee (*blinded for peer-review*). Informed consent was obtained digitally from all participants in all studies.

Study 1

Method

Sample

Data used for the Study 1 stem from the Prosocial Personality Project (PPP). The PPP covers various aspects of pro- and antisocial tendencies through six consecutive primary waves, plus additional follow-up waves. For a full documentation regarding the PPP see the OSF (https://osf.io/m2abp/?view_only=8d8ec046188f41d0a92699d7484b8a37). The current data stem from Wave 1 (T1; November-December 2019), Wave 2 (T2; January 2020), Wave 3 (T3; January-February 2020), follow-up wave 2020-5b (T7b; May 2020) and follow-up wave 2021-1b (T9b; January-February 2021) of the PPP. The overall sample size for the present analyzes is *N*=604 (323 male, 281 female; aged 19-66 years, *M*=45.9, *SD*=11.9). Participants had diverse educational backgrounds, with 35.6% holding a general certificate of secondary education (German: Realschule), 29.8% holding a vocational diploma or university-entrance diploma (German: Fachabitur or Abitur), and 33.4% holding a university/college degree (1.2% holding other degrees). Similarly, the income of participants was quite heterogenous, with 34.6% earning less than 1,500€, 40.2% earning between 1,500€ and 3,000€, 13.9% earning between 3,000€ and 5,000€, and 2.5% earning more than 5,000€ a month (8.8% preferred not to say).

Measures & Procedure

D was measured at T1 with the German item set suggested by Moshagen et al. (2020). For the current analyzes, we chose the 16-item version (D16), a short version of the full 70 items scale, to ensure a more comparable number of items indicating each construct.

The following twelve aversive subclinical traits were measured at T3 or T7b (average time lag between T3 and T7b=3 months) by 118 items in total: Crudelia, Egoism, Frustrallia, Greed, Machiavellianism, Moral Disengagement, Narcissism, Psychological Entitlement, Psychopathy, Sadism, Self-Centeredness, and Spitefulness (for details about the specific scales and their translation used herein, see the OSF repository; https://osf.io/p79bm/?view_only=01e394c7ec8c4165ae7d176d6a794474). The traits were chosen to comprehensively cover all facets of D and ensure consistency with previous research on D (Moshagen et al., 2018; Moshagen et al., 2020). Also, they exhibit satisfactory psychometric properties, including in their German versions employed in the present data set (for details see Zettler et al., 2021). D and the aversive subclinical traits were measured on a five-point Likert-type scale.

AMPD traits were measured at T2 with the German version (Zimmermann et al., 2014) of the Personality Inventory for DSM-5 Short Form (PID-5-SF; Maples et al., 2015), thus with four items each. All AMPD traits, were measured on a four-point Likert-type scale as is common.

IPD was measured with the respective subscale of the The Level of Personality Functioning Scale-Brief Form 2.0 (LPFS-BF 2.0; Weekers et al., 2019) in German (Spitzer et al., 2021), thus with six items on a five-point Likert-type scale. The LPFS-BF 2.0 was chosen as it is commonly used for assessing Criterion A (Morey et al., 2022) and showing good psychometric properties (Morey et al., 2022; Weekers et al., 2019), also in the German version (Spitzer et al., 2021), with a considerably smaller number of items as the full scale.

Analyzes & Inference Criteria

All analyzes are based on structural equation modeling (SEM), using the R package “lavaan” (Rosseel et al., 2023), employing robust ML-estimation and a Satorra-Bentler scaled test statistic.

Raw item scores were used for all analyzes. All latent factors were identified by fixing their variance to one, except for IPD as the outcome, for which we choose a reference indicator approach for model identification.

To test whether and how strongly aversive subclinical and aversive AMPD traits are related to IPD, we estimated one latent model, in which all traits were allowed to correlate. All traits were modeled as single factors. Correlation coefficients were compared using the Fisher's Z procedure as recommended by Hittner et al. (2003).

To test whether any of the traits contributes incremental variance over and above D in accounting for IPD, we ran three latent regressions for each trait. The first and second regression included either D or the aversive trait in question as a predictor of IPD, respectively, so as to obtain zero-order effects. The third model predicted IPD by both D and the aversive trait in question simultaneously, thus allowing to compute incremental effects. Specifically, we calculated what proportion of the total explained variance in the third model (i.e., what is explained both by D *and* the aversive trait in question), is already accounted for by D alone, i.e., $R^2_{D}/R^2_{D+\text{aversive trait}}$.[4]

As supplementary analyzes, we further calculated the incremental variance (ΔR^2) of each trait in the prediction of IPD over and above D. Also, we conducted all analyzes with the self-subscale and the overall score of the LFPS-BF 2.0. As these are, however, only secondary analyzes, we will not discuss them in detail.

Results

For descriptive statistics, observed correlations, and internal consistency estimates (Cronbach's α) of all scales, as well as, exact model fits statistics (median RMSEA=.072, SRMR=043, ω=.82), see Supplementary Table 1 and 2 on the OSF.

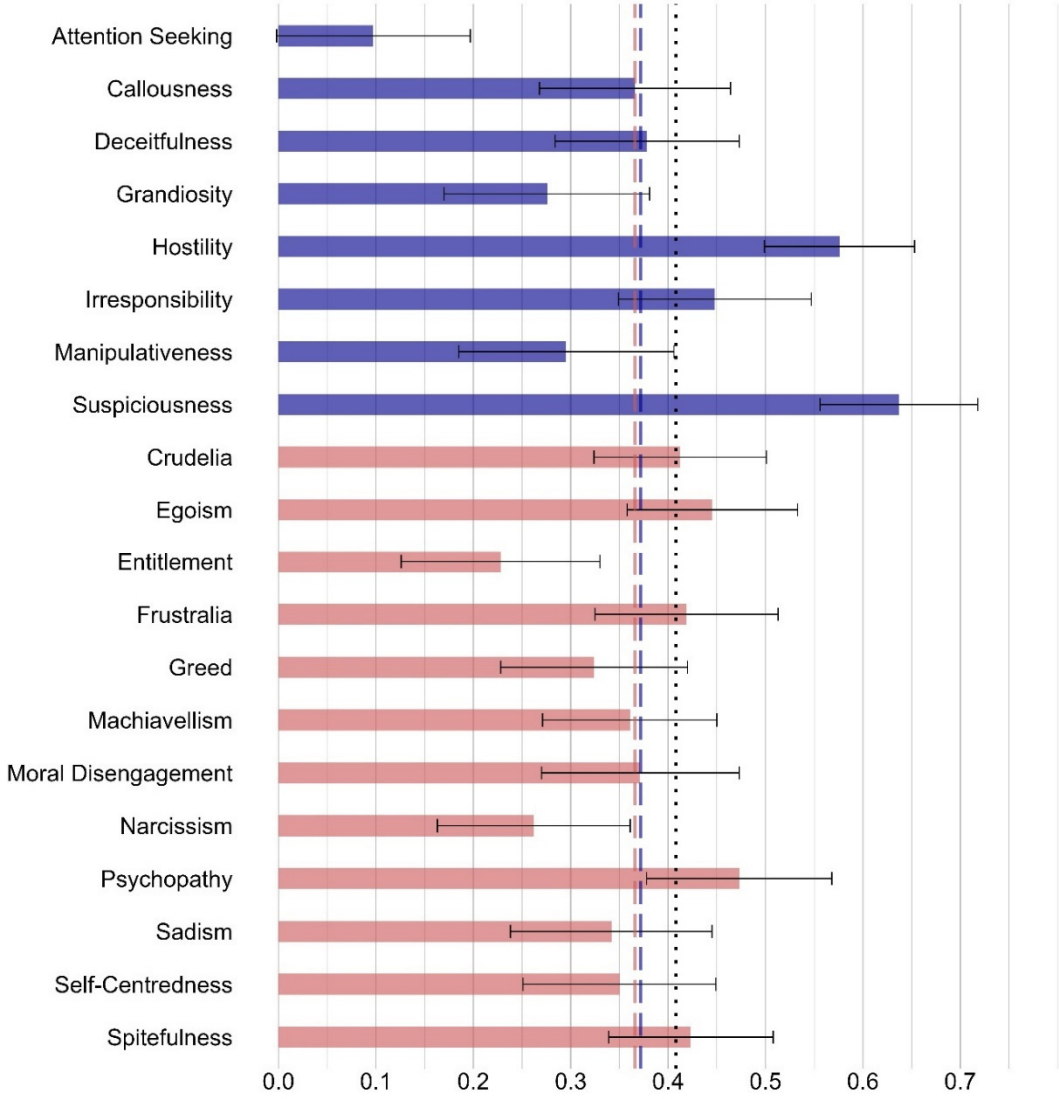
Regarding the latent zero-order correlations of all traits with IPD see Figure 1. Overall, all traits were significantly related to IPD, with the exception of Attention Seeking yielding the weakest,

[4] Due to item overlap between D and some aversive subclinical traits, we conducted additional analyzes in which the respective items were omitted from the measurement model of D. All analyzes lead to the same conclusions.

and the only non-significant, association (r=.10, p=.056, 95%-CI [-.00;.20]). Suspiciousness, in turn yielded the strongest association (r=.64, p<.001, 95%-CI [.56;.72]), being the only aversive trait to correlate significantly stronger with IPD than D (r=.41, p<.001, 95%-CI [.32;.50]). Moreover, on average, aversive AMPD and aversive subclinical traits were similarly related to IPD, both median r=.37, with no significant difference between trait groups, $|\Delta r|$=.01, p=.881, 95%-CI [-.08;.07]. More importantly, turning to the explained variance in IPD, analyzes showed that for 17 out of the 20 aversive traits, D already accounted for most of the total variance explained (>80%) in IPD (see Figure 2), the only exceptions being Hostility (53%), Irresponsibility (71%), Suspiciousness (46%).

Figure 1

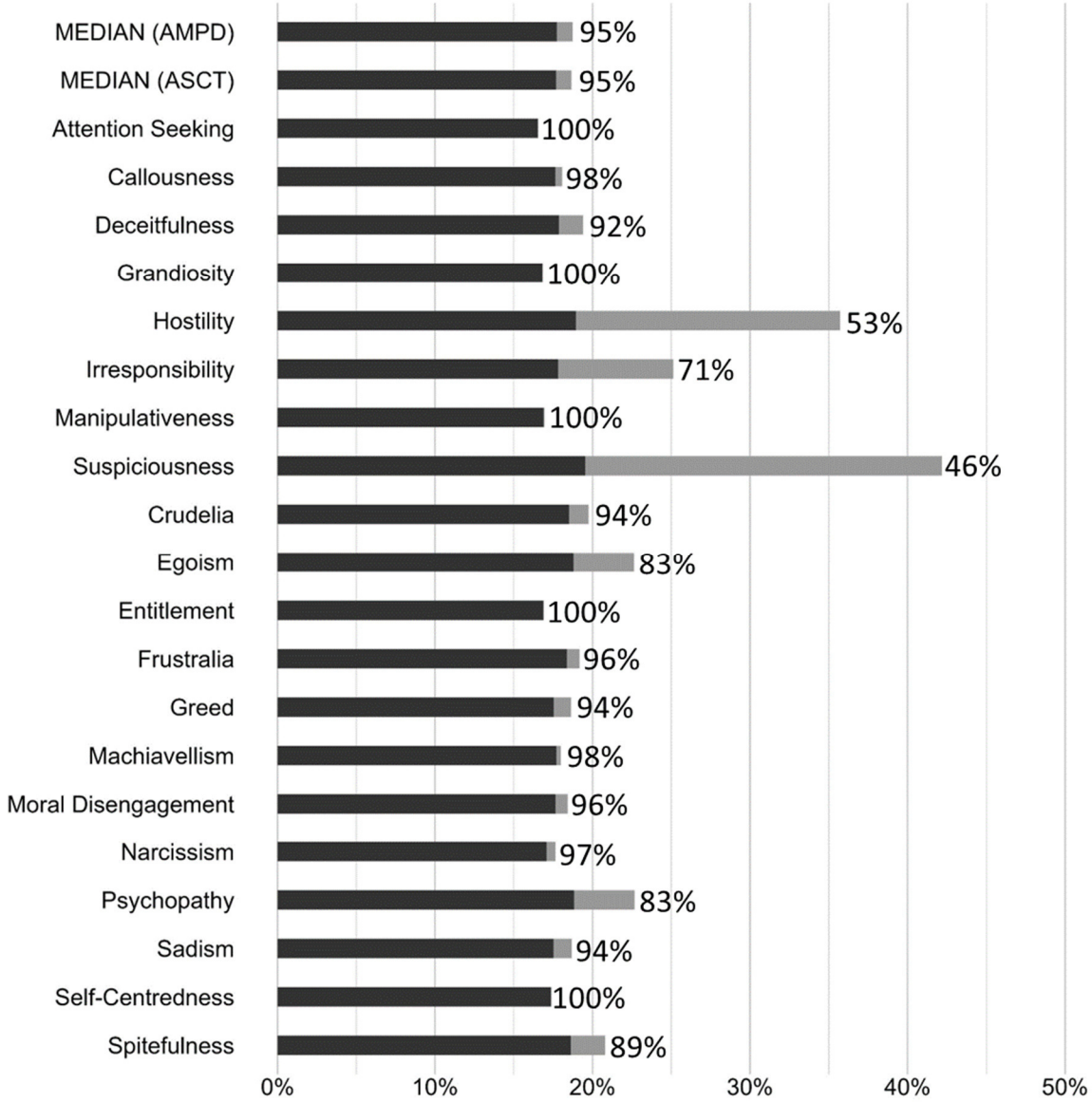
Correlations of all traits with IPD in Study 1



Notes. AMPD traits in dark blue; aversive subclinical traits in light red; dashed lines indicate median correlations across AMPD traits (dark blue) and subclinical (light red) traits with IPD, respectively; the black dotted line indicates the correlation of D with IPD; error bars indicate 95%-CI

Figure 2

Explained variance in IPD already accounted for by D for each aversive trait in Study 1



Notes. The black area of a bar indicates the explained variance of IPD by D (R^2_D), the total bar indicates the total variance explained by D and the respective aversive trait ($R^2_{D+aversive\ trait}$) and thus the light grey area indicates the added variance explained by the respective aversive trait. Percentages at the end indicate the proportion of the total explained variance, which is already accounted for by D alone (R^2_D / $R^2_{D+aversive\ trait}$); AMPD=AMPD traits; ASCT=aversive subclinical traits

Regarding the supplementary analyzes (see for details Supplementary Table 3 on the OSF), the analyzes of the incremental variances (ΔR^2) lead the same conclusion, i.e., only Hostility, Irresponsibility, and Suspiciousness explained small to moderate incremental variance over and above D (ΔR^2>4%). Moreover, using the self-subscale and overall score of the LFPS-BF 2.0 revealed that D covered less of the total variance explained, confirming the general notion that the unique,

non-aversive aspects of aversive traits (beyond D) are more relevant for dysfunction related to the self than for IPD.

Discussion of Study 1

Taken together, the results suggest that (i) there is no inherent difference between aversive AMPD vs. aversive subclinical traits in their relation with IPD, (ii) the association of most traits (AMPD and subclinical) with IPD is due to (can be accounted for by) their common core D, and (iii) some traits – Hostility, Irresponsibility, and Suspiciousness – do explain IPD over and above the common core of all aversive traits and thus seem to encompass unique aspects beyond D that are relevant for IPD. Nevertheless, these results are limited in several ways. First, variables were measured at different time points, with some being temporarily measured closer to IPD than others (e.g., six months between Spitefulness and IPD vs. over one year between D and IPD). On the one hand, this hints at longitudinal effects of the aversive traits on IPD and alleviates concerns of common method variance due to consistent responding. On the other hand, this may have biased the effects of the current analyzes, especially when comparing the effects of traits; arguably, for a fairer comparison, all predictors should ideally be measured at (roughly) the same time and thus with comparable temporal distance from the criterion. Second, the aversive AMPD traits were measured with the short version of the PID, i.e., PID-5-SF (Maples et al., 2015). Thus, they were indicated by a notably smaller number of items than D and some of the aversive subclinical traits which may have led to an underestimation of their (unique) effects on IPD. Finally, hypotheses and analyzes were not preregistered and explorative in nature.

To remedy these limitations and to notably increase confidence in the findings, we therefore conducted a preregistered Study 2 (see https://aspredicted.org/PTN_L91). Based on the results of Study 1, we expected that aversive AMPD and aversive subclinical traits are (on average) similarly correlated with IPD (H1). Moreover, we hypothesized that for all aversive traits, D again would account for most (>80%) of the total variance explained in IPD, with the possible exceptions for Hostility, Irresponsibility, and Suspiciousness (H2). Moreover, we also chose Psychopathy as a

possible exception, as it has been decomposed into a manifestation of D flavored by Disinhibition (Bader et al., 2023) and also showed a noticeable proportion of total variance explained in IPD over D in the current study.

Study 2

Method

Sample

We relied on a dataset, to some parts already used by Scholz et al. (in press), thus there is a partial data overlap (see OSF for details). The sample was collected and financially remunerated through a professional online panel provider. The minimum age for participation was 18 years. Only German citizens with a sufficient level of German were allowed to take part in the study. To ensure an approximately representative distribution of age and gender in the sample, quotas were used during recruiting. We used the same exclusion criteria as Scholz et al. (in press). The final sample size was *N*=2,006 (995 female, 1006 male, 5 diverse; aged 18–74 years, *M*=47.9, *SD*=15.1). Other than age and gender, no further demographic data was collected.

Procedure

Data collection was done at two separate measurement occasion (between May and June 2023) to minimize the impact of consistent responses, as well as the effects of fatigue and boredom. The average time gap between these measurement occasions was seven days. Participants were provided with identical instructions as those in Study 1 for all measures.

Measures & Analyzes

To maximize consistency, we relied on the D16 to measure D, and used the same scales for the twelve aversive subclinical traits and IPD as in Study 1. However, unlike in Study 1, the full version of the PID-5 (Krueger et al., 2012) was used to measure the aversive AMPD traits (Attention Seeking, Deceitfulness, Callousness, Hostility, Irresponsibility, Grandiosity, Manipulativeness, & Suspiciousness). D and the aversive subclinical traits were measured at the first measurement occasion, while the aversive AMPD traits and IPD were measured at the second measurement

occasion. All analyzes of Study 2 were preregistered[5] and were conducted exactly as in Study 1, using the same statistical approaches and the same criteria to evaluate the hypotheses.

Results

For descriptive statistics, observed correlations, and internal consistency estimates of all scales, as well as, exact model fits statistics (median RMSEA=.081, SRMR=047, ω=.85) see Supplementary Table 4 & 5 on the OSF.

Regarding the latent correlations of all traits with IPD see Figure 3. Again, all aversive traits were associated with IPD, this time also including Attention Seeking. Also, once again, subclinical and AMPD traits were similarly related to IPD with median r=.53 for AMPD and median r=.42 for subclinical traits. However, contradicting H1, there was a significant, non-trivial difference between the two groups of traits, $|\Delta r|$=.11, p<.001, 95%-CI [.07;.15], indicating that aversive AMPD traits are, on average, associated somewhat more strongly to IPD than aversive subclinical traits.

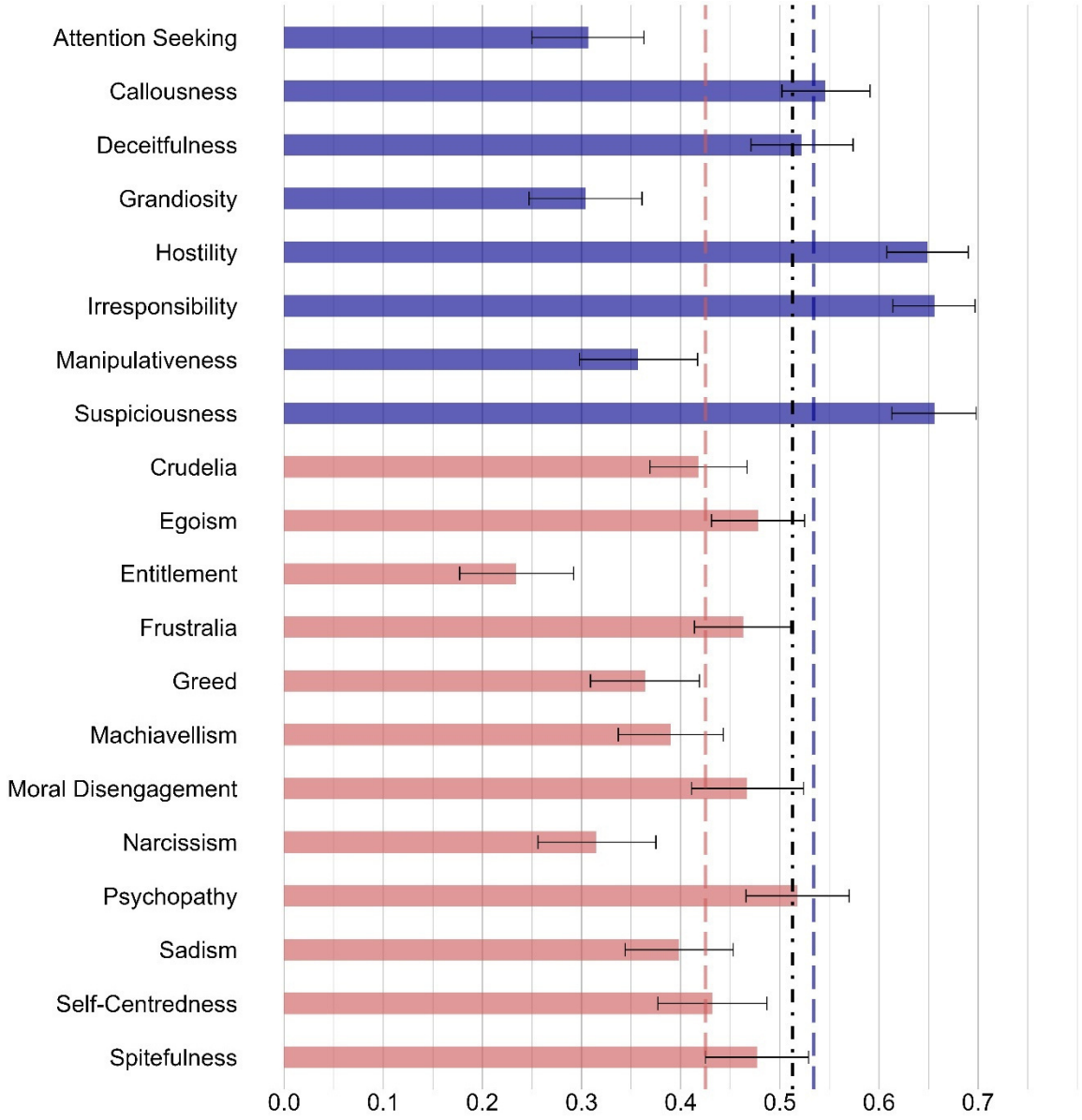
In line with H2, results again revealed that D accounted for a large proportion of the total variance explained (>80%) in 17 out of 20 aversive traits (see Figure 4). Indeed, the proportion of total variance explained in IPD attributable to D across all traits was at least 93%, the only exceptions, in line with the results from Study 1, being Hostility (73%), Irresponsibility (64%), and Suspiciousness (70%).

Again, the supplementary analyzes (see Supplementary Table 6 on the OSF) regarding the incremental contribution of each aversive trait over and above D, were in line with the general picture that only Hostility, Irresponsibility, and Suspiciousness yield meaningful variance beyond D (ΔR^2>2%) that is relevant for IPD. Moreover, the effects were again smaller for the self-subscale and the general personality dysfunctioning score.

[5] Only the analysis omitting the item overlap of D with some of aversive subclinical traits was not preregistered. We decided to additionally include this again to ensure the robustness of the effects of interest while accounting for item overlap. Crucially, these analyzes lead to the same conclusions.

Figure 3

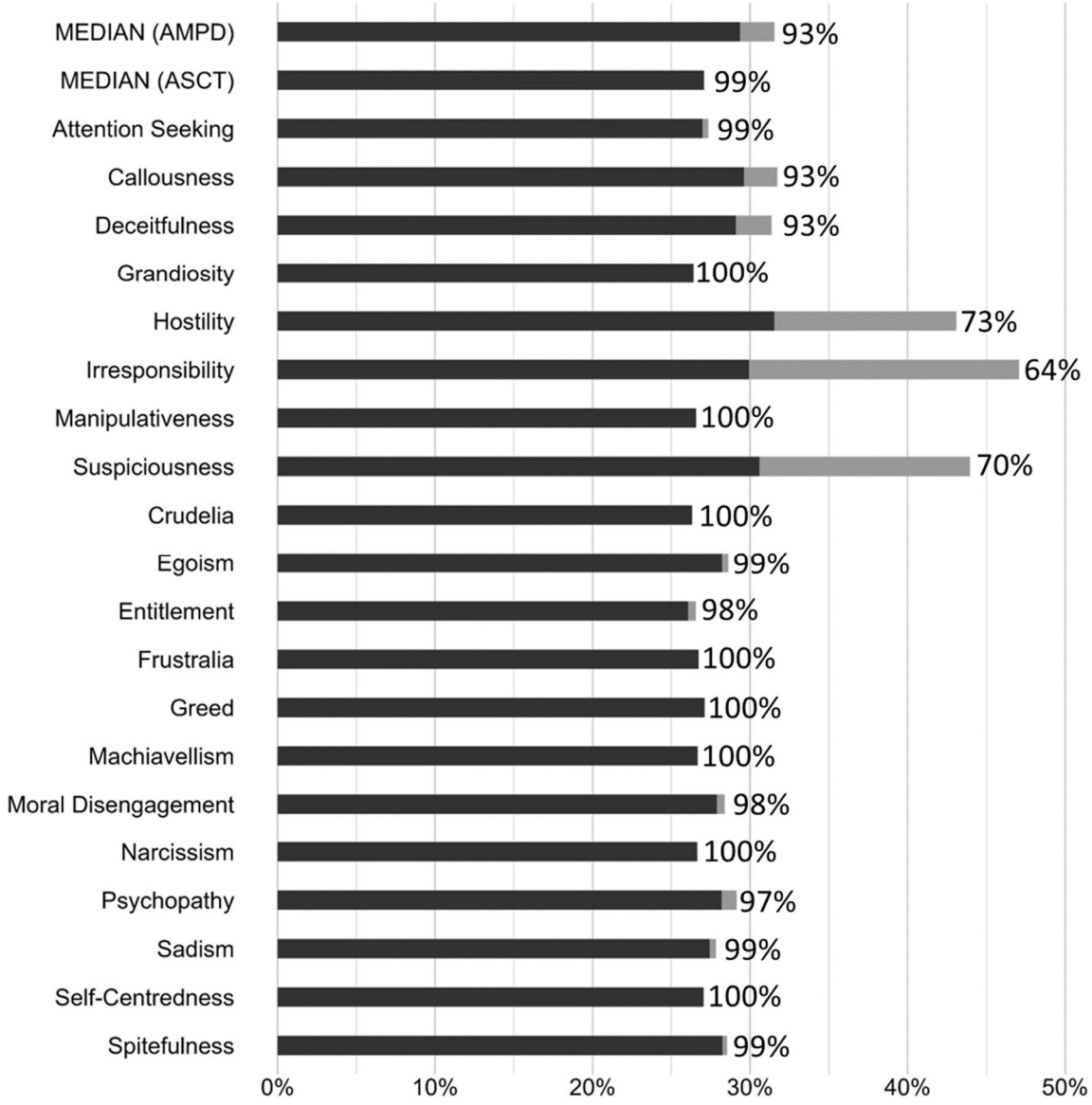
Correlations of all traits with IPD in Study 2



Notes. AMPD traits in dark blue; aversive subclinical traits in light red; dashed lines indicate median correlations across AMPD traits (dark blue) and subclinical (light red) traits with IPD, respectively; the black dotted line indicates the correlation of D with IPD; error bars indicate 95%-CI

Figure 4

Explained variance in IPD already accounted for by D for each aversive trait in Study 2



Notes. The black area of a bar indicates the explained variance of IPD by D (R^2_D), the total bar indicates the total variance explained by D and the respective aversive trait ($R^2_{D+aversive\ trait}$) and thus the light grey area indicates the added variance explained by the respective aversive trait. Percentages at the end indicate the proportion of the total explained variance, which is already accounted for by D alone (R^2_D / $R^2_{D+aversive\ trait}$); AMPD=AMPD traits; ASCT=aversive subclinical traits

Importantly, we also exploratively investigated whether the predictive performance of aversive subclinical and aversive AMPD traits can be attributed to different suitability of the underlying item-sets to differentiate at particular positions along a latent trait continuum. Therefore, we estimated item response models via the R-package *mirt* (Chalmers et al., 2023) to obtain normalized test-information functions for the aversive subclinical trait and aversive AMPD trait

items. Results revealed remarkable similarity in the test information functions for both sets of items (see Supplementary Figure 1 on the OSF), indicating that these item-sets provide maximal information at roughly equivalent levels along the latent trait spectrum. On the one hand, this once more confirms how indistinct clinical and subclinical aversive traits actually are. On the other hand, it does suggest that the small difference found between these two groups of traits in how strongly they are associated with IPD cannot be attributed to differences in item difficulty.

Discussion of Study 2

Overall, the results of Study 2 are mostly in line with Study 1. The single exception, also contrary to the preregistered hypothesis (H1), was that this time, the zero-order association between IPD and aversive AMPD vs. aversive subclinical traits did significantly differ. Even though the difference amounted to a small effect size only, it does suggest that AMPD vs. subclinical aversive traits are not entirely interchangeable. Whereas this relatively small difference could not be explained by differences in item characteristics, it may be due to the fact that AMPD traits and IPD were measure simultaneously, but subclinical traits a week earlier.

Fully in line with Study 1, D accounted for a large proportion of the total variance explained in IPD for 17 out of 20 traits, with only three exceptions, i.e., Hostility, Irresponsibility, and Suspiciousness – exactly as hypothesized. Indeed, the proportion of total variance explained in IPD attributable to D was even larger in size than in Study 1, which may result from the now temporally closer assessment of all variables.

Given the finings so far, it can be concluded that the common core of all aversive traits accounts for most of the association between aversive traits – clinical and subclinical – with IPD. The consistent (relative) exceptions are Hostility, Irresponsibility, and Suspiciousness which appear to subsume at least some aspects beyond D relevant for IPD. In terms of the D-theory, these traits appear to entail a notable “flavor” that is unique to each and beyond the scope of D. However, so far, it remains entirely unclear what exactly these flavors may be.

Therefore, we conducted three more (preregistered) studies, one for each trait respectively, i.e., Study 3a for Hostility (see https://aspredicted.org/KWH_C91), Study 3b for Irresponsibility (see : https://aspredicted.org/CKF_FPJ) , and Study 3c for Suspiciousness (see: https://aspredicted.org/2GY_VRD), to decompose the traits into D and their unique flavors. Given that Hostility shows considerable cross-loadings with the Negative Attachment domain in the AMPD, we hypothesized that Hostility is a manifestation of D, additionally flavored by a lack of strategies for *Emotion Regulation*. For Irresponsibility, given that it primarily loads on the Disinhibition domain in the AMPD, we expected that it is flavored by (a lack of) *Self-control*. Finally, Suspiciousness is a key criterion for the paranoid personality disorder (American Psychiatric Association, 2013). As such, we expected that, beyond D, it is flavored by *Delusional Paranoid Thoughts* including the experience of supernatural phenomena, e.g., a sense that one's own thoughts are controlled by others, and being persecuted by (possibly unknown) others.

Studies 3a-c

Method

Participants & Procedure

Participants were recruited via darkfactor.org, a (free and fully anonymous) website offering self-assessments and individual feedback on D. After participants completed the full 70item version of D (D70; Moshagen et al., 2020) they were asked if they would be willing to answer a few more questions. Participants were assured that they would receive feedback regarding their D-score independent of whether they are willing to answer the additional questions. Those who agreed were presented with the respective additional measures for Study 3a (time of assessment: calendar week 29 and 30), for Study 3b (time of assessment: calendar week 30), or for Study 3c (time of assessment: calendar week 30 and 31). These measures (see Materials for details) were the LFPS-BF 2.0, one of the respective PID-5 subscales (Hostility, Irresponsibility, or Suspiciousness) and one of the scales for the respective flavors (Emotion Regulation, Self-control, or Delusional Paranoid

Thoughts). The scales were presented in random order. Participants received no additional incentives for participating in these studies.

Participants were excluded if they failed one of the a-priori defined plausibility checks (see https://osf.io/93tw6/?view_only=7f95882ef704429f866039350fc25baf). For details on the respective samples see Table 1. Based on a-priori power analyzes, we aimed at a minimum sample size of N=652 for all analyzes to detect even non-trivial increases in R^2=2% with a power of 95%[6].

Materials

In all three studies, D was measured with the full version (D70; Moshagen et al., 2020), thus with 70 items on a five-point Likert-type scale. Correspondingly, in all analyzes, D was modeled as the g-factor over all 70 items (in a bi-factor model as recommended by Bader et al., 2021). All other variables were again modeled as single factors. The three aversive AMPD traits were again measured with the respective subscale of the full PID-5 and IPD with the subscale of the LFPS-BF 2.0, all again on a four-point Likert-type scale.

Table 1

Demographics in Studies 3a-c

Sample	*N*	Sex	Age	Native Speakers	German residents
Study 3a	710	51.4% female, 46.6% male, 1.8% diverse	Range: 18-77, *M*=36.1, *SD*=14.3	97%	90%
Study 3b	639	50.5% female, 47.2% male, 2.2% diverse	Range: 18-74, *M*=33.7, SD=13.2	100%	94%
Study 3c	888	49.5% female, 47.5% male, 2.5% diverse	Range: 18-73, *M*=31.4, *SD*=11.7	96%	88%

Note. Other than the demographic data listed, no further demographic data was collected.

[6] Although Study 3b had a slightly lower sample size than N=652, a post-hoc power analyzes revealed that power was still highly satisfactory (94.6%). Also, we stated in the preregistration that the threshold for triviality as per Cohen (2013, p. 413) of f^2=.02 translates into R^2=1.8%, whereas it actually is R^2=2%. Crucially, both thresholds lead to the same conclusions regarding all hypotheses and power analyzes were based on the f^2-value.

For Study 3a (a lack of) strategies for Emotion Regulation was measured with the respective subscale of the *Difficulties in Emotion Regulation Scale Short Form* (DERS-SF; Kaufman et al., 2016) in German (Gutzweiler & In-Albon, 2018), thus with three items on a five-point Likert-type scale. The scale measures strategies for Emotion Regulation reliably (Cronbach's α = .89 - .92) and shows almost near perfect correlation with the full version (*r* = .91 - .96; Kaufman et al., 2016). It has further been shown to correlate with different outcomes of interpersonal psychopathology, such as aggression, rule breaking, and other (antagonistic) externalizing behavior (Kaufman et al., 2016).

For Study 3b (a lack of) Self-control was measured with the *Brief Self-control Scale* (BSCS; Tangney et al., 2004) in German (Sproesser et al., 2011), thus with 13 items on a five-point Likert-type scale. We chose the scale due to its good psychometric properties, i.e., high internal consistency (Cronbach's α = .83 - .85), strong correlation with its full version (*r* = .92 - .93), and as it has been found to be associated with different outcomes interpersonal psychopathology, e.g., antisocial personality patterns (Freeman et al., 2005).

For Study 3c Delusional Paranoid Thoughts were measured with the conviction version of the *Paranoia Checklist* (Freeman et al., 2005) in German (Lincoln, 2017), thus with 18 items on a five-point Likert-type scale. The scale has indicated excellent psychometric qualities (Cronbach's α < .90; Freeman et al., 2005) and was specifically designed to access conviction in *(Delusional) Paranoid Thoughts* in non-clinical samples. We originally preregistered that we would use the full scale for the analyzes. However, results of Moritz et al. (2012) indicate that the scale actually measures two distinct factors, only one of which (indicated by five items) specifically represents delusional thoughts. As such, we base the main analyzes on the five-item subscale for *Delusional* Paranoid Thoughts but report the results with the full scale in Supplementary Table 6 on the OSF. Importantly, both analyzes lead to same conclusions regarding all preregistered hypotheses.

Analyses

All analyzes used SEM and followed the same (preregistered) pattern for each aversive trait and thus sub-Study: First (H1), it was tested whether each proposed unique flavor (Emotion

Regulation, Self-control, or Delusional Paranoid Thoughts) was (positively) correlated with its respective aversive AMPD trait (Hostility, Irresponsibility, and Suspiciousness). Second, it was tested whether each flavor was also (positively) correlated with IPD (H2). Both tests can be considered preconditions for the main hypotheses of interest. Third, it was tested whether each proposed flavor accounted for a non-trivial (i.e., ΔR^2>2%) incremental proportion of variance in IPD over and above D (H3). Specifically, we compared the explained variance (R^2) in IPD from a regression model with only D as the predictor (model 1) vs. a model with D and the specific flavor as joint predictors (model 2).

Fourth, it was tested whether the respective aversive AMPD trait explains a non-trivial proportion (i.e., ΔR^2>2%) of incremental variance in IPD over and above D (H4). To this end, we compared the explained variance (R^2) in IPD from a latent regression model with only D as the predictor (model 1) vs. a model with D and the respective AMPD trait as the predictors (model 3). This was essentially to replicate the effects found in Study 1 and 2.

Finally, and most importantly for the current research question, we investigated whether the incremental variance in IPD explained by each aversive AMPD trait beyond D (as predicted in H4) can be attributed to the respective proposed flavor (H5). If this were the case, then the incremental variance of the aversive AMPD trait over and above D should be considerably reduced once this flavor is (also) accounted for. To this end, we ran an additional regression (model 4) with D, the respective AMPD trait, and the proposed flavor as predictors of IPD and compared the resulting variance explained to that of model 2 (D & flavor as predictors). We then calculated whether there was a relevant reduction (>2%) in the incremental variance explained by the respective AMPD trait, i.e., ΔR^2 (model2 vs. model4) - ΔR^2 (model1 vs. model3). As supplementary analyzes, we ran the same models with the self-subscale and the overall score of the LFPS-BF 2.0 as outcomes.

Results

For descriptive statistics, observed correlations, internal consistency estimates of all scales, as well as, exact model fits statistics (median RMSEA=.086, SRMR=.046, ω=.83) see Supplementary Table 7 and 8 on the OSF. The explained variances in IPD of all models are depicted in Table 2.

Table 2

Explained variances (R^2) in IPD by each model across Studies 3a-c

#Model	Predictors of IPD	R^2 (Study 3a)	R^2 (Study 3b)	R^2 (Study 3c)
Model 1	D	36%	27%	29%
Model 2	D + flavor	53%	48%	44%
Model 3	D + AMPD trait	51%	50%	48%
Model 4	D + flavor + AMPD trait	56%	50%	48%

All hypotheses across all three studies were supported. Emotion Regulation, Self-control, and Delusional Paranoid Thoughts were all related (all p<.001) with their respective AMPD trait, r=.34, .69, and .68, respectively, supporting H1. Confirming H2, they were also associated with IPD, r=.48, .59, and .51, respectively. Regarding the incremental prediction of IPD of each flavor over and above D (H3), all flavors explained substantial incremental variance, ΔR^2=17%, 21%, and 15%, respectively (all p<.001), confirming that all three yield aspects relevant for IPD beyond D.

Moreover, all three AMPD traits – Hostility, Irresponsibility, and Suspiciousness – also explained notable (and significant, all p<.001) incremental variance in IPD over and above D (H4), replicating the findings from Study 1 & 2 and once more suggesting that they involve flavors beyond D relevant for IPD. Finally, and most importantly, we found clear evidence for H5: In each case, the incremental variance explained in IPD of the aversive AMPD trait over and above D was strongly reduced once the respective hypothesized flavor was accounted for.

Specifically, Hostility, Irresponsibility, and Suspiciousness explained 15%, 23%, and 19% variance in IPD beyond D respectively, but this was reduced by 12%, 21%, and 15%, once the respective flavor, i.e., Emotion Regulation, Self-control, or Delusional Paranoid Thoughts was accounted for. In other words, 95% of the total explained variance in IPD by D, Emotion Regulation and Hostility, is due to D and Emotion Regulation alone, 96% of the total explained variance in IPD by D, Self-control and Irresponsibility, is due to D and Self-control alone, and 92% of the total explained

variance in IPD by D, Delusional Paranoid Thoughts and Suspiciousness, is due to D and Delusional Paranoid Thoughts alone.

Supplementary analyzes with the self-subscale and overall score of the LFPS-BF 2.0 lead to the same conclusion regarding all hypotheses (see Supplementary Table 9 on the OSF for details). This indicates that the relation of the three aversive AMPD traits – Hostility, Irresponsibility, and Suspiciousness – with personality dysfunctioning in general is due to D and their respective flavor.

General Discussion

Summary

Research on socially and/or ethically aversive personality comes in two largely distinct traditions: clinical personality psychopathology and broad personality research. In the Alternative Model for Personality Disorders (AMPD) it is assumed that the *severity* of a personality disorder is indicated by an impairment in personal and/or interpersonal functioning (Criterion A), while the *style is* indicated by 25 maladaptive trait facets (Criterion B). Especially for *interpersonal personality dysfunctioning* (IPD), socially and/or ethically aversive traits are of prime relevance for such impairments. Largely independently, highly similar traits have been established in broad personality research, often denoted as “dark” traits, conceptualized as (more) subclinical. This separation of research traditions, however, is especially questionable given the well-established strong empirical overlap of aversive traits across research traditions (e.g., Rose et al., 2023; Scholz et al., 2022; Sleep et al., 2017; Vize et al., 2020). This begs the question whether the traits from these different traditions actually do relate differently to IPD – as their conceptualization as clinical vs. subclinical may suggest.

More importantly, so far, it has remained an open question to what extent any aversive trait’s association with IPD is due to what all aversive traits share as opposed to what is unique to this trait. As specified in the framework of the D-Factor of Personality (D; Moshagen et al., 2018), any and every aversive trait is a *flavored* manifestation of a single common disposition, D, defined as utility maximization at others’ expense accompanied by justifying beliefs. “Flavored” in this context

means that the respective aversive trait may emphasize certain aspects of D and/or may additionally subsume non-aversive aspects (beyond D). Viewed through this framework, one may test whether and to what extent the association between aversive traits (independent of their research traditions) and IPD is due to D or if they comprise additional, non-aversive flavors beyond their common aversive core relevant for IPD. To this end, we conducted five studies (four of them preregistered).

First, there were only negligible differences in how strongly aversive subclinical as opposed to clinical traits were associated with IPD. Thus, in substantive terms, their separation is highly doubtful. Given that we also found they had near identical test information curves, there is also no psychometric evidence to support their separation as clinical vs. subclinical traits. These findings are in line with previous research indicating that aversive subclinical traits and their common core are related to different forms of psychopathology (Blasco-Belled et al., 2023; Gómez-Leal et al., 2019; Hilbig et al., 2020; Rose et al., 2023). In other words, this is what one ought to expect from a dimensional view of personality disorders (Hopwood et al., 2018): More extreme aversive tendencies, independent of their research traditions, are associated with impairment of a functioning personality. Clearly, theoretical integration of these artificially isolated research traditions is overdue (Miller & Campbell, 2008; Thomaes et al., 2017).

One theoretical position indeed integrating all aversive traits (independent of their tradition) is the D-Factor framework. Its main idea that all aversive traits are manifestations of D and thus traits' association with IPD is largely due to D was supported in Study 1 and 2: Most aversive traits are highly redundant in accounting for IPD. More specifically, 17 out of 20 aversive traits added little to nothing beyond their shared core (D) in accounting for IPD. Only three aversive traits explained a relevant proportion of incremental variance in IPD beyond D, namely Hostility, Irresponsibility, and Suspiciousness. Thus, for these (and only these) traits, one may ask which additional aspects associated with IPD they entail beyond D. In line with theoretical considerations (and confirming pre-registered hypotheses), we identified these as Emotion Regulation (flavor of Hostility), Self-control (flavor of Irresponsibly), and Delusional Paranoid Thoughts (flavor of Suspiciousness).

In summary, the association of 20 aversive traits with IPD can be parsimoniously attributed to four stable tendencies: D (i.e., utility maximization at others' expense accompanied by justifying beliefs), lack of Emotion Regulation (flavor of Hostility), lack of Self-control (flavor of Irresponsibly), and Delusional Paranoid Thoughts (flavor of Suspiciousness). This conclusion, in turn, yields implications for recent discussions in the field.

Implications

For one, some have argued that the AMPD may not cover the full spectrum of socially and/or ethically aversive traits and thus may need to be expanded (Clark & Watson, 2022; Hilbig et al., 2020), e.g., by Greed or Sadism. However, the current study implies that – in accounting for IPD – neither Greed nor Sadism offer incremental value over what they share with all other aversive traits and what is thus already well-covered by the existing AMPD traits (in combination). From a parsimonious view, the inclusion of new traits into the AMPD should be based on their incremental value (arguably especially in Criterion A) over and above already included traits.

The current results are also in line with the four functional systems identified by Wright et al. (2022), based on Contemporary Integrative Interpersonal Theory (CIIT; Pincus, 2005). These four systems describe where dysfunction in an interpersonal context may occur, i.e., self-dysregulation (inability to achieve one's goals), affect dysregulation (inability to regulate emotional experiences), behavioral dysregulation (unable to behave in mutually satisfying and socially appropriate ways), and perceptual distortion (unable to accurately perceive oneself or the other). The four systems described by Wright et al. (2022) bare striking resemblance with the non-redundant features of aversive traits found in the current studies: Self-control aligns with self-dysregulation, Emotion Regulation with affect dysregulation, Delusional Paranoid Thoughts with perpetual distortion, and D, by definition, describes behaviors which only serve the benefit of the individual at others' expense (behavioral dysregulation).

Limitations

Nonetheless, some limitations of the present research deserve attention. First, although we relied on accepted and validated operationalization of Criterion A and B, as common in the field, (Clark & Watson, 2022; Zimmermann et al., 2019), it is possible that (some) findings are limited to the measures chosen. Thus, replication of the present findings with other measures of clinical traits, such as the Computerized Adaptive Test of Personality Disorder (CAT-PD; Simms et al., 2011), would add further confidence. Due to the large overlap of many clinical trait measures (Crego & Widiger, 2016), such a replication ought to be successful. Second, the exclusive reliance on self-reports may have led to an overestimation of what is common across traits – due to common method variance. Although alternative modes of measurement (e.g., observer reports) would thus seem likely to produce smaller effects in absolute terms, we would expect the results of interest – the relative proportions of explained variance due to shared vs. unique aspects – to remain intact.

Conclusion

In conclusion, the current studies avail the integration of aversive clinical and broad personality research, providing further evidence against their artificial separation into “clinical” versus “subclinical”. Moreover, results indicate that most aversive traits are associated with IPD (only) due to their common aversive core, whereas the incremental variance explained by the remaining three can be attributed to (lack of) Emotion Regulation, (lack of) Self-control, and Delusional Paranoid Thoughts, respectively.

References

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

Bader, M., Hilbig, B. E., Zettler, I., & Moshagen, M. (2023). Rethinking Aversive Personality: Decomposing the Dark Triad Traits into their Common Core and Unique Flavors. *Journal of Personality.* Advance online publication. https://doi.org/10.1111/jopy.12785

Blasco-Belled, A., Tejada-Gallardo, C., Alsinet, C., & Rogoza, R. (2023). The links of subjective and psychological well-being with the Dark Triad traits: A meta-analysis. *Journal of Personality.* Advance online publication. https://doi.org/10.1111/jopy.12853

Chalmers, P., Pritikin, J., Robitzsch, A., Zoltak, M., KwonHyun, K., Falk, C. F., & & Meade, A. (2023). *Package 'mirt'* [Computer software]. https://cran.r-project.org/web/packages/mirt/mirt.pdf

Clark, L. A., & Watson, D. (2022). The trait model of the DSM-5 alternative model of personality disorder (AMPD): A structural review. *Personality Disorders: Theory, Research, and Treatment*, *13*(4), 328–336. https://doi.org/10.1037/per0000568

Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences*. Elsevier Science.

Crego, C., & Widiger, T. A. (2016). Convergent and discriminant validity of alternative measures of maladaptive personality traits. *Psychological Assessment*, *28*(12), 1561–1575. https://doi.org/10.1037/pas0000282

Freeman, D., Garety, P. A., Bebbington, P. E., Smith, B., Rollinson, R., Fowler, D., Kuipers, E., Ray, K., & Dunn, G. (2005). Psychological investigation of the structure of paranoia in a non-clinical population. *The British Journal of Psychiatry : The Journal of Mental Science*, *186*, 427–435. https://doi.org/10.1192/bjp.186.5.427

Furnham, A., Richards, S. C., & Paulhus, D. L. (2013). The Dark Triad of Personality: A 10 Year Review. *Social and Personality Psychology Compass*, *7*(3), 199–216. https://doi.org/10.1111/spc3.12018

Gómez-Leal, R., Megías-Robles, A., Gutiérrez-Cobo, M. J., Cabello, R., Fernández-Abascal, E. G., & Fernández-Berrocal, P. (2019). Relationship between the Dark Triad and depressive symptoms. *PeerJ*, *7*, e8120. https://doi.org/10.7717/peerj.8120

Gutzweiler, R., & In-Albon, T. (2018). Überprüfung der Gütekriterien der deutschen Version der Difficulties in Emotion Regulation Scale in einer klinischen und einer Schülerstichprobe Jugendlicher. *Zeitschrift Für Klinische Psychologie Und Psychotherapie*, *47*(4), 274–286. https://doi.org/10.1026/1616-3443/a000506

Hilbig, B. E., Thielmann, I., Klein, S. A., Moshagen, M., & Zettler, I. (2020). The dark core of personality and socially aversive psychopathology. *Journal of Personality.* Advance online publication. https://doi.org/10.1111/jopy.12577

Hopwood, C. J., Kotov, R., Krueger, R. F., Watson, D., Widiger, T. A., Althoff, R. R., Ansell, E. B., Bach, B., Michael Bagby, R., Blais, M. A., Bornovalova, M. A., Chmielewski, M., Cicero, D. C., Conway, C., Clercq, B. de, Fruyt, F. de, Docherty, A. R., Eaton, N. R., Edens, J. F., . . . Zimmermann, J. (2018). The time has come for dimensional personality disorder diagnosis. *Personality and Mental Health*, 82–86. https://doi.org/10.1002/pmh.1408.

Jones, D. N., & Paulhus, D. L. (2023). Short Dark Triad and Short Dark Tetrad. In P. K. Jonason (Ed.), *Shining light on the dark side of personality: Measurement properties and theoretical advances* (pp. 215–230). Hogrefe.

Kaufman, E. A., Xia, M., Fosco, G., Yaptangco, M., Skidmore, C. R., & Crowell, S. E. (2016). The Difficulties in Emotion Regulation Scale Short Form (DERS-SF): Validation and Replication in Adolescent and Adult Samples. *Journal of Psychopathology and Behavioral Assessment*, *38*(3), 443–455. https://doi.org/10.1007/s10862-015-9529-3

Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2012). Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychological Medicine*, *42*(9), 1879–1890. https://doi.org/10.1017/S0033291711002674

Lincoln, T. M. (2017). *PCL - Paranoia Checklist - deutsche Fassung.* https://doi.org/10.23668/psycharchives.4650

Maples, J. L., Carter, N. T., Few, L. R., Crego, C., Gore, W. L., Samuel, D. B., Williamson, R. L., Lynam, D. R., Widiger, T. A., Markon, K. E., Krueger, R. F., & Miller, J. D. (2015). Testing whether the DSM-5 personality disorder trait model can be measured with a reduced set of items: An item response theory investigation of the Personality Inventory for DSM-5. *Psychological Assessment*, *27*(4), 1195–1210. https://doi.org/10.1037/pas0000120

Marcus, D. K., & Zeigler-Hill, V. (2015). A Big Tent of Dark Personality Traits. *Social and Personality Psychology Compass*, *9*(8), 434–446. https://doi.org/10.1111/spc3.12185

Miller, J. D., & Campbell, W. K. (2008). Comparing clinical and social-personality conceptualizations of narcissism. *Journal of Personality*, *76*(3), 449–476. https://doi.org/10.1111/j.1467-6494.2008.00492.x

Miller, J. D., Gentile, B., Wilson, L., & Campbell, W. K. (2013). Grandiose and vulnerable narcissism and the DSM-5 pathological personality trait model. *Journal of Personality Assessment*, *95*(3), 284–290. https://doi.org/10.1080/00223891.2012.685907

Morey, L. C., Good, E. W., & Hopwood, C. J. (2020). Global personality dysfunction and the relationship of pathological and normal trait domains in the DSM-5 alternative model for personality disorders. *Journal of Personality.* Advance online publication. https://doi.org/10.1111/jopy.12560

Morey, L. C., McCredie, M. N., Bender, D. S., & Skodol, A. E. (2022). Criterion A: Level of personality functioning in the alternative DSM-5 model for personality disorders. *Personality Disorders: Theory, Research, and Treatment*, *13*(4), 305–315. https://doi.org/10.1037/per0000551

Moritz, S., van Quaquebeke, N., & Lincoln, T. M. (2012). Jumping to conclusions is associated with paranoia but not general suspiciousness: A comparison of two versions of the probabilistic reasoning paradigm. *Schizophrenia Research and Treatment*, *2012*, 384039. https://doi.org/10.1155/2012/384039

Moshagen, M., Hilbig, B. E., & Zettler, I. (2018). The dark core of personality. *Psychological Review*, *125*(5), 656–688. https://doi.org/10.1037/rev0000111

Moshagen, M., Zettler, I., & Hilbig, B. E. (2020). Measuring the Dark Core of Personality. *Psychological Assessment*, *32*(2), 182–196. https://doi.org/10.1037/pas0000778

Paulhus, D. L., & Williams, K. M. (2002). The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, *36*(6), 556–563. https://doi.org/10.1016/S0092-6566(02)00505-6

Pincus, A. (2005). A Contemporary Integrative Interpersonal Theory of Personality Disorders. In M. F. Lenzenweger & J. F. Clarkin (Eds.), *Major theories of personality disorder* (2nd ed., pp. 282–331). The Guilford Press.

R Core Team. (2024). *R: A language and environment for statistical computing [*Computer software]. R Foundation for Statistical Computing. https://www.r-project.org/

Ringwald, W. R., Hallquist, M. N., Dombrovski, A. Y., & Wright, A. G. C. (in press). Associations between personality (dys) function and general behavioral dysregulation. https://scholar.google.com/citations?user=y7xoq9yaaaaj&hl=de&oi=sra

Rose, L., Lynam, D. R., & Miller, J. D. (in press). Measuring the “Dark” Triad: Comparing the Five-Factor Model Antagonistic Triad Measure to other commonly used self-report instruments. *Assessment*.

Rose, L., Miller, J. D., & Lynam, D. R. (2023). Validation of the Five-Factor Model Antagonistic Triad Measure. *Assessment*, *30*(3), 782–797. https://doi.org/10.1177/10731911211068083

Rosseel, Y., Jorgensen, T. D., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, E., Hallquist, M., Rhemtulla, M., Katsikatsou, M., Barendse, M., Scharf, F., & Du, H. (2023). *Package ‘lavaan’* [Computer software]. https://cran.r-project.org/web/packages/lavaan/lavaan.pdf

Scholz, D. D., Hilbig, B. E., Thielmann, I., Moshagen, M., & Zettler, I. (2022). Beyond (low) Agreeableness: Towards a more comprehensive understanding of antagonistic

psychopathology. *Journal of Personality.* Advance online publication. https://doi.org/10.1111/jopy.12708

Scholz, D. D., Zimmermann, J., Moshagen, M., Zettler, I., & Hilbig, B. E. (in press). Theoretical and empirical integration of 'dark' traits and socially aversive personality psychopathology. *Journal of Personality Disorders*.

Sharp, C. (2022). Fulfilling the promise of the LPF: Comment on Morey et al. (2022). *Personality Disorders: Theory, Research, and Treatment*, *13*(4), 316–320. https://doi.org/10.1037/per0000567

Sharp, C., & Miller, J. D. (2022). Ten-year retrospective on the DSM-5 alternative model of personality disorder: Seeing the forest for the trees. *Personality Disorders: Theory, Research, and Treatment*, *13*(4), 301–304. https://doi.org/10.1037/per0000595

Sharp, C., & Wall, K. (2021). Dsm-5 Level of Personality Functioning: Refocusing Personality Disorder on What It Means to Be Human. *Annual Review of Clinical Psychology*, *17*, 313–337. https://doi.org/10.1146/annurev-clinpsy-081219-105402

Simms, L. J., Goldberg, L. R., Roberts, J. E., Watson, D., Welte, J., & Rotterman, J. H. (2011). Computerized adaptive assessment of personality disorder: Introducing the CAT-PD project. *Journal of Personality Assessment*, *93*(4), 380–389. https://doi.org/10.1080/00223891.2011.577475

Sleep, C. E., Hyatt, C. S., Lamkin, J., Maples-Keller, J. L., & Miller, J. D. (2018). Examining the relations among the DSM-5 alternative model of personality, the five-factor model, and externalizing and internalizing behavior. *Personality Disorders: Theory, Research, and Treatment*, *9*(4), 379–384. https://doi.org/10.1037/per0000240

Sleep, C. E., & Lynam, D. R. (2022). The problems with Criterion A: A comment on Morey et al. (2022). *Personality Disorders: Theory, Research, and Treatment*, *13*(4), 325–327. https://doi.org/10.1037/per0000585

Sleep, C. E., Lynam, D. R., Hyatt, C. S., & Miller, J. D. (2017). Perils of partialing redux: The case of the Dark Triad. *Journal of Abnormal Psychology*, *126*(7), 939–950. https://doi.org/10.1037/abn0000278

Spitzer, C., Müller, S., Kerber, A., Hutsebaut, J., Brähler, E., & Zimmermann, J. (2021). Die deutsche Version der Level of Personality Functioning Scale-Brief Form 2.0 (LPFS-BF): Faktorenstruktur, konvergente Validität und Normwerte in der Allgemeinbevölkerung [The German Version of the Level of Personality Functioning Scale-Brief Form 2.0 (LPFS-BF): Latent Structure, Convergent Validity and Norm Values in the General Population]. *Psychotherapie, Psychosomatik, medizinische Psychologie*, *71*(7), 284–293. https://doi.org/10.1055/a-1343-2396

Sproesser, G., Strohbach, S., Schupp, H., & Renner, B. (2011). Candy or apple? How self-control resources and motives impact dietary healthiness in women. *Appetite*, *56*(3), 784–787. https://doi.org/10.1016/j.appet.2011.01.028

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, *72*(2), 271–324. https://doi.org/10.1111/j.0022-3506.2004.00263.x

Thomaes, S., Brummelman, E., Miller, J. D., & Lilienfeld, S. O. (2017). The dark personality and psychopathology: Toward a brighter future. *Journal of Abnormal Psychology*, *126*(7), 835–842. https://doi.org/10.1037/abn0000305

Vize, C. E., Collison, K. L., Miller, J. D., & Lynam, D. R. (2020). The "core" of the dark triad: A test of competing hypotheses. *Personality Disorders: Theory, Research, and Treatment*, *11*(2), 91–99. https://doi.org/10.1037/per0000386

Weekers, L. C., Hutsebaut, J., & Kamphuis, J. H. (2019). The Level of Personality Functioning Scale-Brief Form 2.0: Update of a brief instrument for assessing level of personality functioning. *Personality and Mental Health*, 3–14. https://doi.org/10.1002/pmh.1434

World Health Organization. (2023). *International statistical classification of diseases and related health problems*. (11th ed.). https://icd.who.int/

Wright, A. G. C., & Ringwald, W. R. (2022). Personality disorders are dead; long live the interpersonal disorders: Comment on Widiger and Hines (2022). *Personality Disorders: Theory, Research, and Treatment*, *13*(4), 364–368. https://doi.org/10.1037/per0000552

Wright, A. G. C., Ringwald, W. R., Hopwood, C. J., & Pincus, A. L. (2022). *It's Time to Replace the Personality Disorders with the Interpersonal Disorders*. Center for Open Science. https://doi.org/10.31234/osf.io/7syvf

Zettler, I., Moshagen, M., & Hilbig, B. E. (2021). Stability and Change: The Dark Factor of Personality Shapes Dark Traits. *Social Psychological and Personality Science*, *12*(6), 974–983. https://doi.org/10.1177/1948550620953288

Zimmermann, J., Altenstein, D., Krieger, T., Holtforth, M. G., Pretsch, J., Alexopoulos, J., Spitzer, C., Benecke, C., Krueger, R. F., Markon, K. E., & Leising, D. (2014). The structure and correlates of self-reported DSM-5 maladaptive personality traits: Findings from two German-speaking samples. *Journal of Personality Disorders*, 518–540. https://doi.org/10.1521/pedi_2014_28_130

Zimmermann, J., Kerber, A., Rek, K., Hopwood, C. J., & Krueger, R. F. (2019). A Brief but Comprehensive Review of Research on the Alternative DSM-5 Model for Personality Disorders. *Current Psychiatry Reports*, *21*(9), 21–92. https://doi.org/10.1007/s11920-019-1079-z

Zimmermann, J., Müller, S., Bach, B., Hutsebaut, J., Hummelen, B., & Fischer, F. (2020). A Common Metric for Self-Reported Severity of Personality Disorder. *Psychopathology*, *53*(3-4), 168–178. https://doi.org/10.1159/000507377