Measuring the Dark Core of Personality

Morten Moshagen

Ulm University, Germany

Ingo Zettler

University of Copenhagen, Denmark

Benjamin E. Hilbig

University of Koblenz-Landau, Germany

Author Note

Correspondence concerning this manuscript should be directed to Morten Moshagen, Research Methods, Institute of Psychology and Education, Ulm University, Albert-Einstein-Allee 4, 89081 Ulm, Germany. E-Mail: morten.moshagen@uni-ulm.de.

The data have not been published previously, neither in whole nor in part. Preparation of this manuscript was supported by a grant from the Carlsberg Foundation (CF16-0444) to the second author, and Grants HI 1600/1-2 and HI 1600/6-1 to the final author, who was also supported by the research-training group Statistical Modeling in Psychology (GRK 2277), all funded by the German Research Foundation (DFG).

Abstract

The Dark Factor of Personality (D) is the basic disposition that gives rise to specific personality traits related to antagonistic, malevolent, or socially aversive behavior, thereby representing the common core of dark personality. Whereas existing evidence clearly supports the conceptualization and utility of D, the assessment of D was possible only indirectly and with extensive effort, so far. Applying rational item selection techniques to seven large and highly heterogeneous samples (total $N > 165{,}000$), we herein identified sets of items (comprising 70, 35, and 16 items, respectively) that allow for a psychometrically sound and more concise assessment of D. Results indicate that all identified item sets are characterized by high internal consistencies and high retest-reliabilities, clearly map on a single factor in line with the definition of D, and exhibit substantial associations to various relevant criteria, including three assessments of actual behavior. In particular, the item sets showed substantial associations with behavioral measures of individual utility maximization disregarding, accepting, or malevolently provoking disutility on others, and were also related to various justifying beliefs, thereby mirroring the defining features of D. In sum, the identified item sets allow for a concise, reliable, and valid assessment of D.

Keywords: Dark Factor of Personality; D factor; dark traits; dark core

Public Significance Statement

The Dark Factor of Personality (D) is the basic personality disposition from which specific dark traits related to malevolent behavior (such as Psychopathy) arise as manifestations. Improving on previous indirect assessment attempts, we herein identify and validate sets of items differing in length that are optimally suited to measure D.

Measuring the Dark Core of Personality

Researchers, practitioners, and laypeople alike have been interested in stable dispositions related to antagonistic, malevolent, or socially aversive behavior. These traits are often subsumed under the umbrella term “dark traits”, with Machiavellianism, Narcissism, and Psychopathy—the components of the Dark Triad (Paulhus & Williams, 2002)—as the most prominent examples. There is ample evidence indicating that elevated levels in such traits bear important consequences for everyday functioning in a variety of domains. Specifically, dark traits have been associated with mental health and psychopathology (e.g., Blonigen, Hicks, Krueger, Patrick, & Iacono, 2005; Harrop et al., 2017; Monaghan, Bizumic, & Sellbom, 2016) as well as with aggression, delinquency, and many other consequential behaviors (Muris, Merckelbach, Otgaar, & Meijer, 2017; O’Boyle, Forsyth, Banks, & McDaniel, 2012). Given their importance, a plethora of dark traits with often rather subtle theoretical differences has been introduced in the past decades (Marcus & Zeigler-Hill, 2015; Paulhus & Jones, 2014). Indeed, operationally (in terms of item content) and empirically (in terms of associations) dark traits show a substantial overlap (Muris et al., 2017; O’Boyle, Forsyth, Banks, Story, & White, 2015), giving rise to the question how to conceptually describe and explain their commonalities.

In this respect, Moshagen, Hilbig, and Zettler (2018) provided a theoretical framework unifying and extending other notions about the commonalities of dark traits (e.g., Diebels, Leary, & Chon, 2018; Jonason, Li, Webster, & Schmitt, 2009; Jones & Figueredo, 2013). Moshagen et al. (2018) proposed that dark traits arise as flavored manifestations of a general underlying dispositional tendency, which thereby represents the common core of all dark traits. This underlying tendency, the Dark Factor of Personality (D), is 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” (p. 657).

Utility maximization may refer to monetary- or status-related goals, but likewise to emotional gains such as feelings of superiority or pleasure. Similarly, disutility of others refers to any type of cost an individual may bear, including material, emotional, or physical disutilities. Finally, justifying beliefs comprise any implicit or explicit belief an individual might rely on to justify malevolent behaviors. Examples of such beliefs include a sense of entitlement for oneself or one's group, viewing the world as a competitive jungle, or believing that others are stupid and, in turn, deserve to be exploited. Importantly, the concept of D does not imply that individuals must hold any one particular belief or set of beliefs; instead, the main idea is that individuals hold some belief(s) that they deem appropriate to justify malevolent acts.

In analogy to the *g*-factor of intelligence, D is conceptualized as the underlying disposition responsible for the emergence of any particular dark trait, so that any dark trait can be regarded as a specific, flavored manifestation of D. Correspondingly, elevated levels in D may become evident in one or more of the Dark Triad components or in any other dark trait(s), such as Sadism or Spitefulness. However, any specific manifestation of D (i.e., any specific dark trait) may be uniquely flavored by including aspects that are not shared by other dark traits and are thus rather unique to this particular trait (such as agentic extraversion in Narcissism, e.g., Crowe, Lynam, Campbell, & Miller, in press; or disinhibition in Psychopathy, e.g., Patrick, Fowles, & Krueger, 2009) and/or by placing a different emphasis on the defining features of D. For example, high-D individuals might hardly notice that their behavior inflicts disutility on others (arguably pronounced in Psychopathy), or they might notice without caring (arguably pronounced in Machiavellianism), or they might actually derive own utility from the very act of inflicting disutility on others (arguably pronounced in Sadism). Nonetheless, the common characteristic of dark traits is that disutility *is* inflicted on others in pursuing one's own goals – as is the very definition of D.

The concept of D received considerable empirical support in a series of studies considering 9 different dark traits (Egoism, Machiavellianism, Narcissism, Moral Disengagement, Psychological Entitlement, Psychopathy, Sadism, Self-Interest, and Spitefulness; Moshagen et al., 2018; Zettler, Moshagen, & Hilbig, 2019). Briefly, bifactor modeling supported the existence of a single general factor in line with the definition of D. This factor exhibited a high degree of rank-order stability over four years, predicted behavioral outcomes, and was substantially related to a host of criterion measures, including aggression and dominance as well as lack of empathy and nurturance. In contrast, specific dark traits rarely explained incremental variance in the criteria, in particular with respect to the behavioral outcomes. Furthermore, D was shown to determine how dark traits develop over time, i.e., D longitudinally predicted individual differences in dark traits measured four years later (often to a similar extent as the stability of the to-be-predicted dark traits themselves), and individual change in dark traits could be traced back to a change in D. In tandem, evidence suggests that D captures much of the behaviorally relevant variance of most of the dark traits considered and indeed shapes their development, indicating that D can be considered the very basic disposition from which dark traits arise; that is, the core of dark personality.

Measuring D

D is explicitly conceptualized as a fluid construct, implying that D conforms to the principle of "indifference of the indicator" (Spearman, 1927). This means that D does not crucially depend on a specific choice of indicator variables, as long as a sufficient number of indicator variables related to dark personality are included. Empirical support for this notion has been provided by means of several resampling studies (Moshagen et al., 2018; Zettler et al., 2019). For instance, measuring D by random subsets of the available items only marginally affected the meaning of the resulting factors, as indicated by highly similar

coefficients in the longitudinal prediction of dark traits and by very high correlations between such “reduced” versions of D and D measured via all available items.

Given the conceptualization of D as a fluid construct with dark traits as its specific manifestations, any particular measurement instrument designed to assess a dark trait also will reflect D. In other words, any dark trait inventory acts as a “vehicle” (Jensen, 1992) to measure D, so that, in principle, D could be assessed by employing any arbitrary dark trait measure. However, any specific operational definition of D obviously depends on the indicator variables used. For example, if it is attempted to measure D using an inventory designed to assess Psychopathy, the resulting scores will of course primarily reflect Psychopathy and only secondarily reflect D. Also, they may reflect aspects of Psychopathy that largely lie beyond the theoretical scope of D (such as dishinbition). Thus, although any specific dark trait inventory will also measure D to a certain extent, appropriately measuring D itself requires a sufficiently large number of diverse indicators in line with the theoretical scope of D.

So far, D has been assessed by items included in established scales measuring various specific dark traits, regardless of whether the particular vehicle or a particular item is actually directly suited to indicate D. For instance, Moshagen et al. (2018) measured D by 93 items of 9 different scales. However, about 25% of the considered items exhibited low loadings on D, indicating that these items only marginally contributed to the measurement of D and thus could be dropped with little loss of information. In addition, a shortcoming of most existing dark trait measures is the use of predominantly positively keyed items. Specifically, only 10 out of the 93 items used in Moshagen et al. (2018) were negatively keyed, in turn amplifying issues resulting from response biases such as aquiescence. Such response biases might distort the factor structure (DiStefano & Motl, 2006), bias longitudinal assessments (Wetzel, Lüdtke, Zettler, & Böhnke, 2016) and relationships to other measures (Podsakoff, MacKenzie, Lee, &

Podsakoff, 2003), or render comparisons across groups of individuals meaningless (e.g., strata defined by education or culture; Cheung & Rensvold, 2000). Obviously, any appropriate measure of D (or indeed the more specific dark traits) must attempt to avoid such biases.

Given the above, a theoretically adequate, concise, and psychometrically sound measurement of D requires a sufficient number of indicators representing a diverse set of dark traits to cover D in full breadth, using (sets of) items that are well suited to indicate D and meet contemporary psychometric standards. Correspondingly, the purpose of the present studies is to identify sets of items derived from existing dark trait inventories using rational item selection techniques to obtain a concise and psychometrically sound measurement of D. In doing so, we also aimed at identifying item sets differing in length to allow for the measurement of D when assessment times must be kept to minimum.

Methods

All studies were approved by the ethics committee of the University of Koblenz-Landau (#154_2018). Additional information on the samples, items, and analyses is provided in the online supplementary materials available at the Open Science Framework (https://osf.io/ag835/).

Participants and Procedure

The studies are based on seven different samples. Samples A1-A3 were used to inform item selection, whereas samples B1-B4 were used to determine the factor structure, psychometric properties, and validity of the finally retained items. All participants provided informed consent and were debriefed after completing the respective study.

Participants of samples A1, A2, A3, and B1 were attracted by international media coverage about D. We set up a website (darkfactor.org) providing general information about D and a questionnaire allowing individuals to determine their level in D. The participants completed the questionnaire on an anonymous and voluntary basis without any compensation

apart from feedback on their scores. In samples A1-A3, items were administered according to the Synthetic Aperture Personality Assessment procedure (Revelle, Wilt, & Rosenthal, 2010): Participants were free to choose to complete 30, 60, or 80 items drawn randomly (and presented in random order) from the itempool (which initially comprised 182 items, and was reduced in two consecutive steps to 162 and 118 items, respectively). Participants opting to complete 80 items received 10 additional questions not pertinent to the present study immediately after completing the items referring to D. Participants in validation sample B1 always completed the same set of 70 ultimately retained items.

The remaining validation samples (B2-B4) were recruited through professionally managed online panels (B2 using MTurk, B3 and B4 using Prolific). In these samples, participants received a flat fee for their participation and, in case of sample B3, an additional payoff (with a maximum of 4 GBP) depending on their choices during the study.

All studies were conducted online. In all samples, we excluded participants who showed suspicious response behaviour (such as failing attention check items, requiring less than 2 sec for each item on average, or always selecting the same response option), who indicated an insufficient proficiency in the English language, or who did not provide information on age and gender. We also excluded records associated with the same ip-address to remove data potentially resulting from multiple participations of the same individual.

The preliminary item evaluation and refinement was based on sample A1 (N = 7,226; 64% male, 77 other; mean age 34.75 years, SD = 13.1). Participants originated from more than 100 countries, though mostly residents of the US (39%) or Denmark (32%; all other ≤ 6%). Participants completed a subset of either 30 (31%) or 60 (69%) items drawn randomly from the initial itempool of 182 items, resulting in a mean covariance coverage of .08.

The first item selection was performed using sample A2 (N = 74,477; 59% male, 596 other; mean age 35.18 years, SD = 12.1). The participants were from over 150 countries, with

most participants living in the US (32%) or Belgium (22%; all other ≤ 5%). Participants completed a subset of either 30 (34%), 60 (26%), or 80 (40%) items drawn at random from the revised itempool (after the preliminary item evaluation) comprising 162 items, resulting in a mean covariance coverage of .14.

The final item selection was performed using sample A3 (N = 77,218; 54% male, 564 other; mean age 35.13 years, SD = 12.5). Again, participants originated from over 150 countries, with relatively large numbers stemming from the US (21%), Germany (20%), and Denmark (14%; all other ≤ 7%). Participants completed a subset of 30 (40%), 60 (11%), or 80 (49%) items drawn randomly from the reduced itempool (after the first item selection) of 118 items, which led to a mean covariance coverage of .28.

In the first validation sample (B1), participants always completed the 70 finally retained items (presented in random order). Of the 5,190 participants, 51% were male (50 other). Mean age was 29.28 years (SD = 10.7). Most participants originated from the US (20%) or Norway (19%; all other ≤ 6%).

The second validation sample (B2) was used to determine test-retest reliability. Participants completed the 70 finally retained items (presented in random order) twice with a lag of at least 34 days. Of the 130 participants at the first measurement occasion, 77 (60% male, 1 other; mean age = 34.71 years, SD = 10.0) completed the same items again, on average 34.61 (SD = 0.6) days later (drop-out rate of 41%). The majority of the participants originated from the US (72%) or India (17%; all other ≤ 2%).

The third validation sample (B3) comprised 537 individuals (50% female; mean age 29.93 years, SD = 9.9) who completed the pool of 118 items as obtained after the first item selection (presented in random order) and afterwards an incentivized behavioral measure of pro-social (vs. pro-self) value orientations and a similar behavioral task measuring sadism.

For the purpose of validation, we only analyze the responses to the 70 finally retained items. Participants primarily originated from the UK (26%) or Poland (12%; all other ≤ 8%).

Participants in the fourth validation sample (B4) completed the 70 finally retained items (again, in random order) at the first measurement occasion (N = 978; 50% male, 4 other; mean age 31.08 years, SD = 9.7). Ten days later, participants were reinvited to complete a battery of self-report questionnaires serving as criteria, which were also presented in random order. At the end of the second measurement occasion, participants engaged in a cheating task as a behavioral measure of dishonesty. Depending on their behavior in the cheating task, participants either completed a boring search task or were immediately directed to the end of the study. Thus, the incentive for cheating was to avoid tedious work (see Hilbig & Zettler, 2015, Study 3). Data collection at the second measurement occasion was stopped upon N = 498 respondents (52% male, 2 other; mean age 30.69 years, SD = 9.1) successfully completed the study. The sample was once more diverse in representing participants originating from the UK (24%), Portugal (13%), or Poland (10%; all other ≤ 8%).

Initial Itempool

To derive item sets suited for the measurement of D, we evaluated items from established measures of 12 different dark traits with respect to their ability to indicate D. We attempted to be highly inclusive by considering eight of the nine dark traits also included in Moshagen et al. (2018) and four additional dark traits (Amoralism-Crudelia, Amoralism-Frustralia, Greed, and Self-Centeredness). With regard to the additional dark traits, we—like Moshagen et al. (2018)—searched for articles referring to dark traits published between 2014 and early 2018 in journals related to Personality Psychology (namely, *Assessment*, *European Journal of Personality*, *Journal of Personality*, *Journal of Personality and Social Psychology*, *Journal of Personality Assessment*, *Journal of Research in Personality*, *Personality and Individual Differences*, *Personnel Psychology*, *Psychological Assessment*, *Psychological*

Science), considered relevant reviews (e.g., Marcus & Zeigler-Hill, 2015; Paulhus & Jones, 2014), and considered traits that were otherwise brought to our attention (e.g., via correspondences with other researchers).[1]

We obtained an initial pool of 184 items by considering various measures assessing the 12 dark traits (see Table 1 for the number of items per dark trait). In selecting the measures, we opted for established full-length versions (over short forms) whenever possible. We often considered multiple measures for the same dark trait to ensure that it is represented in adequate breadth. At the same time, we also attempted to ensure that all traits are represented by an approximately equal number of items, so that no single trait strongly dominates the item pool.[2] As detailed below, about one third of the original items were modified to achieve a balance between positively and negatively keyed items. Participants indicated their level of agreement to each item on a five-point Likert response scale ranging from 1 (strongly disagree) to 5 (strongly agree).

Amoralism-Crudelia and Amoralism-Frustralia. Crudelia and Frustralia are two components of the Amoralism concept by Knežević (2003). Crudelia represents Amoralism involving brutality, whereas Frustralia represents Amoralism caused by frustration. Crudelia and Frustralia were represented by the respective subscales (comprising 13 and 14 items, respectively) from the AMR40 inventory (Knežević, 2003). We modified 3 items of the scale

[1] Unlike Moshagen et al. (2018), we did not consider Self-Interest, which is defined as the motivation to pursue gains in "socially valued domains, including material goods, social status, recognition, academic or occupational achievement, and happiness" (Gerbasi & Prentice, 2013, p. 496). From this definition it becomes clear that Self-Interest describes a form of utility maximization that rarely implies disutility for others and thus largely lies outside the scope of D. We also did not consider Dispositional Envy (e.g., Rentzsch & Gross, 2015), because we consider envy as an emotional state that may act as antecedent or moderator of malevolent acts rather than being an instance of D (cf., Lange, Paulhus, & Crusius, 2018).

[2] The only exception pertained to Self-Centredness, because there is only a single measure comprising 4 items.

measuring Frustralia, so that 7 items of this scale were negatively keyed. From the scale measuring Crudelia, 6 items were negatively keyed.

Egoism. Egoism is „the excessive concern with one's own pleasure or advantage at the expense of community well-being" (Weigel, Hessing, & Elffers, 1999, p. 349). It was represented by the corresponding 20-item scale by Weigel et al. (1999). We modified 8 items, so that 10 items were negatively keyed.

Greed. Greed is the dissatisfaction of not having enough, combined with the insatiable desire to acquire more. It was represented by 6 and 5 items from the Dispositional Greed Scales by Seuntjens, Zeelenberg, van de Ven, and Breugelman (2015) and Krekels and Pandelaere (2015), respectively. Two items of the original scales were excluded because of duplicate item content. We additionally modified 4 items, so that 6 of the 11 items were negatively keyed.

Machiavellianism. Machiavellianism can be defined as a "duplicitous interpersonal style assumed to emerge from a broader network of cynical beliefs and pragmatic morality" (Jones & Paulhus, 2014, p. 93). It was represented by the respective subscale of the Short Dark Triad (Jones & Paulhus, 2014), 6 items from the Mach IV (Christie & Geis, 1970) related to a cynical worldview and manipulative tactics, and 5 items from the Amorality and Desire for Control subscales of the Machiavellian Personality Scale (Dahling, Whitaker, & Levy, 2008). We modified 4 items, so that half of the 20 items were negatively keyed.

Moral Disengagement. Moral Disengagement describes a set of cognitive processing styles (e.g., dehumanization, misattribution of responsibility and blame) that allow for behaving unethically without feeling distress (Bandura, 1999). It was represented via the 16-item Propensity to Morally Disengage Scale (Moore, Detert, Treviño, Baker, & Mayer, 2012). We modified 8 items, so that half of the items were negatively keyed.

Narcissism. Narcissism can be characterized by an exaggerated sense of self-importance. Whereas strong feelings of self-importance are shared among all narcissism conceptualizations, recent research suggests to distinguish between agentic, antagonistic, and neurotic narcissism, with antagonistic narcissism being most relevant for aggressive, exploitative acts at the costs of others (Back & Morf, in press; Crowe et al., in press). We considered items referring to antagonistic narcissism in particular, because this aspect of narcissism shows the strongest theoretical overlap with D. Narcissism was thus represented by the 9-item Narcissistic Rivalry subscale of the Narcissistic Admiration and Rivalry Questionnaire (Back et al., 2013), 5 items referring to grandiosity from the respective subscale of the Short Dark Triad, and 4 items related to leadership/authority and exploitativeness/entitlement from the Narcissistic Personality Inventory (Raskin & Hall, 1981). We modified 7 items, so that 8 of the 18 items were negatively keyed.

Psychological Entitlement. Psychological Entitlement is „a stable and pervasive sense that one deserves more and is entitled to more than others" (Campbell, Bonacci, Shelton, Exline, & Bushman, 2004, p. 31). It was represented using the respective 9-item scale by Campbell et al. (2004) and additional 3 items from the Entitlement Attitudes Questionnaire referring to revenge entitlement (Żemojtel-Piotrowska et al., 2017). We modified 5 items, so that half of the 12 items were negatively keyed.

Psychopathy. Psychopathy is characterized by impulsive hostility, i.e., deficits in affect and self-control. It was represented by 12 items from the Levenson Self-Report Psychopathy Inventory (LSRP; Levenson, Kiehl, & Fitzpatrick, 1995) tapping egocentricity (5), callousness (3), and anti-social tendencies (4), as well by 8 items from the respective subscale of the Short Dark Triad, leading to a total of 20 items. We did not consider items that describe consequences of Psychopathy (Skeem & Cooke, 2010) such as "I have never gotten into trouble with the law". We likewise did not consider items from the LSRP that either

imply greed (e.g. “Making a lot of money is my most important goal.”) or that refer to strategic manipulation (e.g., “I tell other people what they want to hear so that they will do what I want them to do.”), which is rather aligned with Machiavellianism (McHoskey, Worzel, & Szyarto, 1998). We modified 7 items, so that half of the 20 items were negatively keyed.

Sadism. Sadism is the tendency to engage in cruel, demeaning, or aggressive behaviors to promote one's own pleasure. It was represented by 13 items from the Assessment of Sadistic Personality Inventory (Plouffe, Saklofske, & Smith, 2017) and 7 items from the Short Sadistic Impulse Scale (O’Meara, Davies, & Hammond, 2011), leading to a total of 20 items. We excluded 3 items of the original scales because of duplicate item content and one item (“I have stolen from others without regard for the consequences.”) because its relation to Sadism is unclear. We modified 11 items, so that 11 items were negatively keyed.

Self-Centeredness. Self-Centeredness is the indifference or insensitiveness to the suffering and needs of others. It was represented by the respective 4-item measure provided by Arneklev, Grasmick, Tittle, and Bursik (1993). We modified 2 items, so that half of the items were negatively keyed.

Spitefulness. Spitefulness is the tendency to harm others for pleasure, even if this entails harm to oneself. It was represented via 16 items of the corresponding inventory by Marcus, Zeigler-Hill, Mercer, and Norris (2014). One item of the original scale was excluded, because it exhibited negative loadings on both D and the specific factor for Spitefulness in Moshagen et al. (2018). We modified 6 items, so that 7 items were negatively keyed.

Criterion Measures

Selection of the criteria was guided by the rationale that measures ought to represent (a) actual behavior involving utility maximization at the expense of others (such as cheating), (b) possible justifying beliefs (such as viewing the world as a competitive jungle), (c) relevant

outcomes in the realm of antagonistic, malevolent behavior (such as Crime/Delinquency), or (d) important psychological correlates (such as lack of Empathy). We assessed three behavioral measures (cheating, allocation decisions measuring Social Value Orientations, and a structurally similar, newly developed behavioral measure of sadism) in samples B3 and B4, respectively, and 8 self-report measures in sample B4. For all self-report measures (except for Crime and Analogous Behavior, which is scored dichotomously), we used a five-point Likert scale ranging from 1 = strongly disagree to 5 = strongly agree to maintain consistency.

Cheating Task. We employed a probabilistic cheating task as used widely in the behavioral ethics literature, the coin-toss task in the variant employed by Hilbig and Zettler (2015, Study 3). Specifically, following all other measures, participants were asked to perform three tedious and unenjoyable search tasks involving three sets of 100 English pseudowords (e.g., counting how often the letter "y" appeared). The three sets of pseudowords were shown on the screen to make participants aware of the unpleasantness of the task. Participants were informed that they could skip the search task if they reported to have won in a game of chance: Participants were instructed to take a coin, to toss it exactly twice, and that they may skip the boring search tasks if they reported having tossed exactly two heads. Participants could thus cheat by just reporting two heads, despite having obtained a different outcome (or not having tossed a coin at all). Importantly, participants were informed that they would always receive the complete payment for the study, regardless of whether they completed the boring search tasks or not. The observed outcome in this task is whether participants reported to have won in the game of chance (so that they skipped the boring search task) or whether they completed the boring search task consciously.

Social Value Orientations (SVO). SVO describe preferences for joint outcomes and cooperation vs. a pro-self orientation (Murphy & Ackermann, 2014). In each of the 6 primary tasks of the SVO slider measure (Murphy, Ackermann, & Handgraaf, 2011), participants

allocate points (worth 4 GBP per 100 points) between themselves and an unknown other. The extent of pro-social vs. pro-self behavior is measured as angle with lower degrees expressing more individualistic choices. Choices were fully incentive-compatible, that is, consequential for participants' bonus payment.

Sadism-SVO. To obtain a behavioral measure of sadistic tendencies, we created a SVO-type measure comprising 9 tasks specifically designed to discriminate competitive (i.e., allocations that maximize the difference between one's own and the other's outcome – at the cost of minimizing the joint outcome) from individualistic choices (allocations that maximize the own outcome). As in the SVO described above, each task required participants to allocate points (worth 4 GBP per 100 points) between themselves and an unknown other. Higher scores on this measure represent a stronger willingness to forgo own outcomes for the sake of even further reducing the outcome of the unknown other (i.e., more sadistic behavior). As with the SVO, this measure was fully consequential for participants' bonus payment.

Aggression. Aggression was measured via the physical (9 items; e.g., "If somebody hits me, I hit back.") and verbal (5 items; e.g. "When people annoy me, I may tell them what I think of them.") subscales of the Aggression Questionnaire (Buss & Perry, 1992).

Crime/Delinquency. Crime/Delinquency was measured by 10 items of the Crime and Analogous Behavior Scale (Miller & Lynam, 2003, p. 173). Each statement asks respondents to indicate whether they ever engaged in a certain behavior (such as stealing a car, attacking someone with the intent to injure, being arrested) by responding yes or no.

Competitive and Dangerous Worldviews. Competitive Jungle Worldviews are beliefs characterizing the world as a "ruthless, amoral struggle for resources and power in which might is right and winning everything" (Duckitt, Wagner, du Plessis, & Birum, 2002, p. 78). Dangerous and Threatening Worldviews are beliefs that "the social world is a dangerous and threatening place in which good, decent people's values and way of life are threatened by bad

people" (Duckitt et al., 2002, p. 78). We measured Competitive (e.g., "It's a dog-eat-dog world where you have to be ruthless at times.") and Dangerous (e.g., "There are many dangerous people in our society who will attack someone out of pure meanness, for no reason at all.") Worldviews by 6 items each according to Sibley and Duckit (2009).

Distrust. We assessed Distrust using the corresponding 10-item subscale (e.g., "I distrust people.") of the IPIP 16PF (Goldberg et al., 2006).

Dominance. Within the concept of social dominance orientation, the Dominance subdimension represents "a preference for group-based dominance hierarchies in which dominant groups actively oppress subordinate groups" (Ho et al., 2015, p. 1004). It was assessed using the respective 4-item subscale (e.g., "Some groups of people are simply inferior to other groups.") of the SDO_{7S} (Ho et al., 2015).

Empathy. Empathy was assessed via the 22-item version (e.g., "I really enjoy caring for other people.") of the Empathy Quotient (Baron-Cohen & Wheelwright, 2004; Wakabayashi et al., 2006).

Internalized Moral Identity. Internalized Moral Identity is a "self-conception organized around specific moral traits" (Aquino & Reed, 2002, p. 1424) and was assessed via the respective 5-item subscale of the measure of Moral Identity (Aquino & Reed, 2002). Participants are provided with a description of a "moral" person (e.g., caring, fair, kind) and are asked to indicate the importance of this description for participants' self-concept (e.g., "I strongly desire to have these characteristics.").

Statistical Analyses

Given the planned-missingness design in samples A1-A3, missing data can be assumed to be missing completely at random (MCAR), so we based all exploratory factor analyses (EFAs; using maximum-likelihood estimation) on a covariance matrix obtained using pairwise deletion (which is unbiased in the case of MCAR; Little & Rubin, 2002). In

the structural equation models, missing data in samples A2-A3 were addressed by using full information maximum likelihood estimation. There were no missing data in samples B1-B4, so we employed plain maximum likelihood (samples B1-B3) and robust-weighted least squares (to address the dichotomous response format of the Crime/Delinquency measure in sample B4), respectively. Huber-White sandwich estimated standard errors and corrected test-statistics ($T_{RML}^{(e1)}$ in Yang, Jiang, and Yuan, 2018) were applied to account for the effects of model size (e.g., Moshagen, 2012) and non-normality (e.g., Enders, 2001). In light of the extreme statistical power of the model chi-square test-statistic with sample sizes as large as in the present studies,[3] model fit was evaluated through the *SRMR* and the *RMSEA*. Note that we did not consider the *CFI*, because it conflates model misfit with (lack of) indicator reliability (i.e., loading magnitude), so that a superior approach is to consider indices that are pure indicators of model fit or pure indicators of construct reliability (Moshagen & Auerswald, 2018).

We specified a bifactor model (e.g., Reise, 2012) to model the hypothesized structure of D in line with the one employed in Moshagen et al. (2018). Briefly, bifactor models assume a general factor that affects all items (representing D) along with a set of specific factors that capture the remaining systematic covariance between items representing each specific dark trait. The specific factors and D were constrained to orthogonality, as were all item uniquenesses. The *ECV* (explained common variance) was used to gauge the relative proportion of common variance explained by D versus the specific dark traits (Ten Berge & Sočan, 2004), so that an ECV of 1 indicates that all common variance between the items representing a particular dark trait is explained by D.

[3] With N = 75,000, the power to detect misspecifications corresponding to $RMSEA \geq .001$ on $\alpha = .05$ is > 99% (Moshagen & Erdfelder, 2016).

The final item set was obtained based on a rational item-selection approach using meta-heuristics (Eisenbarth, Lilienfeld, & Yarkoni, 2015; Schroeders, Wilhelm, & Olaru, 2016), which have been shown to outperform traditional approaches (Olaru, Witthöft, & Wilhelm, 2015). The selection of items out a pool of possible candidates to optimize a criterion can be regarded as a combinatorial problem. Given that it is rarely possible to evaluate all permutations of items, suitable algorithms (such as genetic algorithms, GA; Holland, 1992) heuristically find a viable solution in a reasonable time by applying evolutionary principles. In brief, the GA optimizes fitness (a fitting function) over various generations (iterations) that change according to selection, cross-over, and mutation. At each iteration, various item sets are evaluated with respect to their value on the fitting function (fitness), the best (fittest) item sets (selection) are recombined (cross-over), some additional items are added and/or removed (mutations), and the resulting item sets reevaluated. This process continues until a maximum number of iterations is reached.

To estimate correlations involving the cheating variable, we needed to account for the fact that a certain proportion of those participants who indicated to have skipped the boring search tasks did so because they actually obtained two heads in the game of chance (i.e., they did not cheat). The observed outcome is therefore contaminated by honest (lucky) individuals, so that correlations involving this outcome need to be disattenuated using adapted procedures (Heck, Thielmann, Moshagen, & Hilbig, 2018; Moshagen & Hilbig, 2017).

Structural equation modeling was performed using Mplus (version 7.11; Muthén & Muthén, 2015), all other analyses in R. We used EFA as implemented in the psych package (Revelle, 2018), the RRreg package (Heck & Moshagen, 2018) to estimate correlations (and bootstrapped *p*-values) involving the cheating variable, an implementation of parallel analysis by Auerswald and Moshagen (in press), and the GA package (Scrucca, 2013).

Results

Item selection proceeded in a three-step process based on samples A1-A3. The resulting item sets were then evaluated and validated in samples B1-B4.

Preliminary Item Evaluation

Items from the initial pool of 184 items were considered for further inclusion based on their intercorrelations and their item-total correlations in sample A1 (N = 7,226). This led to the revision of 4 items and the exclusion of 27 items due to item-total correlations < .30. Concerning the items representing Psychopathy, we excluded 4 items related to disinhibition (e.g., “I avoid dangerous situations”), because this theme was only loosely related to the total score. To maintain a sufficient number of items representing Psychopathy, we added 5 newly developed items focusing on (low) impulse control. The modified item pool thus contained 162 items, of which 81 were negatively keyed (see Table 1 for the distribution of items across traits and the supplementary materials at https://osf.io/ag835 for a complete list of items).

First Item Selection

To determine the factorial structure of the modified item pool of 162 items, we performed an EFA using the data from sample A2 (N = 74,477). Parallel analysis based on the correlation matrix with 95th percentile reference eigenvalues (as recommended by Auerswald & Moshagen, in press) indicated the presence of 14 factors, which is closely aligned with the expectation of finding a general factor (D) along with a specific factor for each of the dark traits. The scree plot (Figure 1) indicates a strong first factor (eigenvalue 43.3) explaining 27% of the variance along with 13 minor factors (eigenvalues 1.3 – 4.7) explaining between 1% and 3% of the variance. Given that we were interested in measuring D, we extracted the first factor only, leading to item loadings ranging from .19 to .73 (mean = .50; see supplement for details).

Because employing the GA on the pool of 162 items did not yield reasonable solutions,[4] the item pool was further refined based on the EFA results by dropping 25 items with low loadings and another 19 items due to very high similarity to other items, while at the same time maintaining a balance between positively and negatively keyed items. The reduced item pool thus comprised 118 items (59 negatively keyed).

Final Item Selection

An EFA using sample A3 (N = 77,218) on the reduced itempool of 118 items yielded a strong general factor (eigenvalue 35.0) explaining 30% of the variance and 11 minor factors (eigenvalues 1.1 – 3.4) explaining 1% - 3% of the variance. Extracting the first factor led to item loadings ranging from .35 to .71 (mean = .53; see supplement). We then estimated a bifactor model for the 118 items by specifying a general factor for D and a specific factor for each dark trait capturing the remaining commonalities of the items representing that dark trait. The model yielded a good fit to the data, $\chi^2(6{,}667) = 226{,}089$, $p < .01$; *SRMR* = .040; *RMSEA* = .021 (90%-CI: .021 - .021); note, however, that the *RMSEA* is downwards biased in the presence of missing data, e.g., Davey, Savla, & Luo, 2005). All items significantly loaded on D (range: .34 - .73; mean .52; see supplement for details). The *ECV*s (Table 2) indicate that D is strongly reflected in the items of all dark traits, in particular in those representing Psychopathy (*ECV* = .92), Amoralism-Frustralia (*ECV* = .90), and Machiavellianism (*ECV* = .90). Indeed, the variance of the specific factor for Psychopathy was very small and did not

[4] Employing the GA on a very large number of items (with the goal to construct a single unidimensional measure) tends to lead to instable solutions due to the high dimensionality of the problem (i.e., the number of possible permutations). In addition, it proved difficult to define optimization criteria that minimize redundancies between the selected items. We thus further refined the item pool by removing items highly similar in content to other items and items that are unlikely to get selected due to low loadings, so that the resulting reduced item pool was better suited to employ the GA.

differ significantly from zero (p = .10), despite the very high power resulting from the large sample. On average over all items, D explained 78% of the common variance. These results verify that the majority of the retained items are well-suited indicators of D.

To obtain the final set of items, we used a rational item-selection approach based on the GA (as described above) using a custom fitness-function (Olaru, Schroeders, Hartung, & Wilhelm, in press) optimizing the following criteria: (1) maximize the correlation to the total score across the 118 items in the pool (to ensure breadth of content); (2) maximize the loadings on D (and, thus, reliability); (3) minimize the *RMSEA* and the *SRMR* to optimize model fit; (4) balance positively and negatively keyed items, both with respect of the number of items and with respect to the average loading; (5) ensure that items stem from diverse dark traits measures (require at least 3 items per dark trait measure to maintain breadth and limit to a maximum of 10 items per dark trait measure so that no single dark traits dominates the final measure); (6) ensure that items work similarly in different groups (i.e., across gender and age groups as well as across different countries).

We used several approximations in determining the criteria, because it did not prove feasible to estimate a bifactor model in the GA. We estimated an EFA model forcing the extraction of a single factor to determine loadings and model fit. Whereas doing so obviously neglects item information that is specific to a particular dark trait, this is appropriate and even necessary given that our intention was not to develop short scales for the specific dark traits, but rather a sound measure of D. Thus, items carrying content rather unrelated to or beyond D are of limited value for this purpose and can be dropped. In fact, using a single factor is well aligned with the purpose to identify items suited to measure D itself, as D is theoretically conceptualized as a unidimensional construct. To approximate (metric) measurement invariance across groups, we estimated a single factor model separately for the groups of interest (e.g., separately for females and males) and considered the standard deviation of each

item loading across groups as an indicator of differential item functioning (see Hartung, Doebler, Schroeders, & Wilhelm, 2018, for a similar approach).

Given that the GA does not necessarily yield the best (i.e., optimal) solution, we applied the algorithm 10 times to determine selection frequencies of each item (Olaru et al., in press). Beyond the GA results, final item selection was also guided by theoretical rationales to avoid redundancies and to ensure that D is measured in sufficient breath. This procedure resulted in a final set of 70 items (35 negatively keyed), which we refer to as D70. The finally retained items exhibited very high correlations to the pool of 118 (r = .99) and 162 items (r = .97), respectively, indicating that only little content was lost in item selection.

Further, we applied the GA in the same way as above using the items retained in the D70 as the initial item pool to obtain two shorter versions. Based on these results (and additionally applying theoretical rationales) we obtained a set of 35 items (17 negatively keyed), which we refer to as D35. The same approach was pursued (using the D35 as a basis) to obtain a 16-item version (D16; 8 negatively keyed).

Confirmatory Factor Analyses

We first verified the factor structure of the D70 via bifactor modeling using new data. To this end, we pooled the data of the four validation samples (B1 to B4; using the first measurement occasion of sample B2) yielding a combined sample size of N = 6,838.

Estimation of the full bifactor model specifying a general (D) and 12 specific factors did not converge successfully because some specific factors failed to exhibit any variance. Specific factors with (near) zero variance occur when the common variance of the associated items is entirely explained by D, so that no covariance remains to be explained by the specific factor. We thus omitted the specific factors for the items representing Egoism, Machiavellianism, Moral Disengagement, Psychological Entitlement, Psychopathy, and Spitefulness. The revised model (which now comprised a general factor along with six

specific factors) yielded an adequate fit to the data, $\chi^2(2{,}310) = 31{,}085$, $p < .01$; *SRMR* = .040; *RMSEA* = .043 (90%-CI: .042 - .043). Most loadings on D were of an average magnitude (range: .39 - .70; mean .56; Table 3). The items exhibiting the 10 strongest loadings on D stemmed from seven different dark trait measures. The *ECV*s (Table 2) are slightly higher compared to those obtained with the pool of 118 items, which was to be expected, given that items were selected with respect to their ability to indicate D rather than the more specific dark traits (which also leads to some specific traits exhibiting no variance at all). Across all items, D explained an average of 85% of the common variance.

We also used the pooled validation sample to evaluate the factor structure of the D35 and the D16 by only considering the responses to items retained in these versions. In light of the comparatively small number of items that arguably do not carry sufficient information to obtain meaningful specific factors, we fitted a standard single-factor confirmatory factor model to the data. Model fit was acceptable for both, the D35, $\chi^2(560) = 10{,}366$, $p < .01$; *SRMR* = .038; *RMSEA* = .050 (90%-CI: .050 - .051), and the D16, $\chi^2(104) = 2{,}238$, $p < .01$; *SRMR* = .034; *RMSEA* = .055 (90%-CI: .053 - .057). All items exhibited substantial loadings on D in both versions (Table 3; mean loading D35 = .59; mean loading D16 = .62). In addition, the observed correlations of the D35 and the D16 to the D70 (r = .98 and r = .95, respectively) show that the shorter versions are able to provide a close approximation of the D70, in sum suggesting that both versions are reasonable shorter measures of D.

Internal Consistencies

We computed Cronbach's α estimates of internal consistency in the pooled validation sample for the D70 as well as for the D35 and the D16 by considering the respective relevant items. Internal consistency of the D70 was excellent (α = .970 [95%-CI: .969 - .971]) and was very high for the shorter versions (D35: α = .950 [.948 - .952]; D16: α = .906 [.903 - .909]).

Retest Reliability

Retest-reliability over approximately 34 days was determined using participants from sample B2 who completed both measurement occasions (N = 77). Retest-reliability of the D70 was r_{tt} = .95 (95%-CI: .92 - .97). We further considered only the responses to the items contained in the D35 and the D16, respectively, to obtain an estimate of retest-reliability of the shorter versions, resulting in r_{tt} = .93 [.89 - .96] for the D35 and r_{tt} = .90 [.85 - .94] for the D16. Retest-reliability of all versions thus proved to be high.

Criterion-related Validity

We first consider the behavioral outcomes assessed herein, i.e., the ability of the D70 to predict own utility maximization disregarding (cheating), accepting (SVO), or malevolently provoking (Sadism-SVO) disutility for others. In line with the conceptualization of D, the D70 was significantly associated with dishonest behavior, r = .34, a pro-social vs. pro-self orientation, r = -.43, and with the behavioral measure of sadism, r = .22. Highly similar correlations to the behavioral outcomes were obtained when using the D35 or the D16 (Table 4), indicating that all versions are well suited to predict actual behavior.

The D70 also correlated substantially with all self-reported criteria under consideration in the expected direction (Table 4; $.28 \leq |r| \leq .92$). Again, the correlations to the criteria were similar when computed based on the subsets of items contained in the D35 and D16, respectively, thereby supporting nomological consistency (Hilbig, Moshagen, & Zettler, 2016; Thielmann & Hilbig, in press). More specifically, the justifying beliefs aspect of D was reflected in substantial correlations with a preference for group-based dominance hierarchies ($.59 \leq r \leq .61$), Vile World Beliefs ($.21 \leq r \leq .28$), and, in particular, Competitive Jungle Worldviews ($.91 \leq r \leq .93$). The importance of D for real-word behavior was supported by substantial relations to Aggression ($.65 \leq r \leq .67$) and Crime/Delinquency ($.32 \leq r \leq .37$). Finally, all item sets displayed theoretically sound relationships to relevant psychological

covariates, suggesting that D is characterized by Distrust ($.49 \leq r \leq .51$), (lack of) Empathy ($-.31 \geq r \geq -.37$), and a low importance of Moral Identity for the self ($-.65 \geq r \geq -.73$).

General Discussion

One way to explain and describe patterns of antagonistic, malevolent, and socially aversive behavior is to assume stable predispositions towards such behaviors. A large body of evidence indicates that such dark traits are closely linked to psychological adjustment and interpersonal behavior in a variety of domains. In light of the empirical and theoretical overlap between dark traits, Moshagen et al. (2018) argued that all dark traits arise from the same underlying dispositional tendency "to maximize one's individual utility—disregarding, accepting, or malevolently provoking disutility for others—, accompanied by beliefs that serve as justifications" (p. 657), which they called the Dark Factor of Personality (D). Whereas Moshagen et al. (2018) and Zettler et al. (2019) provided ample evidence in favor of the conceptualization of D, the actual assessment of D was a rather cumbersome matter so far. We tackled this issue in the present research by identifying sets of items that can be used to measure D more concisely while exhibiting superior psychometric properties compared to the assessment relied upon in previous studies.

The resulting item sets—which we refer to as D70, D35, and D16—are characterized by high internal consistencies and high retest-reliabilities, balance positively and negatively keyed items, exhibit substantial associations to a variety of relevant criteria, including three behavioral outcomes, and clearly map on a single factor representing D. Indeed, exploratory factor analyses as well as bifactor modeling on a large number of items strongly suggested the presence of a single factor that captures most of the communalities of the indicators, thus once more providing evidence in favor of the notion of D as the basic disposition underlying dark traits. Likewise, the specific factors exhibited only little or no variance beyond D. Note, however, that the latter is likely also a consequence of selecting items with respect to their

ability to indicate D rather than the more specific dark traits. That is, items carrying content specific to a particular trait were less likely retained in the final item sets as ought to be the case given the present goal (namely, providing a sound measure for D). In turn, the specific dark traits when measured via the identified item sets may lack features that are relevant to said traits (but largely unrelated to D). For example, several conceptualizations of Psychopathy involve the aspect of disinhibition (e.g., Patrick, Fowles, & Krueger, 2009). This aspect, however, largely lies beyond the theoretical scope of D, so that corresponding items were not retained for the measures of D. By implication, measuring Psychopathy with this particular set of items will not represent Psychopathy in its full breadth (though it may serve as a sufficient approximation). In any case, we do not recommend using the item sets identified herein as measures of the more specific traits. Rather, if researchers are primarily interested in a specific dark trait (e.g., Psychopathy), the most reasonable approach is to seek an appropriate capturing said construct in particular.

In this context, the question concerning the role of the more specific dark traits beyond D arises. Whereas D is assumed to represent the basic disposition responsible for the emergence of any more specific dark trait, so that the malevolent parts of any dark trait are due to D, specific dark traits typically also comprise features that lie outside the scope of D. Sticking to the example from above, disinhibition is only partly reflected in D, because neither a lack of planfulness, a preference for immediate rewards, nor a lack of inhibitory control imply malevolent behavior per se (and are thus not well aligned with the conceptualization of D). Nevertheless, disinhibition may of course contribute to the way individuals exhibit a particular malevolent behavior. For example, whereas all Dark Triad components are related to cheating behavior under a low risk of getting caught, only Psychopathy has been found to continue predicting cheating behavior when there is a high risk of receiving punishment (Jones & Paulhus, 2017). It thus seems plausible that more

specific dark traits (such as Psychopathy) outperform D in predicting more specific outcomes. At the same time, however, the malevolent parts of any dark trait are comprised in D.

It should also be noted that the question of the dimensionality of many dark traits is far from settled (e.g., Miller, Vize, Crowe, & Lynam, in press; Muris et al., 2017). However, this issue is largely irrelevant with respect to (the measurement of) D, because D is agnostic regarding the dimensionality of a specific manifestation. For reasons of parsimony, we assumed unidimensionality for each specific dark trait in the bifactor models and likewise assumed that each specific factor reflects the specifics of a single dark trait. Nevertheless, it is entirely possible that, say, Psychopathy (when measured broadly) comprises several distinguishable subdimensions beyond D, so it might also be reasonable to specify more than one specific factor representing various subdimensions of a single dark trait. Concerning D and the measurement thereof, however, the crucial aspect is the unidimensionality of D itself.

A related issue pertains to the selection of instruments (and items) for our initial itempool. For some of the considered traits, in particular the Dark Triad components, various measures exist that often differ substantially in length, scope, and even their theoretical underpinning. For example, there is a notable number of established and validated self-report measures of Psychopathy in adulthood, including the LSRP (Levenson et al., 1995), the Elemental Psychopathy Assessment (Lynam et al., 2011), the Psychopathic Personality Inventory-Revised (Lilienfeld & Widows, 2005), the Self-Report Psychopathy Scale (Paulhus, Neumann, & Hare, in press), and the Triarchic Psychopathy Measure (Drislane, Patrick, & Arsal, 2014), as well as scales measuring Psychopathy embedded in inventories additionally assessing other traits, such as the Short-Dark Triad (Jones & Paulhus, 2014), the Dirty Dozen (Jonason & Webster, 2010), or the antisocial features subscale in the Personality Assessment Inventory (Morey, 2007). Although many (but not all) of these measures show adequate convergent validity (e.g., Miller & Lynam, 2012; Tsang, Salekin, Coffey, & Cox,

2018), each represents a unique approach to the assessment of Psychopathy, which also reflects "competing conceptualizations of the psychopathy construct itself" (Watts, Waldman, Smith, Poore, & Lilienfeld, 2017, p. 953). Similar issues are evident in other dark traits as well, in particular concerning the remaining components of the Dark Triad, i.e., Narcissism (e.g., Cain, Pincus, & Ansell, 2008; Crowe et al., in press; Kaufman, Weiss, Miller, & Campbell, in press; Krizan & Herlache, 2018) and Machiavellianism (see Miller et al., in press). Consequently, the approach taken herein to merely consider a single or a few instruments to obtain items representing a particular dark trait is—necessarily—somewhat arbitrary. However, the fluid nature of D implies that the indicators to assess D are ultimately interchangeable, so that any dark trait measure (with sound psychometric properties and sufficient breadth in content) is generally suited to indicate D.

Validity

The D70, and the shorter versions as well, exhibited a high degree of criterion-related validity. Most importantly, substantial relations to all three behavioral outcomes reflecting the utility-based aspect of D were observed. D was related to behavioral dishonesty as measured in a cheating task (representing the tendency to maximize own utility *disregarding* other's disutility), the SVO expressing the tendency to prefer pro-social over pro-self outcomes (the tendency to maximize own utility *accepting* other's disutility), and behavioral sadism (the tendency to derive own utility from malevolently *provoking* disutility on others). Similarly, the notion that individuals with elevated levels on D hold beliefs that serve to justify malevolent behavior was mirrored by the tendency of such individuals to endorse beliefs characterizing the world as a competitive jungle or as a dangerous and threatening place, and to favor dominance hierarchies of social groups. Thus, the obtained item sets exhibit associations to outcomes consistent with the very theoretical definition of D.

Beyond the links to outcomes representing key characteristics of the theoretical definition of D, its relevance for real-world behavior was further evident in substantial associations with (self-reported) aggression as well as with (self-reported) criminal and deviant behaviors. The obtained item sets also exhibited theoretically reasonable links to important psychological correlates, characterizing high-D individuals by a lack of empathy, a lower importance of moral-identity for the self-concept, and elevated distrust towards others. In sum, the present studies provide ample evidence concerning the validity of D and the identified item sets to measure D.

Modeling D

Several latent variable models can be reasonably applied to model the D70. In general, the theoretical model purported herein (and detailed in Moshagen et al., 2018) is based on a bifactor model with the general factor representing D and one specific factor for each specific dark trait capturing the remaining covariances of the items designed to assess said trait. However, the results indicate that such a model may lead to some dark traits empirically not exhibiting any specific variance beyond D at all, so that reduced models omitting these specific factors should be considered (as we have done herein). Note that omitting specific factors that exhibit no variance does not change the meaning of the general factor representing D.

An alternative approach to bifactor modeling are higher-order factor models, where the lower-order factors (representing each specific dark trait) act as indicator variables for the second-order factor (representing D). Higher-order models are nested special cases of bifactor models, so that both are closely related and difficult to distinguish empirically (e.g., Brunner, Nagy, & Wilhelm, 2012; Mansolf & Reise, 2017). Indeed, the general factor arising in a bifactor model is often virtually identical to the second order factor arising in a higher-order model. Correspondingly, although the theoretical conceptualization of D is closer aligned

with a bifactor approach for reasons outlined in Moshagen et al. (2018, pp. 661-662), both modeling approaches are generally suited when the prime interest lies on D. However, bifactor and higher-order models strongly differ regarding the meaning of the specific versus the lower-order factors. In bifactor models, the specific dark traits emerge residualized for the general (D) factor, so that these often differ greatly in meaning from the non-residualized primary factors as obtained in higher order models (e.g., Sellbom & Tellegen, 2019). Thus, higher-order models can be of value whenever one is interested in obtaining estimates for the specific dark traits not residualized for D.

Largely unrelated to whether a bifactor or a higher-order model is considered, a further question pertains to the definition of the specific or lower-order factors. As elaborated above, we assumed that each specific factor reflects the specifics of a single dark trait in line with the approach pursued in Moshagen et al. (2018). However, alternative specifications are possible. For example, various subdimensions of a single dark (say, boldness, meanness, and disinhibition according to the triarchic conceptualization of Psychopathy; Patrick et al., 2009) could be represented by one factor each. Also, one may attempt to determine common themes that are shared across various specific dark traits (e.g., Watts et al., 2017), so that one specific factor represents a certain theme or facet, rather than a specific dark trait. A useful approach in this regard might rely on a series of nested exploratory bifactor models to unfold the structure of the specifics beyond D (Goldberg et al., 2006), which clearly is an interesting avenue for further research.

The considerations above apply to the D70. Concerning the D35 and the D16, however, it seems most reasonable to apply a simple single factor model representing D, given that the retained items arguably do not allow for specifying meaningful specific (in a bifactor model) or lower-order (in a higher-order model) factors. It might be reasonable to allow correlations between some item residuals, because some items in the shorter version

still refer to the same specific dark trait, however, for reasons of parsimony we recommend doing so only in light of substantial evidence.

Limitations

Some limitations should also be considered. Although online data collection is considered largely unproblematic (Miller, Crowe, Weiss, Maples-Keller, & Lynam, 2017; Weigold, Weigold, & Russell, 2013), future studies should generalize the results to offline samples assessing D in a paper-pencil format. Moreover, D was always assessed using self-reports, so that some of the results might be distorted owing to socially desirable responding (e.g., Tourangeau & Yan, 2007). While we are currently in the process of obtaining data allowing for the determination of the extent of self-observer agreement, it would be beneficial to consider peer-reports as a complementary data source (McAbee & Connelly, 2016).

A particular strength of present studies is the use of very large and highly diverse samples, including a substantial proportion of individuals from societies that are not characterized as western, educated, industrialized, rich, and democratic (WEIRD; Henrich, Heine, & Norenzayan, 2010). For instance, across samples, more than 3,000 individuals originated from Thailand, more than each 2,000 from Brazil and India, and more than 1,000 each from Latvia, the Philippines, Romania, and Singapore. This strength, however, also comes at the cost that many participants were not native in English, which might have led to subtle shifts in the comprehension of some items. In this regard, it would be important to provide the item sets in languages other than English. Likewise, one criterion during item selection explicitly took differential item functioning across groups, including cultures, into account. However, we did not conduct a formal test of measurement invariance across groups; neither did we investigate cultural differences in detail.

Given the focus of the present work to identify items that can be used to measure D, we also did not explicitly consider alternative theoretical propositions concerning the

communalities of dark traits beyond the one by Moshagen et al. (2018). For instance, Jones and Figueredo (2013) suggested that callous affect and interpersonal manipulation are the shared features among the dark triad components. On a theoretical level, D would be expected to cover additional aspects beyond callousness and manipulation, in particular by accounting for malevolent behavior that does not involve manipulation of others (such as exploitation), by including antagonist behavior even at own costs, and by emphasizing the importance of justifying beliefs (Moshagen et al., 2018, pp. 659-660). However, these theoretical predictions require additional empirical tests which lie outside the present scope and arguably require a sound measure of D first. Similarly, we herein refrained from relating D to more basic dimensions of personality or psychopathology, in particular regarding Agreeableness according to the Five-Factor Model (e.g., McCrae & Costa, 2008), Honesty-Humility according to the HEXACO Model (e.g., Ashton & Lee, 2007), and Antagonism according to the DSM-V (e.g., Krueger, Derringer, Markon, Watson, & Skodol, 2012). It would be interesting to examine these concepts in conjunction with each other, additionally considering several behavioral outcomes to establish their similarities and differences. Clearly, tackling any of these issues requires a psychometrically and more concise measure of D—as provided in the present work.

Conclusion

These issues notwithstanding, we herein identified sets of items that are well suited to indicate D in a psychometrically sound way, clearly map on a single factor in line with the definition of D, and exhibit a high degree of validity as evidenced by the relations to various criterion measures, including three behavioral outcomes. Thereby, we go far beyond the rather cumbersome assessment of D as relied upon in previous studies. We are confident that the provision of item sets with different length—with the D70 providing the most comprehensive

representation of D and the shorter D35 and D16 as reasonable proxies when there are constraints on the assessment time—may stipulate further research on dark traits and D.

References

Aquino, K., & Reed, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, *83*, 1423–1440.

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

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review, 11*, 150–166.

Auerswald, M., & Moshagen, M. (in press). How to determine the number of factors to retain in exploratory factor analysis: A comparison of extraction methods under realistic conditions. *Psychological Methods*. https://doi.org/10.1037/met0000200

Back, M. D., Küfner, A. C., Dufner, M., Gerlach, T. M., Rauthmann, J. F., & Denissen, J. J. (2013). Narcissistic admiration and rivalry: Disentangling the bright and dark sides of narcissism. *Journal of Personality and Social Psychology*, *105*, 1013–1037.

Back, M. D., & Morf, C. C. (in press). Narcissism. In V. Zeigler-Hill & T. K. Shackelford, *Encyclopedia of personality and individual differences*. New York: Springer.

Bandura, A. (1999). Moral Disengagement in the Perpetration of Inhumanities. *Personality and Social Psychology Review*, *3*, 193–209. https://doi.org/10.1207/s15327957pspr0303_3

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

Blonigen, D. M., Hicks, B. M., Krueger, R. F., Patrick, C. J., & Iacono, W. G. (2005). Psychopathic personality traits: Heritability and genetic overlap with internalizing and externalizing psychopathology. *Psychological Medicine*, *35*, 637–648. https://doi.org/10.1017/S0033291704004180

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality*, *80*, 796–846.

Buss, A. H., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology*, *63*, 452–459.

Cain, N. M., Pincus, A. L., & Ansell, E. B. (2008). Narcissism at the crossroads: Phenotypic description of pathological narcissism across clinical theory, social/personality psychology, and psychiatric diagnosis. *Clinical Psychology Review*, *28*, 638–656. https://doi.org/10.1016/j.cpr.2007.09.006

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*, 29–45.

Cheung, G. W., & Rensvold, R. B. (2000). Assessing Extreme and Acquiescence Response Sets in Cross-Cultural Research Using Structural Equations Modeling. *Journal of Cross-Cultural Psychology*, *31*, 187–212. https://doi.org/10.1177/0022022100031002003

Christie, R., & Geis, F. L. (1970). *Studies in machiavellianism*. New York: Academic Press.

Crowe, M. L., Lynam, D. R., Campbell, W. K., & Miller, J. D. (in press). Exploring the structure of narcissism: Toward an integrated solution. *Journal of Personality*. https://doi.org/10.1111/jopy.12464

Dahling, J. J., Whitaker, B. G., & Levy, P. E. (2008). The development and validation of a new Machiavellianism scale. *Journal of Management, 35*, 219–257.

Davey, A., Savla, J., & Luo, Z. (2005). Issues in evaluating model fit with missing data. *Structural Equation Modeling, 12*, 578–597.

Diebels, K. J., Leary, M. R., & Chon, D. (2018). Individual Differences in Selfishness as a Major Dimension of Personality: A Reinterpretation of the Sixth Personality Factor. *Review of General Psychology, 22*, 367–376.

DiStefano, C., & Motl, R. W. (2006). Further Investigating Method Effects Associated With Negatively Worded Items on Self-Report Surveys. *Structural Equation Modeling, 13*, 440–464. https://doi.org/10.1207/s15328007sem1303_6

Drislane, L. E., Patrick, C. J., & Arsal, G. (2014). Clarifying the content coverage of differing psychopathy inventories through reference to the Triarchic Psychopathy Measure. *Psychological Assessment, 26*, 350–362. https://doi.org/10.1037/a0035152

Duckitt, J., Wagner, C., du Plessis, I., & Birum, I. (2002). The psychological bases of ideology and prejudice: testing a dual process model. *Journal of Personality and Social Psychology, 83*, 75–93.

Eisenbarth, H., Lilienfeld, S. O., & Yarkoni, T. (2015). Using a genetic algorithm to abbreviate the Psychopathic Personality Inventory–Revised (PPI-R). *Psychological Assessment, 27*, 194–202. https://doi.org/10.1037/pas0000032

Enders, C. K. (2001). The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychological Methods, 6*, 352–370. https://doi.org/10.1037/1082-989X.6.4.352

Gerbasi, M. E., & Prentice, D. A. (2013). The Self-and Other-Interest Inventory. *Journal of Personality and Social Psychology*, *105*, 495–514.

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

Harrop, T. M., Preston, O. C., Khazem, L. R., Anestis, M. D., Junearick, R., Green, B. A., & Anestis, J. C. (2017). Dark traits and suicide: Associations between psychopathy, narcissism, and components of the interpersonal–psychological theory of suicide. *Journal of Abnormal Psychology*, *126*, 928–938. https://doi.org/10.1037/abn0000300

Hartung, J., Doebler, P., Schroeders, U., & Wilhelm, O. (2018). Dedifferentiation and differentiation of intelligence in adults across age and years of education. *Intelligence*, *69*, 37–49. https://doi.org/10.1016/j.intell.2018.04.003

Heck, D. W., & Moshagen, M. (2018). RRreg: An R Package for Correlation and Regression Analyses of Randomized Response Data. *Journal of Statistical Software*, *85*(2), 1–29. https://doi.org/10.18637/jss.v085.i02

Heck, D. W., 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*, 356–371.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*(7302), 29–29. https://doi.org/10.1038/466029a

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

Hilbig, B. E., & Zettler, I. (2015). When the cat's away, some mice will play: A basic trait account of dishonest behavior. *Journal of Research in Personality*, *57*, 72–88.

Ho, A. K., Sidanius, J., Kteily, N., Sheehy-Skeffington, J., Pratto, F., Henkel, K. E., … Stewart, A. L. (2015). The nature of social dominance orientation: Theorizing and measuring preferences for intergroup inequality using the new SDO$_7$ scale. *Journal of Personality and Social Psychology*, *109*, 1003–1028. https://doi.org/10.1037/pspi0000033

Holland, J. H. (1992). Genetic Algorithms. *Scientific American*, *267*, 66–73.

Jensen, A. R. (1992). Commentary: Vehicles of g. *Psychological Science*, *3*, 275–279. https://doi.org/10.1111/j.1467-9280.1992.tb00671.x

Jonason, P. K., Li, N. P., Webster, G. D., & Schmitt, D. P. (2009). The dark triad: Facilitating a short-term mating strategy in men. *European Journal of Personality*, *23*, 5–18.

Jonason, P. K., & Webster, G. D. (2010). The Dirty Dozen: A Concise Measure of the Dark Triad. *Psychological Assessment*, *22*, 420–432.

Jones, D. N., & Figueredo, A. J. (2013). The Core of Darkness: Uncovering the Heart of the Dark Triad. *European Journal of Personality*, *27*, 521–531.

Jones, D. N., & Paulhus, D. L. (2014). Introducing the Short Dark Triad (SD3): A Brief Measure of Dark Personality Traits. *Assessment*, *21*, 28–41.

Jones, D. N., & Paulhus, D. L. (2017). Duplicity among the dark triad: Three faces of deceit. *Journal of Personality and Social Psychology*, *113*, 329–342. https://doi.org/10.1037/pspp0000139

Kaufman, S. B., Weiss, B., Miller, J. D., & Campbell, W. K. (in press). Clinical Correlates of Vulnerable and Grandiose Narcissism: A Personality Perspective. *Journal of Personality Disorders*. https://doi.org/10.1521/pedi_2018_32_384

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

Krekels, G., & Pandelaere, M. (2015). Dispositional greed. *Personality and Individual Differences*, *74*, 225–230.

Krizan, Z., & Herlache, A. D. (2018). The Narcissism Spectrum Model: A Synthetic View of Narcissistic Personality. *Personality and Social Psychology Review*, *22*, 3–31. https://doi.org/10.1177/1088868316685018

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

Lange, J., Paulhus, D. L., & Crusius, J. (2018). Elucidating the Dark Side of Envy: Distinctive Links of Benign and Malicious Envy With Dark Personalities. *Personality and Social Psychology Bulletin*, *44*, 601–614. https://doi.org/10.1177/0146167217746340

Levenson, M. R., Kiehl, K. A., & Fitzpatrick, C. M. (1995). Assessing psychopathic attributes in a noninstitutionalized population. *Journal of Personality and Social Psychology*, *68*, 151–158. https://doi.org/10.1037//0022-3514.68.1.151

Lilienfeld, S. O., & Widows, M. R. (2005). *Psychopathy Personality Inventory Revised (PPI-R) Manual*. Lutz: Psychological Assessment Resources.

Little, R. J., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York: Wiley.

Lynam, D. R., Gaughan, E. T., Miller, J. D., Miller, D. J., Mullins-Sweatt, S., & Widiger, T. A. (2011). Assessing the basic traits associated with psychopathy: Development and validation of the Elemental Psychopathy Assessment. *Psychological Assessment*, *23*(1), 108–124. https://doi.org/10.1037/a0021146

Mansolf, M., & Reise, S. P. (2017). When and why the second-order and bifactor models are distinguishable. *Intelligence*, *61*, 120–129.

Marcus, D. K., & Zeigler-Hill, V. (2015). A Big Tent of Dark Personality Traits. *Social and Personality Psychology Compass*, *98*, 434–446.

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*, 563–574.

McAbee, S. T., & Connelly, B. S. (2016). A multi-rater framework for studying personality: The trait-reputation-identity model. *Psychological Review*, *123*, 569–591. https://doi.org/10.1037/rev0000035

McCrae, R. R., & Costa, P. T. (2008). The five-factor theory of personality. In O. P. John, R. W. Robins, & L. A. Pervin, *Handbook of personality psychology: Theory and research* (3rd ed., pp. 159–181). New York, NY: Guilford Press.

McHoskey, J. W., Worzel, W., & Szyarto, C. (1998). Machiavellianism and psychopathy. *Journal of Personality and Social Psychology*, *74*, 192–210.

Miller, J. D., Crowe, M., Weiss, B., Maples-Keller, J. L., & Lynam, D. R. (2017). Using online, crowdsourcing platforms for data collection in personality disorder research: The example of Amazon's Mechanical Turk. *Personality Disorders: Theory, Research, and Treatment*, *8*, 26–34.

Miller, J. D., & Lynam, D. R. (2003). Psychopathy and the Five-Factor Model of Personality: A Replication and Extension. *Journal of Personality Assessment*, *81*, 168–178. https://doi.org/10.1207/S15327752JPA8102_08

Miller, J. D., & Lynam, D. R. (2012). An examination of the Psychopathic Personality Inventory's nomological network: A meta-analytic review. *Personality Disorders: Theory, Research, and Treatment*, *3*, 305–326. https://doi.org/10.1037/a0024567

Miller, J., Vize, C., Crowe, M. L., & Lynam, D. (in press). A critical appraisal of the Dark Triad literature and suggestions for moving forward. *Current Directions in Psychological Science*.

Monaghan, C., Bizumic, B., & Sellbom, M. (2016). The role of Machiavellian views and tactics in psychopathology. *Personality and Individual Differences*, *94*, 72–81.

Moore, C., Detert, J. R., Treviño, L., Baker, V. L., & Mayer, D. M. (2012). Why Employees Do Bad Things: Moral Disengagement and Unethical Organizational Behavior. *Personnel Psychology*, *65*, 1–48.

Morey, L. C. (2007). *Personality Assessment Inventory: Professional Manual*. Odessa, FL: Psychological Assessment Resources.

Moshagen, M. (2012). The Model Size Effect in SEM: Inflated Goodness-of-Fit Statistics Are Due to the Size of the Covariance Matrix. *Structural Equation Modeling*, *19*, 86–98.

Moshagen, M., & Auerswald, M. (2018). On congruence and incongruence of measures of fit in structural equation modeling. *Psychological Methods*, *23*, 318–336.

Moshagen, M., & Erdfelder, E. (2016). A New Strategy for Testing Structural Equation Models. *Structural Equation Modeling*, *23*, 54–60.

Moshagen, M., & Hilbig, B. E. (2017). The statistical analysis of cheating paradigms. *Behavior Research Methods*, *49*, 724–732. https://doi.org/10.3758/s13428-016-0729-x

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

Muris, P., Merckelbach, H., Otgaar, H., & Meijer, E. (2017). The malevolent side of human nature: A meta-analysis and critical review of the literature on the Dark Triad (Narcissism, Machiavellianism, and Psychopathy). *Perspectives on Psychological Science*, *12*, 183–204.

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*, *18*, 13–41.

Murphy, R. O., Ackermann, K. A., & Handgraaf, M. (2011). Measuring Social Value Orientation. *Judgment and Decision Making*, *6*, 771–781. https://doi.org/10.2139/ssrn.1804189

Muthén, L. K., & Muthén, B. (2015). *Mplus user's guide* (7th ed.). Los Angeles: Muthén & Muthén.

O'Boyle, E. H., Forsyth, D. R., Banks, G. C., & McDaniel, M. A. (2012). A meta-analysis of the dark triad and work behavior: A social exchange perspective. *Journal of Applied Psychology*, *97*, 557–579.

O'Boyle, E. H., Forsyth, D. R., Banks, G. C., Story, P. A., & White, C. D. (2015). A Meta-Analytic Test of Redundancy and Relative Importance of the Dark Triad and Five-Factor Model of Personality. *Journal of Personality*, *83*, 644–664.

Olaru, G., Schroeders, U., Hartung, J., & Wilhelm, O. (in press). Ant Colony Optimization and Local Weighted Structural Equation Modeling. A Tutorial on Novel Item and Person Sampling Procedures for Personality Research. *European Journal of Personality*. https://doi.org/10.1002/per.2195

Olaru, G., Witthöft, M., & Wilhelm, O. (2015). Methods matter: Testing competing models for designing short-scale Big-Five assessments. *Journal of Research in Personality*, *59*, 56–68. https://doi.org/10.1016/j.jrp.2015.09.001

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

Patrick, C. J., Fowles, D. C., & Krueger, R. F. (2009). Triarchic conceptualization of psychopathy: Developmental origins of disinhibition, boldness, and meanness.

Development and Psychopathology, *21*(3), 913–938. https://doi.org/10.1017/S0954579409000492

Paulhus, D. L., & Jones, D. N. (2014). Measures of Dark Personalities. In G. J. Boyle, D. H. Saklosfske, & G. Matthews, *Measures of personality and social psychological constructs* (pp. 562–594). San Diego: Academic Press.

Paulhus, D. L., Neumann, C. S., & Hare, R. D. (in press). *Manual for the Hare Self-Report Psychopathy scale*. Toronto: Multi-Health Systems.

Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, *36*, 556–563.

Plouffe, R. A., Saklofske, D. H., & Smith, M. M. (2017). The Assessment of Sadistic Personality: Preliminary psychometric evidence for a new measure. *Personality and Individual Differences*, *104*, 166–171. https://doi.org/10.1016/j.paid.2016.07.043

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*, 879–903.

Raskin, R., & Hall, C. S. (1981). The Narcissistic Personality Inventory: Alternative Form Reliability and Further Evidence of Construct Validity. *Journal of Personality Assessment*, *45*, 159–162. https://doi.org/10.1207/s15327752jpa4502_10

Reise, S. P. (2012). The Rediscovery of Bifactor Measurement Models. *Multivariate Behavioral Research*, *47*, 667–696.

Rentzsch, K., & Gross, J. J. (2015). Who Turns Green with Envy? Conceptual and Empirical Perspectives on Dispositional Envy: Dispositional envy. *European Journal of Personality*, *29*, 530–547. https://doi.org/10.1002/per.2012

Revelle, W. (2018). *psych: Procedures for Personality and Psychological Research [R package]*. Retrieved from https://cran.r-project.org/package=psych

Revelle, W., Wilt, J., & Rosenthal, A. (2010). Individual Differences in Cognition: New Methods for Examining the Personality-Cognition Link. In A. Gruszka, G. Matthews, & B. Szymura (Eds.), *Handbook of Individual Differences in Cognition* (pp. 27–49). https://doi.org/10.1007/978-1-4419-1210-7_2

Schroeders, U., Wilhelm, O., & Olaru, G. (2016). Meta-Heuristics in Short Scale Construction: Ant Colony Optimization and Genetic Algorithm. *PLOS ONE*, *11*, e0167110. https://doi.org/10.1371/journal.pone.0167110

Scrucca, L. (2013). GA: A Package for Genetic Algorithms in R. *Journal of Statistical Software*, *53*(4). https://doi.org/10.18637/jss.v053.i04

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

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

Sibley, C. G., & Duckitt, J. (2009). Big-Five Personality, Social Worldviews, and Ideological Attitudes: Further Tests of a Dual Process Cognitive-Motivational Model. *Journal of Social Psychology*, *149*, 545–561. https://doi.org/10.1080/00224540903232308

Skeem, J. L., & Cooke, D. J. (2010). Is criminal behavior a central component of psychopathy? Conceptual directions for resolving the debate. *Psychological Assessment*, *22*, 433–445.

Spearman, C. (1927). *The abilities of man. Their nature and measurement*. London: Macmillan.

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

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

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*, 859–883.

Tsang, S., Salekin, R. T., Coffey, C. A., & Cox, J. (2018). A comparison of self-report measures of psychopathy among nonforensic samples using item response theory analyses. *Psychological Assessment, 30*, 311–327. https://doi.org/10.1037/pas0000481

Wakabayashi, A., Baron-Cohen, S., Wheelwright, S., Goldenfeld, N., Delaney, J., Fine, D., … Weil, L. (2006). Development of short forms of the Empathy Quotient (EQ-Short) and the Systemizing Quotient (SQ-Short). *Personality and Individual Differences, 41*, 929–940. https://doi.org/10.1016/j.paid.2006.03.017

Watts, A. L., Waldman, I. D., Smith, S. F., Poore, H. E., & Lilienfeld, S. O. (2017). The nature and correlates of the dark triad: The answers depend on the questions. *Journal of Abnormal Psychology, 126*, 951–968. https://doi.org/10.1037/abn0000296

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

Weigold, A., Weigold, I. K., & Russell, E. J. (2013). Examination of the equivalence of self-report survey-based paper-and-pencil and internet data collection methods. *Psychological Methods*, *18*, 53–70. https://doi.org/10.1037/a0031607

Wetzel, E., Lüdtke, O., Zettler, I., & Böhnke, J. R. (2016). The Stability of Extreme Response Style and Acquiescence Over 8 Years. *Assessment*, *23*, 279–291. https://doi.org/10.1177/1073191115583714

Yang, M., Jiang, G., & Yuan, K.-H. (2018). The Performance of Ten Modified Rescaled Statistics as the Number of Variables Increases. *Structural Equation Modeling*, *25*, 414–438. https://doi.org/10.1080/10705511.2017.1389612

Żemojtel-Piotrowska, M. A., Piotrowski, J. P., Cieciuch, J., Calogero, R. M., Van Hiel, A., Argentero, P., … Wills-Herrera, E. (2017). Measurement of Psychological Entitlement in 28 Countries. *European Journal of Psychological Assessment*, *33*, 207–217. https://doi.org/10.1027/1015-5759/a000286

Zettler, I., Moshagen, M., & Hilbig, B. E. (2019). Stability and Change: The Dark Factor of Personality Shapes Dark Traits. *Submitted for Publication*.

Tables

Table 1.

Number of Items by Dark Trait and Item Selection Stage

Dark Trait	Itempool	Initial Item Evaluation	First Item Selection	Final Item Sets		
				D70	D35	D16
Amoralism-Crudelia	13 (7)	12 (6)	10 (4)	7 (3)	4 (1)	3 (1)
Amoralism-Frustralia	14 (7)	12 (6)	9 (4)	6 (2)	3 (0)	1 (0)
Egoism	20 (10)	15 (8)	10 (5)	5 (2)	3 (1)	1 (1)
Greed	11 (6)	10 (5)	5 (2)	4 (1)	1 (1)	0
Machiavellianism	20 (10)	17 (8)	13 (7)	7 (4)	4 (2)	2 (1)
Moral Disengagement	16 (8)	13 (7)	10 (5)	5 (2)	1 (0)	1 (0)
Narcissism	18 (8)	14 (7)	10 (5)	6 (4)	3 (3)	0
Psychological Entitlement	12 (6)	12 (6)	9 (6)	5 (4)	2 (1)	0
Psychopathy	20 (10)	18 (8)	12 (4)	7 (3)	5 (2)	3 (1)
Sadism	20 (11)	20 (11)	15 (9)	8 (6)	5 (5)	3 (3)
Self-Centeredness	4 (2)	4 (2)	4 (2)	4 (2)	2 (1)	1 (1)
Spitefulness	16 (7)	15 (7)	11 (5)	6 (2)	2 (0)	1 (0)

Note. Number of negatively keyed items in parenthesis.

Table 2

Explained Common Variance by D

	ECV	
Items	Sample A3 (N = 77,218)	Samples B1-B4 (N = 6,838)
Amoralism-Crudelia	0.73	0.76
Amoralism-Frustralia	0.90	0.83
Egoism	0.71	1.00 [a]
Greed	0.53	0.57
Machiavellianism	0.90	1.00 [a]
Moral Disengagement	0.71	1.00 [a]
Narcissism	0.77	0.76
Psychological Entitlement	0.65	1.00 [a]
Psychopathy	0.92	1.00 [a]
Sadism	0.75	0.75
Self-Centeredness	0.78	0.83
Spitefulness	0.81	1.00 [a]

Note. *ECV*s in sample A3 are based on a bifactor model involving 118 items, whereas the *ECV*s in the pooled validation samples B1-B4 are based on a bifactor model only involving the items included in the D70.

[a] The associated specific factor exhibited no variance, implying an *ECV* of 1.00.

Table 3

Item Loadings on D

#	D70	D35	D16	#	D70	D35	D16	#	D70	D35	D16
1*	.60	.63	.65	25*	.49			49	.46		
2*	.66	.65	.62	26	.57			50	.68	.66	.64
3	.60			27*	.49			51*	.59	.57	
4*	.48	.47		28	.55	.57		52*	.55	.57	
5*	.48	.47		29	.60			53	.51	.50	
6	.57	.57	.56	30	.66	.66	.64	54*	.58	.60	.62
7*	.52			31*	.43			55	.49	.50	.47
8	.60			32	.61			56*	.55		
9*	.48			33*	.64			57	.54	.54	.54
10*	.47			34*	.58	.59	.61	58	.63		
11	.64	.63	.60	35*	.47			59	.47		
12*	.54			36	.58	.58		60	.59		
13	.50			37*	.59			61*	.39	.39	
14	.67	.65		38*	.60	.61	.61	62	.59		
15*	.51	.52		39	.44			63*	.51		
16	.68	.67		40*	.61	.62	.62	64	.69	.69	
17*	.69	.72	.75	41*	.54	.53		65*	.45		
18	.64	.63		42*	.67			66*	.63	.67	
19	.58	.60		43	.48			67	.61		
20	.60	.60		44	.63			68*	.65	.68	.70
21	.70	.69		45	.52			69*	.49		
22	.46			46*	.53			70*	.59	.61	
23	.55			47*	.47						
24	.57	.57	.56	48	.63	.64	.65				

Note. Starred items are negatively keyed. All loadings differ significantly from zero at $p <$ .05. Estimates based on the pooled validation samples B1-B4 (N = 6,838). See supplement for the verbatim items (https://osf.io/ag835).

Table 4

Internal Consistencies and Correlations to Criterion Measures

	Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	D70	(.97)													
2	D35	.98	(.95)												
3	D16	.95	.97	(.91)											
4	Aggression	.65	.65	.67	(.74)										
5	Cheating	.34	.35	.36	.02	-									
6	Crime/Delinquency	.32	.35	.37	.63	.09	(.62)								
7	Competitive Jungle	.92	.93	.91	.67	.19	.37	(.70)							
8	Dangerous World	.28	.26	.21	.30	.01	-.02	.34	(.79)						
9	Distrust	.51	.51	.49	.49	-.08	.15	.57	.66	(.85)					
10	Dominance	.59	.61	.59	.41	.16	.29	.73	.12	.28	(.71)				
11	Empathy	-.31	-.37	-.36	-.26	-.21	-.18	-.33	.04	-.20	-.21	(.89)			
12	Moral Identity	-.65	-.71	-.73	-.44	-.26	-.25	-.67	-.01	-.37	-.39	.42	(.81)		
13	Sadism-SVO	.22	.24	.23	-	-	-	-	-	-	-	-	-	(.93)	
14	SVO	-.43	-.45	-.46	-	-	-	-	-	-	-	-	-	.16	-

Note. Cronbach's alpha estimates of internal consistency on the diagonal. Correlations between the D70, D35, and D16 are based on the pooled validation samples B1-B4 (N = 6,838); all r differ significantly from zero at $p < .05$. Correlations between D70, D35, and D16 to 4-12 are based on sample B4 (N = 498); $|r| > .10$ ($|r| > .125$ concerning correlations involving the cheating task) differ significantly from zero at $p < .05$. Correlations between D70, D35, and D16 to 13-14 are based on sample B3 (N = 537); all r differ significantly from zero at $p < .05$.

Figures

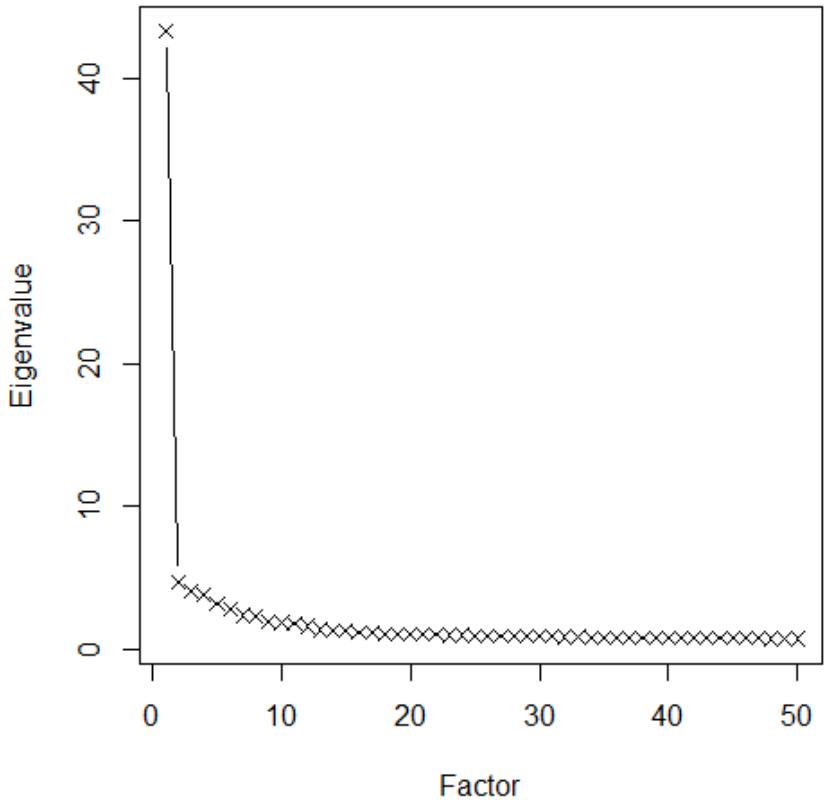

Figure 1. Screeplot (showing the first 50 factors) obtained during the initial item selection with 162 items and sample A2 (N = 74,477).