This is a preprint of an article in press in *European Journal of Personality*. Please cite as:

Hartung, J., Bader, M., Moshagen, M., & Wilhelm, O. (in press). Age and gender differences in socially aversive ("dark") personality traits. *European Journal of Personality.*

Age and Gender Differences in Socially Aversive ("Dark") Personality Traits

Johanna Hartung

Martina Bader

Morten Moshagen

Oliver Wilhelm

Department of Psychology, Ulm University

Author Note

Johanna Hartung, Institute of Psychology and Education, Ulm University, Germany, johanna.hartung@uni-ulm.de; Martina Bader, Institute of Psychology and Education, Ulm University, Germany, martina.bader@uni-ulm.de; Morten Moshagen, Institute of Psychology and Education, Ulm University, Germany, morten.moshagen@uni-ulm.de; Oliver Wilhelm, Institute of Psychology and Education, Ulm University, Germany, oliver.wilhelm@uni-ulm.de.

Correspondence concerning this article should be addressed to Johanna Hartung, Institute of Psychology and Education, Ulm University, Albert-Einstein-Allee 47, 89081 Ulm, Germany, Email: johanna.hartung@uni-ulm.de, Tel: (+49) 731 5031145

Abstract

The strong overlap of personality traits discussed under the label of "dark personality" (e.g., psychopathy, spitefulness, moral disengagement) endorses a common framework for socially aversive traits over and beyond the dark triad. Despite the rapidly growing research on socially aversive traits, there is a lack of studies addressing age-associated differences in these traits. In the present study (N = 12,501), we investigated the structure of the D Factor of Personality across age and gender using Local Structural Equation Modeling, thereby expressing the model parameters as a quasi-continuous, non-parametric function of age. Specifically, we evaluated loadings, reliabilities, factor (co-)variances, and means across 35 locally weighted age groups (from 20 to 54 years), separately for females and males. Results indicated that measurement models were highly stable, thereby supporting the conceptualization of the D factor independent of age and gender. Men exhibited uniformly higher latent means than females and all latent means decreased with increasing age. Overall, D and its themes were invariant across age and gender. Therefore, future studies can meaningfully pursue causes of mean differences across age and between genders.

Keywords: age and gender differences in personality; dark personality; Local structural equation models; age differentiation; D factor

Age and Gender Differences in Socially Aversive ("Dark") Personality Traits

Understanding which personality traits are associated with socially aversive behavior is an important research area. Personality traits that can have dysfunctional, potentially destructive behavioral manifestations and vary within nonclinical populations, often discussed under the label of "dark personality" (e.g., psychopathy, spitefulness, moral disengagement, narcissism, greed, etc.), are promising candidates to be studied in this context. Despite the rapidly growing research in this area, only a few studies investigated age differences (e.g., Ali & Chamorro-Premuzic, 2010; Götz et al., 2020). A further limitation is that existing studies often considered a limited age-range (e.g., Craker & March, 2016) and typically focused only on single traits (e.g., Grijalva et al., 2015), thus the literature is lacking studies on age and gender differences in multiple socially aversive traits. Importantly, beyond age-associated mean differences, potential structural differences of these traits across the life span (e.g., varying correlations between traits depending on the age of the sample) have rarely been studied so far (Carter et al., 2015; Klimstra et al., 2020), but are of paramount importance to contribute to a better understanding of socially aversive personality traits. Correspondingly, the present study explores gender- and age-related mean and structural differences in various socially aversive personality traits organized in the framework of the D-factor of personality (Moshagen et al., 2018) in a large sample of 12,501 individuals covering a wide age range from 21 to 53 years.

Socially-Aversive Personality Traits

Since the introduction of the so called "dark triad"—a group of three overlapping socially aversive traits (narcissism, Machiavellianism, and psychopathy)—in personality psychology (Paulhus & Williams, 2002), the research corpus dealing with the malevolent part of personality has exploded. Hundreds of research articles have been published and several meta-analyses have emerged from this large body of studies (e.g., Howard & Van Zandt, 2020; Muris et al., 2017; O'Boyle et al., 2015; Schreiber & Marcus, 2020; Vize et al., 2018)

as well as special issues devoted to tendencies towards ethically, morally, and socially questionable behavior (e.g., Veselka & Vernon, 2014). In recent years, personality research on subclinical socially aversive traits has been extended from the "dark triad" (Paulhus & Williams, 2002), over the "dark tetrad", additionally including sadism (Međedović & Petrović, 2015), to an overarching assembly of socially aversive personality traits in the framework of the D Factor of Personality (Moshagen et al., 2018).

Aside from the number of socially aversive traits, their nature – most importantly their convergent and discriminant validity – has been controversial. There are theoretical considerations (e.g., Diebels et al., 2018; Marcus & Zeigler-Hill, 2015) as well as empirical findings about the considerable overlap among said traits (e.g., Book et al., 2016; Jones & Figueredo, 2013). Both theory and data suggest that the high overlap between the members of the "dark triad" (Bertl et al., 2017; Muris et al., 2017; O'Boyle et al., 2015) and also between all socially aversive personality traits in general is mostly due to a general tendency towards ethically, morally, and/or socially questionable behavior. Previous notions on the commonalities of a subset of dark traits (e.g., Diebels et al, 2018; Hodson et al., 2018; Jones & Figueredo, 2013) have recently been unified and expanded by introducing a common disposition underlying all socially aversive personality traits (and thus representing their commonalities), integrating key features from definitions of all socially aversive traits. The *D Factor of Personality* 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" (Moshagen et al., 2018, p. 657). This conceptualization has been empirically tested in a series of studies that investigated the factor structure and predictive validity of a common D-factor and specific factors of socially aversive personality traits (Hilbig et al., 2020; Moshagen et al., 2018; Moshagen, Zettler, & Hilbig, 2020; Moshagen, Zettler, Horsten, et al., 2020). D is defined in analogy to the *g* factor of intelligence as "a general dispositional tendency underlying all dark traits". According to

this definition, specific socially aversive traits represent flavored manifestations of D and differ in terms of which defining aspects of D are most important, as well as in terms of the unique aspects that they carry beyond D. The way that the measurement instrument of D was developed follows these theoretical considerations. A large number of instruments, which convey similar measurement intentions, namely measuring socially aversive or so called "dark" traits, were collected and items best representing a common core were selected using meta-heuristics (Moshagen, Zettler, & Hilbig, 2020). D is thus a parsimonious description of self-reported individual differences that can help to find sufficient explanations of differences in criterion measures of interest (Bader et al., 2020; Hilbig et al., 2020; Moshagen, Zettler, & Hilbig, 2020).

Previous studies (Moshagen et al., 2018; Moshagen, Zettler, & Hilbig, 2020) modeled D as a general factor arising in a bifactor model along with specific factors representing the unique aspects of each specific socially aversive trait (such as psychopathy or spitefulness). However, deriving specific factors based on item-scale assignments in specific instruments is not likely to deliver sustainable specific factors (Međedović & Petrović, 2015; Volmer, Koch, & Wolff, 2019), in particular in light of the debated distinctiveness of some socially aversive traits (e.g., Miller et al., 2017). To determine a superior representation of the structure of D, Bader et al. (2020; see https://osf.io/bc4et/) used a slightly modified "bass-ackwards" approach (Goldberg, 2006) to identify themes that reflect reliable and theoretically sound variance beyond a very strong general factor that supports the idea of a common core of all socially aversive traits. The thereby exploratively identified themes were inductively labeled callousness (lack of concern for others), deceitfulness (tendency to deceive), narcissistic entitlement (sense of self-importance and inflated deservingness), sadism (tendency to inflict harm for personal gain), and vindictiveness (tendency to seek retaliation). Note that these themes immediately map on the theoretical definition of D as the tendency to maximize one's individual utility while disregarding (callousness), accepting (deceitfulness), or malevolently

provoking (sadism, vindictiveness) disutility for others, which is accompanied by justifying beliefs (narcissistic entitlement). Hence, this mapping between empirically derived themes and theoretical definition could be interpreted as evidence for the successful item selection.

Studies locating D in the models of basic personality (Moshagen et al., 2018; Moshagen, Zettler, & Hilbig, 2020) indicated—besides a substantial (negative) correlation to conscientiousness—large correlations with honesty-humility from the HEXACO model and agreeableness (from both the HEXACO model and the Five Factor Model; FFM). This is rather unsurprising given that the theoretical conceptualization of D also implies features thought to reflect honesty-humility or agreeableness (for details on theoretical similarities and differences, see Moshagen et al., 2018; Moshagen, Zettler, Horsten, et al., 2020) and given that item contents used to measure these constructs are partly similar (see Vize et al., 2020). Notably, recent meta-analytic evidence on the "dark triad" traits indicates that their communalities neither reduce to honesty-humility or agreeableness, respectively, nor that honesty-humility or agreeableness can be perfectly predicted by the communalities of the dark triad traits (Schreiber & Marcus, 2020), which is mirrored by meta-analytic evidence indicating different nomological nets (Howard & van Zandt, 2020). By extension, it seems likely that the same also holds when considering D (representing not only the commonalities of the dark triad components, but the basic disposition underlying all socially aversive traits), as hinted by evidence suggesting that D incrementally predicts a host of criterion measures (including behavioral outcomes) over both honesty-humility (Hilbig et al., 2020; Moshagen et al., 2018; Zettler et al., 2020) and agreeableness (Moshagen et al., 2018; Moshagen, Zettler, Horsten, et al., 2020) . Nevertheless, it is clear that among the dimensions of basic personality, D is best approximated by honesty-humility or FFM-agreeableness.

Age-Associated Mean Trends in Socially Aversive Personality

In general, personality development is a complex issue with biological, genetic and environmental factors interacting (Specht et al., 2014; Tucker-Drob et al., 2019). Age-

associated differences concerning mean levels of certain personality traits are often thought to reflect maturational processes shared by a population. The maturity principle (Roberts et al., 2006) states that an increase in adjustment (Staudinger & Kessler, 2009) and age-associated improvements in emotion regulation (Carstensen et al., 2003) result in individuals becoming more agreeable, conscientious, and emotionally stable with increasing age (Caspi et al., 2005).

The development of socially aversive personality traits is often explained resorting to life-history theory (Ellis et al., 2009), which states that individual differences in life-history strategies are influenced by the environment people find themselves in, with unpredictable environments leading to "fast" strategies (e.g., early reproduction, impulsivity, aggression, etc.). For example, the job environment might become more predictable with increasing age based on tenure positions, better social networks, or policies like "who came last has to go first when economic problems arise". Similarly, the relationship environment might appear less predictable for singles than for someone in a long-term relationship.

In a similar vein, one factor that is believed to drive personality development are changes in social roles. During adolescence and early adulthood, individuals attempt to establish themselves in the workforce (negotiate salary, positions, and demands, as well as to prevail against peers) and begin romantic relationships. A certain level of competitive or antagonistic behavior might prove beneficial (Spurk et al., 2016) and might be viewed as more desirable at this stage of life (Berenson et al., 2017). However, being responsible, reliable, and socially integrated becomes more important with increasing age (Roberts & Wood, 2006) or, in other words, transgressive and norm-violating behavior could have negative consequences at the workplace (O'Boyle et al., 2012; but see Wille et al., 2019) as well as in romantic relationships (Ali & Chamorro-Premuzic, 2010; Jonason et al., 2012).

In sum, theories stating that the change in social roles and beneficial or disadvantageous effects of socially aversive personality traits drive personality development give a point to a decrease in mean levels of these personality traits. This also mirrors

biological evidence suggesting that endocrinological changes could inform personality differences. Testosterone levels – which have been found to be associated with a variety of malevolent behaviors (Book et al., 2001; Eisenegger et al., 2011; Mazur & Booth, 1998) – decrease across the adult age range (Fabbri et al., 2016; Handelsman et al., 2016). Lower testosterone levels in older adults could form the biological basis for the lower tendency for aggressive, antisocial behavior.

Empirically, age-associated differences in personality are mostly investigated by (cross-sectional) comparisons of means in different age-groups or by correlations of personality characteristics with age. Concerning basic dimensions of personality, most studies report an increase in (mean-level) FFM-agreeableness with age, although the slopes are not very steep (Anusic et al., 2012; Klimstra et al., 2020; Soto et al., 2011; Srivastava et al., 2003). In the HEXACO framework, a moderate increase in honesty-humility across the adult age range is evident (for a meta-analysis, see Moshagen, Thielmann, et al., 2019). Given that these dimensions of basic personality provide the closest proxy of socially aversive traits, a similar (inverse) association of the latter with age can be expected. Indeed, research has repeatedly indicated that age typically shows small to moderate negative correlations with all "dark triad" components (Ali & Chamorro-Premuzic, 2010; Barlett & Barlett, 2015; Craker & March, 2016; Fox & Rooney, 2015; Götz et al., 2020; Kowalski et al., 2017; Spurk et al., 2016).

Longitudinal evidence largely corroborates the cross-sectional results summarized above. Regarding dimensions of basic personality, agreeableness increases slightly with increasing age (Roberts et al., 2006; Terracciano et al., 2005), but results between studies are mixed, partly depending on the operationalization employed and the age range considered (Graham et al., 2020). The developmental pattern of honesty-humility follows a pronounced and consistent increase across the adult age range (Milojev & Sibley, 2017).

Regarding longitudinal evidence involving socially aversive traits in particular, the

majority of existing studies either focused on narcissism, relied on data on (early) adolescence or clinical samples, or realized only short intervals between measurements (e.g., Frick et al., 2003; Greitemeyer & Sagioglou, 2017; Sijtsema et al., 2019). The few findings regarding traits different from narcissism in community samples essentially point to a decreasing trend in socially aversive personality traits (e.g., Blonigen et al., 2006; Grosz et al., 2019; Ingo Zettler et al., 2020), including a mean-level decrease of about a third of a standard deviation of the D factor over a 4-year period (Zettler et al., 2020). Longitudinal studies on narcissism show that individual age trends vary a lot (e.g., Wille et al., 2019). Findings on mean level changes range from significant increase to significant decrease depending on the age range and the facets of narcissism that were studied (Carlson & Gjerde, 2009; Chopik & Grimm, 2019; Grosz et al., 2019; Wetzel et al., 2019; Wille et al., 2019). More specifically, narcissistic admiration was found to be stable in young adults (Grosz et al., 2019), individual autonomy was found to increase from adolescence to old age (Chopik & Grimm, 2019), maladaptive forms of narcissism (e.g., hypersensitivity, willfulness) declined from adolescence to old age (Chopik & Grimm, 2019), and leadership, vanity, and entitlement decreased from young adulthood to midlife (Wetzel et al., 2019). Interestingly, using an observer-based methodology, no significant changes in overall narcissism were found in young adults (Carlson & Gjerde, 2009) and over a 23-year interval mean levels of narcissism decreased while on an individual level the majority of participants did not show reliable change (Wetzel et al., 2019). These differences underscore one again the importance of interpreting results in the context of the statistical methods and the measurement.

Most studies test for a linear relationship of personality trait levels and age (e.g., correlations) are likely bound by the common statistical methods used. However, the possibility of non-linear relationships should not be disregarded. In light of findings from personality development, it appears possible that there is a steeper downward trend in early adulthood compared to old age (Schwaba & Bleidorn, 2018). In older age, increasing health

problems might constitute a less predictable environment and trying to increase one's own benefits gets more socially accepted and beneficial again, which could lead to an reverse maturation process (Kornadt et al., 2018).

In sum, empirical evidence is mostly in line with the maturity principle suggesting a decrease in socially aversive personality traits. However, results concerning specific socially aversive traits are rather heterogeneous and mostly limited to only few traits. Importantly, there is hardly any evidence for structural stability although being the foundation of interpretable age difference comparisons, which will be discussed in the following section.

Age-Associated Differences in the Factor Structure of Socially Aversive Personality

Comparing mean level differences of psychological constructs across age necessitates a stable structure of the constructs under consideration in the first place, because mean trends are not interpretable if structural differences occur. Specifically, meaningful group comparisons can only be made when strong invariance holds (Meredith, 1993). However, age-associated personality differences might also manifest in factor configurations rather than (or in addition to) factor means (Olaru et al., 2018). Ignoring possible differences on the structural level jeopardizes the interpretability of mean trends (Borsboom, 2006).

Indeed, whereas socially aversive trait measures were mostly developed based on rather young samples (Jonason & Webster, 2010; Jones & Paulhus, 2014), many studies investigating the prediction of outcome variables were conducted in adult samples (e.g., Craker & March, 2016; Fox & Rooney, 2015). These studies assume that the scales measure the same constructs independently of age so that mean scores can be compared independently of participants' age. However, measurement invariance across age as a prerequisite of mean value comparisons has not been tested for most of these scales and, to the present day, it is not clear if the manifestations of D are the same across the whole life span.

To the best of our knowledge, there are only three studies so far that examined

structural age-related differences in *multiple* socially aversive traits.[1] In a 4-year longitudinal study involving the D factor and nine socially aversive traits (including the "dark tetrad" components), Zettler et al. (2020) provided evidence for both strict longitudinal measurement invariance of the traits under scrutiny and for structural invariance concerning the intercorrelations between the traits across time, thus suggesting remarkable structural stability over a 4-year period. However, Zettler et al. (2020) merely assumed the same structure holds across the full age range and did not investigate possible structural differences depending on the age of the participants.

Using the Dirty Dozen (Jonason & Webster, 2010) as a measure of the dark triad components, Carter et al. (2015) found a varying number of factors depending on gender and age of the subsample when modelled with an IRT-based analysis. However, this study has some limitations: 1) the Dirty Dozen is a questionable measure of "dark" personality because of its validity issues (Jones & Paulhus, 2014; Maples et al., 2014; Miller et al., 2012), 2) the age range of the two samples overlapped, and 3) the two samples were recruited differently, which makes it likely that these samples differ in other aspects as well. It is therefore very possible that the differences in the "dark triad" structure were driven by other factors beyond age. Nevertheless, this study, as well as a related study by (Spurk & Hirschi, 2018), indicates that the assumption that socially aversive personality measures are invariant or that manifestations of these traits are the same across the age range might be false, which could result in misleading interpretations of age trends in socially aversive personality traits.

A study that is conceptually similar to the present study used a cross-sectional dataset of studies using the Dirty Dozen (Klimstra et al., 2020). The one-factor model fitted the data better in younger age groups than in older age groups, thus pointing to a differentiation of the structure of the Dirty Dozen. The results regarding measurement invariance revealed at least

[1] Please note that there are studies investigating measurement invariance of instruments measuring a single socially aversive trait, e.g. narcissism (Berenson et al., 2017), or basic prosocial personality traits, i.e. agreeableness (e.g., Lucas & Donnellan, 2011; Olaru et al., 2018).

partial scalar invariance between adjacent age groups, but often no scalar invariance between non-adjacent age groups. These results indicate that the Dirty Dozen is not invariant across the whole age range, but that invariance might be a question of the magnitude of age differences. They found lower levels of manipulation, egocentricity, callous affect, and a general factor in late adulthood age groups compared to the young adulthood age group. Although this study has several strengths, there are also some shortcomings: 1) like the study described above, it is based on the Dirty Dozen with only twelve items, 2) age was not treated as continuous variable, and 3) correlations between the "dark personality features" were not investigated. The latter information could be helpful to understand the age-related structural differences of the Dirty Dozen.

In summary, these studies indicate that the factor structure of socially aversive personality traits might differ across age. However, the small number of studies testing structural age differences relied on an instrument of questionable quality and categorization of age. Although the Dirty Dozen results could be interpreted as *pars pro toto*, it is arguably not a very good representation of the nomothetically broader D factor. It is plausible to assume that differentiation in personality may only be evident when considering a broad item set, which actually allows to observe potential structural differences, rather than relying on only a few and arguably rather similar items that might obscure structural differences. In contrast to this sparse, plagued by methodologically shortcomings previous research (Carter et al., 2015; Klimstra et al., 2020), ensuring measurement equivalency is crucial to avoid capitalizing on methodological artifacts. Clearly, if the factor structure of personality traits differs as functions of age, this would not only render mean comparisons meaningless, but also implies differentiated relations with other personality traits or criterion measures. To close this gap in research, we studied the influence of age on the structure of socially aversive personality traits within an integrative framework.

The Present Investigation

It is noteworthy that age-associated differences in socially aversive personality traits, in particular concerning possible structural differences, have received little attention in the literature so far. In order to shed light on potential age-associated differences in the structure and mean levels of socially aversive personality traits, the present study relied on a large sample of participants covering a wide age range and employed an innovative method for estimating continuous age moderation of factor model parameters. Local structural equation modeling (LSEM; Hildebrandt et al., 2009) allows for investigating structural differences in the measurement model and trends in factor means across age as a continuous variable. Hence, the current investigation attempts to overcome the limitation of previous studies that treated age as a discrete variable (see Wagner et al., 2019 for a discussion on how to model age-related changes) or did not investigate measurement invariance at all and simply assumed structural equivalence (e.g., Craker & March, 2016; Fox & Rooney, 2015; Kowalski et al., 2017). The nonparametric attributes of the LSEM analysis are especially beneficial as the shape of the moderation effects is unknown, considering that age-related trends may turn out non-linear. In addition, as hinted by robust gender differences in socially aversive personality (Muris et al., 2017; Paulhus & Williams, 2002; Wilson & Sibley, 2011) in conjunction with some studies indicating interactive effects of gender and age (Carter et al., 2015; James et al., 2014), we conducted all analyses separately for female and male subsamples to address potential differences in age-related trajectories in women and men. Further going beyond prior research, we rely on an integrative model with items sampled much broader as compared to the “dark triad” components to evaluate whether and how structural differences in socially aversive personality characteristics occur as function of age and gender. In summary, the current study moved beyond previous studies on age and gender differences in socially aversive traits, namely by our ability to investigate factor structure across continuously measured age in a sophisticated analysis based on a large sample that answered items with a broad content coverage.

Concerning possible age-related differentiation of the structure of socially aversive personality characteristics, we hypothesized that intercorrelations between the themes of socially aversive personality and second-order loadings, respectively, decrease with increasing age, in line with the results presented in Klimstra et al. (2020). In early adulthood, the basic dispositional tendency to maximize one's individual utility at the expense of others might be rather unspecialized in different domains (e.g., without a specific pattern in the type of behavior). By contrast, as individuals grow older, individual utility maximization at the expense of others might become more domain-specific (e.g., limited to the workplace) or they might develop a preferred type of malevolent behavior (e.g., hurting other people to maximize their own utility). This specialization or differentiation of socially aversive personality traits should be reflected in decreasing intercorrelations between the themes and second-order loadings, respectively. With respect to mean trends, in line with the maturity principle (Roberts et al., 2006) and the age-related decrease in testosterone levels (Fabbri et al., 2016; Handelsman et al., 2016), we expected that latent means decrease with increasing age.

Methods

Sample

Analyses were based on data from 12,501 participants native or fluent in English collected online between February and July 2019 via darkfactor.org. Parts of the data have been used in Bader et al. (2020) and Moshagen, Zettler, and Hilbig (2020). The dataset used for the current analysis is accessible at OSF (https://osf.io/4hgub/). The study was not preregistered.

The website was advertised via reports and interviews in the international media (see darkfactor.org for links). Participants could choose to complete short versions of the questionnaire (16 and 35 items) or the complete 70-item version (D70; see below). For the present analyses, only participants completing the D70 were considered. They received feedback on their D score as incentive.

To remove potentially biased observations, further exclusion criteria were used: records of identical IP addresses in combination with identical age and gender (potentially resulting from repeated participation of the same subject), suspicious response behavior (i.e., requiring less than 2 seconds per item), and participants refusing to disclose information on age and gender. Additionally, two participants were excluded due to missing data on the D70 items, which occurred due to technical issues. Another 170 participants stated "other" as their gender, yielding a group-size too small to support an age-related analysis. Therefore, only female and male participants will be compared.

Participants originated from more than 100 countries. About 41% of the sample stated an English-speaking country as their country of residence. The three most frequent countries of origin were USA (29%), Norway (15%), and Poland (9%).

The overall sample was divided into a female and a male sample (n_{female} =5,982 and n_{male} = 6,519). Both samples comprised adults between 18 and 80 years old. Mean age was 30.48 years (*SD* = 11.33) and 30.24 years (*SD* = 11.37) in the female and male sample, respectively.

Measures

D and its themes were measured with the D70 (Moshagen, Zettler, & Hilbig, 2020). The questionnaire was derived as an efficient yet valid and inclusive measure of D using a rational-item selection approach (Schroeders et al., 2016). Specifically, a genetic algorithm (Holland, 1992) was applied to select 70 items with high loadings on D from a large item pool covering a broad set of 12 different socially aversive traits, thereby ensuring psychometric soundness and maintaining nomological breadth. Participants indicated their agreement with each item on a 5-point scale, ranging from *strongly disagree* (1) to *strongly agree* (5). The items and their allocation to the five themes are available at an OSF repository (https://osf.io/4hgub/).

Model-based reliabilities (Brunner et al., 2012) for the five themes are indicated by McDonald's (2013) omega, expressing the proportion of variance accounted for by the latent variable relative to observed score variance. Omega is based on the standardized loadings which take the standard deviation of the factor and the standard deviation of the indicator into account ($\lambda_{standardized} = (\lambda_{unstandardized} * SD_{factor})/SD_{indicator}$). Omega for the themes was .91 for callousness (15 items), .85 for deceitfulness (9 items), .86 for narcissistic entitlement (11 items), .93 for sadism (18 items), and .92 for vindictiveness (17 items). Descriptive statistics for the latent variables can be found in Table 1 and descriptive statistics for the items can be found in Table A.1 in the Appendix.

Analyses

Measurement Model

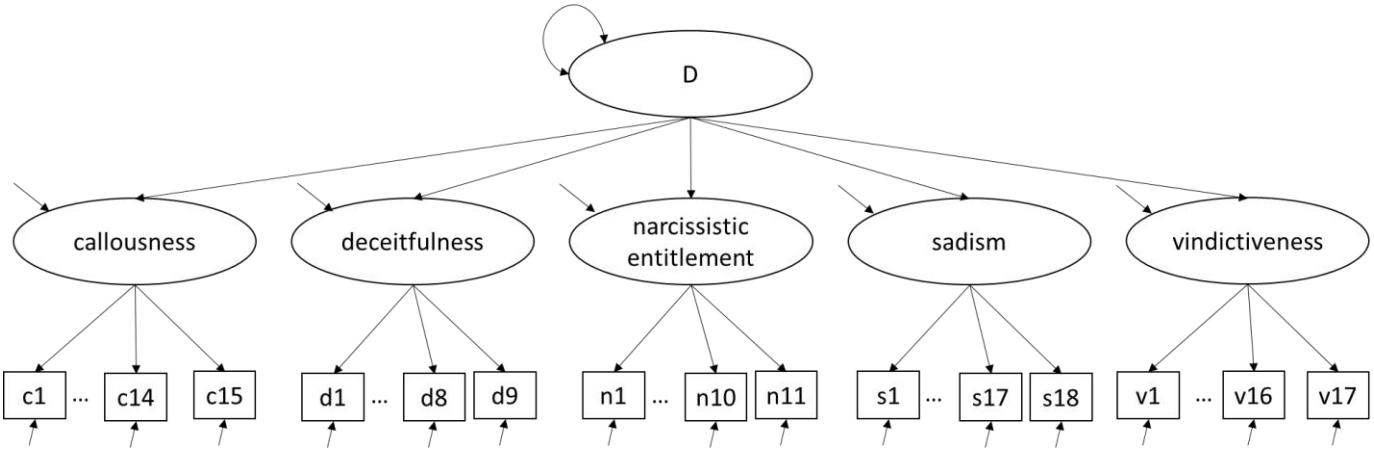
We specified a higher-order model assuming a primary factor for each of the themes according to Bader et al. (2020) and a second-order common D factor (see Figure 1).[2] The likert-type items were treated as continuous because MLR produces better estimates of interfactor correlations (Li, 2016), the rescaling reduces bias effectively (DiStefano, 2002), and in the context of multiple-group CFA does not lead to an inflated Type I error rate (Koh & Zumbo, 2008). Factors were identified using the effects-coding method (Little et al., 2006). That is, factor loadings of the same factor were constrained to be 1 on average, and item intercepts were constrained to be 0 on average. By using this identification method, the latent variances reflect the average of the variances of the indicators accounted for by the construct, and the latent means are the optimally weighted averages of the set of indicator means for the respective construct. In other words, factor means are estimated as the average of indicator means weighted by their respective factor loading and factor variances represent the average of the variance of the indicators explained by the latent factor. In the higher-order model,

[2] Note that higher-order and bifactor models yield a general factor which is essentially equivalent, see Brunner et al. (2012) and Moshagen et al (2020) for a discussion.

however, factor means of the first-order factors are also constrained to be 0 on average and are hence not easy to interpret. This is why we report factor means of the first-order factors as estimated in a correlated factor model. Results from both models were highly similar. Detailed results from the correlated factor model can be found on the OSF repository (https://osf.io/4hgub/).

Figure 1

Second-order model for five themes and the common D factor



Note. D captures the shared variance between the five themes. To enhance clarity, only a subset of three indicators per theme (first-order factor) is displayed.

Maximum likelihood estimation with robust standard errors (Huber-White) and scaled test-statistics (Yuan-Bentler) was used to account for sampling weights and the effects of non-normality (Yuan & Bentler, 2000). In line with recent recommendations, we considered the point estimate values for the Standardized Root Mean Square Residual (*SRMR*) and the robust Root Mean Squared Error of Approximation (*RMSEA*) around .08 as indicative of acceptable model fit (Browne & Cudeck, 1992; Moshagen & Auerswald, 2018). The Comparative Fit Index (*CFI*) was not used as an indicator of pointwise model fit as it is more dependent on loading magnitude than on discrepancy from the null model. Therefore, strict cutoffs based on the *CFI* could be misleading. The *CFI* is, however, useful for model comparisons, as all

models are compared with respect to the same reference (Moshagen & Auerswald, 2018). Indeed, simulation studies demonstrated the usefulness of the *CFI* for invariance testing (Chen, 2007). We thus considered differences between models with increasing invariance constraints as meaningful when $\Delta CFI \geq -.01$ and $\Delta RMSEA \geq .015$ (Chen, 2007; Cheung & Rensvold, 2002).

All analyses were conducted using R 3.5.1 (R Development Core Team, 2001). The main R package used for the LSEM analyses was *sirt 3.9-4* (Robitzsch, 2019). All R scripts including R code to produce the figures are available on the OSF repository.

Multiple Group Analysis (MGCFA) and Local Structural Equation Modeling (LSEM)

In order to test for a main effect of gender, multiple-group confirmatory factor analysis (MGCFA; Vandenberg & Lance, 2000) with increasing levels of measurement invariance across gender were tested (Byrne & Stewart, 2006; Chen et al., 2005; Meredith, 1993). First, the model structure was estimated simultaneously in both groups (configural variance), then first-order factor loadings were constrained to be equal in both groups (metric invariance), followed by additionally restricting second-order factor loadings to be equal across groups. Next, indicator intercepts were additionally equated across groups (scalar invariance). Then, residual variances of the first-order factors were constrained to be equal across groups. This step was followed by additionally restricting the residual variances of the indicators to be equal across groups (strict invariance).

To investigate age-related structural differences, we resorted to LSEM. LSEM is an approach for modeling parameters of a structural equation model as a continuous, non-parametric function of a moderator (in this case, age; Hildebrandt et al., 2009; Olaru et al., 2019), so that all model parameters can be investigated without *a priori* specifying their relation with the context variable. The main advantage of LSEM is the ability to avoid the negative consequences of artificially categorizing a continuous context variable (Hildebrandt et al., 2009; MacCallum et al., 2002). Because of the nonparametric character of LSEM,

nonlinear trends and onset of transformation can be detected by visualizing transformations in model parameters (Olaru et al., 2019). Therefore, LSEM is particularly well suited to investigate questions about differences in model parameters including patterns of age-associated differences.

LSEM estimates the measurement model at each specified value of the age variable, called focal points, with observations weighted as a function of their proximity to a focal age point. Based on recommendations of Hildebrandt et al. (2009), a Gaussian kernel function with a bandwidth factor of $h = 2$ was used to weight estimations, and focal age points ranging from 20 to 54 years with increments of 1 year were defined. These upper and lower limits were selected to reduce boundary bias based on the age distribution of the sample (Hartung et al., 2018; Hildebrandt et al., 2016). In total, we estimated 35 weighted models for each the female and the male subsamples. Due to the weighting of the participants at every focal point, the mean context variable value does not exactly match the focal point value. This shift in mean values of the weighted samples especially affects the ranges of the context value distribution. The weighted mean age at every focal point is referred to as effective age (see also Hartung et al., 2018). The effective sample size at each focal point of the context variable depends on the weighting and the context variable distribution because observations near the focal points are partially included. Note that the LSEM results need to be interpreted descriptively, because no direct inference statistical test of model parameter comparisons was performed.

We expanded the investigation of factor means by using a Multiple Indicator Multiple Causes (MIMIC) model for methodological cross-validation. The MIMIC model analysis was conducted by regressing the latent mean of D or the five themes on age and gender. The interaction of age and gender was calculated by including the product of centered age and effect-coded gender with females as reference group as additional predictor. Although LSEM is superior to MIMIC models because every model parameter is dependent on age and it does

not assume a linear relationship, a methodological cross-validation is useful to determine whether the pattern of results was unduly influenced by the method applied.

Results

Multiple-Group CFA to Investigate Gender Differences

First, we tested measurement invariance of D and the five themes across gender (Table 2). The configural invariance model yielded an adequate fit to the data. At every level of invariance, constraining the respective parameters to be equal across gender led to a significant decrease in model fit, but both ΔCFI and $\Delta RMSEA$ suggested that the differences in model fit were not meaningful. Consequently, strict invariance across gender was supported.[3] Standardized second-order loadings of the five themes on D were very high, ranging from $\lambda = .84$ to $\lambda = .95$ (mean $\lambda = .92$), thus supporting the notion of a strong single underlying factor (D). The standardized latent mean difference for D was .42, indicating higher scores for males amounting to about half of a standard deviation.

In sum, these results confirm (a) gender-invariant measurement models, (b) a strong common factor, and (c) gender differences on the latent mean, with men exhibiting higher scores.

LSEM with Effects Coding to Investigate Age-Associated Differences

We first report sample attributes for the weighted samples and model fit. To evaluate our first hypothesis concerning age-associated difference of the factor structure, we consider loadings, reliabilities, and first-order factor variances. Finally, we investigate latent mean differences to test the hypothesis of decreasing factor means with increasing age.

Effective Sample Size and Model Fit

The average age of the weighted samples (effective age) ranged from 21.5 to 53 years. The effective sample size in the female (male) sample ranged from $n = 328$ ($n = 306$) at focal

[3] In the correlated factors model, strict invariance was supported as well and factor variances and covariances could be restricted to be equal across gender.

age point 54 to *n* = 2,693 (*n* = 2,988) at focal age 23 (see Figure A.1**Fehler! Verweisquelle konnte nicht gefunden werden.**). Thus, the weighted sample sizes were sufficiently large to support model estimation.

The robust *RMSEA* in the female and male samples ranged from .049 to .063 and .050 to .069, respectively, with increasing values with increasing age. Model fit across the age range was thus acceptable (see Appendix Figure A.2).

Factor Loadings and Reliabilities

Using the effects-coding method, loadings are estimated freely at each focal point, resulting in the optimal weighting at each focal point. Note that this identification method requires that unstandardized loadings are constrained to an average of one at each focal point for each factor. Therefore, age trends in unstandardized loadings should be interpreted in the context of the scaling procedure.

Most first-order loadings were highly similar across age (see Appendix Figure A.3). Similarly, the representativeness of the indicator – as assessed by the rank order and magnitude of loadings – for the latent factors were similar across gender groups. For example, the indicator d30 “Why should I care about other people, when no one cares about me?” had the highest loading on callousness across the whole age range in the female as well as in the male sample (note that starting in the mid-30s for females or beginning-40s for males, 95% confidence intervals overlap with the ones of other items). One of the few exceptions was the unstandardized loading of item d37 “There is poor comfort in revenge.”, which increased in the female sample rather steeply until an effective age of about 43 from .9 [.85; .96] to 1.3 [1.20; 1.37] and remained quite stable afterwards. In the male sample, however, it was rather stable across the age range, with estimates between about 1 [.94; 1.02] and 1.1 [1.08; 1.23].

The second-order loadings of the themes were also rather stable across age in both gender groups (see Figure A.4), thereby disconfirming our first hypothesis. The largest difference across the age range in the female sample was evident for the second-order loading

of vindictiveness that increased, with estimates between 1.08 [1.06; 1.10] and 1.21 [1.16; 2.25]. In the male sample, all second-order loadings appeared to be rather stable and all loadings were of similar height except the loading of narcissistic entitlement, which was lower than the others. However, in general, second-order loadings of the five themes were uniformly strong and very stable across the whole age range.[4]

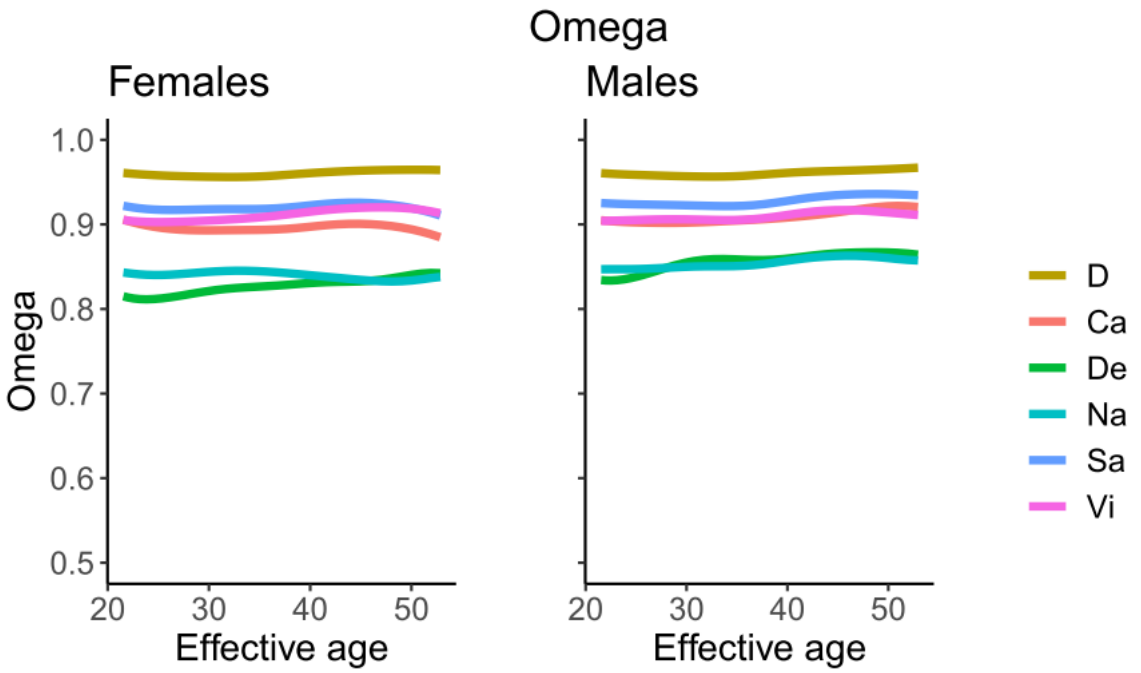
Omega estimates of factor reliability for the D-factor and all five themes were good to very good across the whole age range (ranging from .81 to .97). Furthermore, omegas proved to be highly stable across age and gender, indicating that D and the five themes were measured with similar accuracy across groups (see Figure 2).

In sum, there was no evidence for substantial differences in loadings or factor reliabilities as function of age and gender, thereby indicating that the measurement models were quite stable across the considered groups and, thus, that D and the themes have a similar meaning in all groups.

[4] Factor correlations in the correlated factors model were also very stable across the whole age range.

Figure 2

Age Gradients of Omegas Estimated with Local Structural Equation Modeling



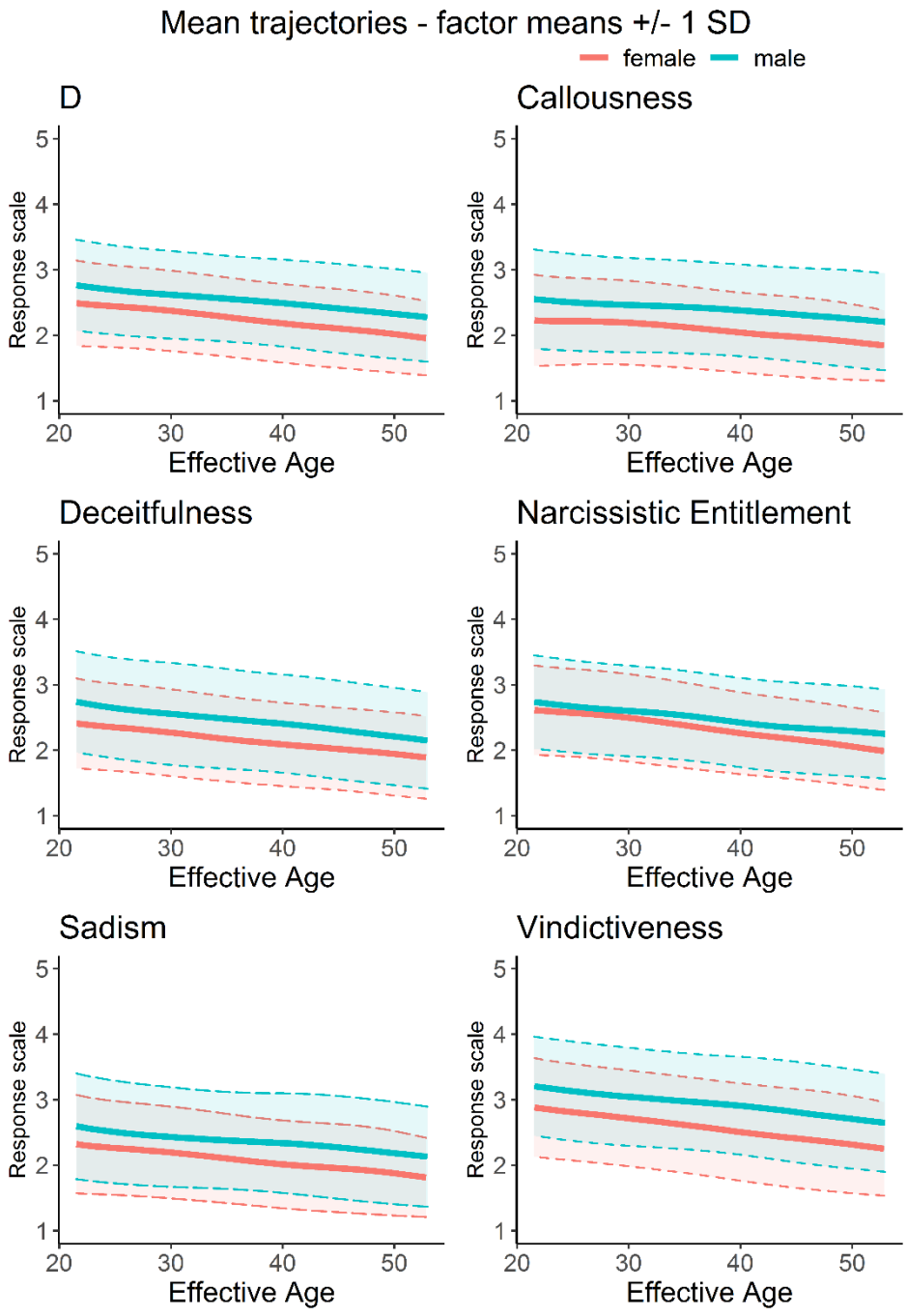
Note. Ca = callousness, De = deceitfulness, Na = narcissistic entitlement, Sa = sadism, Vi = vindictiveness.

Factor Variances and Means

Factor mean and variance of the D factor were estimated in the higher-order model. In order to obtain factor means in the metric of the scale, the factor means and variances for the five themes were estimated in a correlated factors model. Factor variances remained rather stable across age groups (Figure 3). Trajectories of the factor means in the metric of the scale are illustrated in Figure 3 (unstandardized and standardized factor means with confidence intervals can be found in the Appendix Figure A.5 and A.6). In line with our second hypothesis, all latent means decreased with increasing age. Interestingly, despite substantial mean differences, the age-related trajectories were quite similar for both gender groups.

Figure 3

Age Gradients for the Latent Means of D and the Themes Estimated with Local Structural Equation Modeling

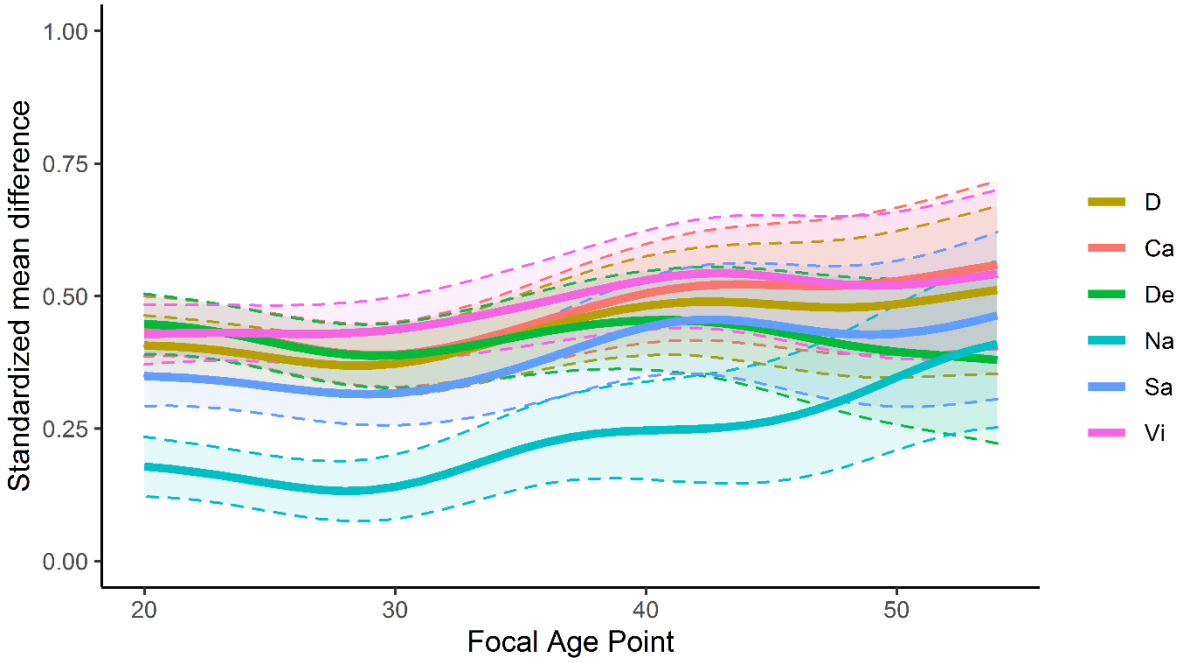


Note. Estimates are in the metric of the response scales (1 = strongly disagree, 5 = strongly agree). Latent means of the themes were estimated in a correlated factors model; latent means of D were estimated in a higher-order factor model. Solid lines represent the estimated latent mean; dashed lines depict an interval of +/- 1 standard deviations (square root of estimated factor variances at the respective focal point).

To investigate main effects of gender and interaction effects of gender and age in greater detail, we determined standardized latent mean differences for each theme and D at each focal point.[5] As can be seen in Figure 4, men exhibited higher latent means across the whole age range. Whereas gender differences in deceitfulness were stable, differences in D, callousness, vindictiveness, and sadism showed a slight increase with age. However, confidence intervals of the minimum and maximum value overlapped for all of them. The most pronounced age-related difference concerning gender difference was found for narcissistic entitlement (ranging from .13 [.08; .19] at age 27 to .40 [.25; .57] at age 52).

Figure 4

Standardized Latent Mean Differences on D and the Themes Between Gender as a Function of Age Based on the Estimates from Local Structural Equation Modeling



Note. Latent means for D were estimated in a higher-order model. Latent means for the themes were estimated in a correlated factors model. Positive values indicate higher values for males. Solid lines represent the estimated factor loadings from CFA. Dotted lines represent the 95% confidence intervals. Ca = callousness, De = deceitfulness, Na = narcissistic entitlement, Sa = sadism, Vi = vindictiveness.

[5] An additional analysis, in which LSEM was based on a MIMIC model including gender as manifest variable, yielded the same shape of mean differences (see online supplement).

Cross-Validation of Mean Trends with MIMIC Models

Table 3 shows the results from the MIMIC analysis using the overall sample and the gender subsamples. In line with our expectations, regression weights for age on D and the five themes were all negative and in a similar height. Like in LSEM, the smallest regression estimate was the linear influence of age on callousness. As can be seen in Table 3, all regression weights of gender were positive, indicating higher scores for males. The smallest gender difference occurred for narcissistic entitlement. Interaction effects of age and gender were not significant except for a small positive interaction effect for narcissistic entitlement. In summary, the results from the cross-validation indicated that the pattern of results did not differ substantially when MIMIC models (that simply assume rather than investigate structural invariance) were used instead of LSEM.

Supplemental Analysis: Skewness of Manifest Variables

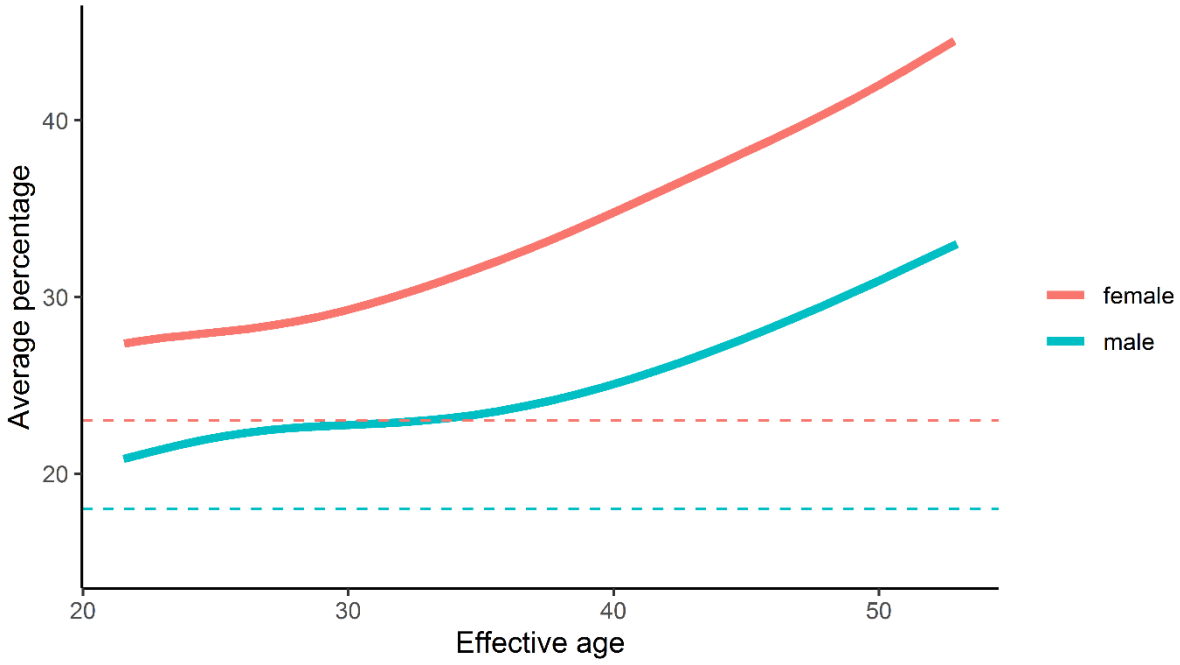
Manifest indicator means, as well as latent factor means, decreased across the age range. The low mean values raise the question about floor effects, particularly for the older age groups. Typically, a floor effect is defined as a substantial proportion of participants receiving the lowest scores (e.g., Terwee et al., 2007). Consequently, participants with the lowest possible score cannot be distinguished from one another, so that age trends might be underestimated.

If the data were normally distributed with a mean of 3 and a standard deviation (*SD*) of 1, about 7% of participants would be expected to have a value below 1.5. Using the observed average means and *SD*s (females: $M = 2.35$, $SD = 1.17$; males: $M = 2.63$, $SD = 1.23$), 23% of the female subsample and 18% of the male subsample would be expected to have a value below 1.5. In fact, 31% of female participants and 23% of male participants have chosen the lowest answer option for one item on average (Table A.1). Figure 5 shows the average percentage of participants with scores of 1 (calculated by averaging the quotient of the number of participants in the weighted sample who scored 1 for an item by the effective

sample size) across effective age. In both gender groups, there was a clear increase in skewness with increasing age: older participants mainly used the lower end of the scale.

Figure 5

Average Percentage of Participants with Scores = 1



Note. Average percentage was calculated by averaging the quotient of the number of participants in the weighted sample who scored 1 for an item by the effective sample size. Dashed lines represent the expected percentage if data were normally distributed around the observed mean value and with the observed SD in the overall female and male samples.

Discussion

Given that age-associated differences in the structure of "dark" personality have received little attention thus far, we herein examined age and gender differences in socially aversive personality traits from early to middle adulthood – between effective ages of 21.5 to 53 years – in a large sample based on a cross-sectional study design using an innovative analytical approach. We thereby relied on a model that includes an overarching factor, which

reflects the general tendency to socially aversive or morally questionable behavior as well as lower-order themes.

Age and gender differences in personality are often discussed without making sure that the measurement works the same way and mostly using raw scores (e.g., Barlett & Barlett, 2015). Thus, the literature might be biased and it is important to test whether previous results hold up when using measurement invariance techniques and latent score comparisons. For the discussion about convergence and divergence of various personality traits as well as which traits are needed to get the full picture of personality features relevant in different contexts, one helpful step is to find out which features are more or less sensitive to age and gender effects. The present investigation asks important questions about the measurement and structure of "dark" personality across age, which could not only be useful for research on dark personality per se but could also help to shed light in the most robust result in criminology, the age-crime curve (Braithwaite, 1989; Farrington, 1986).

The results indicated that measurement and structural properties were mostly stable across age and gender groups. In particular (and contrary to our expectations), the second-order loadings of the themes underlying D were highly stable across the entire age range, indicating no differentiation by age. In line with our second hypothesis, socially aversive personality characteristics uniformly decreased with increasing age. As such, the parallel age-associated differences in means and uniformly strong correlations between the themes in the correlated factors model and the second-order loadings in the higher-order model across the whole age range support the conceptualization and usefulness of D as the common underlying disposition. Factor loadings and reliabilities proved to be generally stable across age in both the female and male subsamples, thereby indicating that the same themes emerge, are measured in the same way, and are equally conceptualized across the groups being studied.

Given the apparent equivalence of the measurement models, it is meaningful to compare structural properties across groups. The D factor and the themes exhibited highly

stable variances across both gender and age groups. The stable omega estimates suggest that roughly the same amount of construct-relevant variance was captured. With respect to the structure of D, we expected an increased specialization of transgressive and norm-violating personality characteristics such that the intercorrelations between the themes in the correlated factor model or the second-order loadings in the higher-order model decrease with increasing age. However, disconfirming this hypothesis, the intercorrelations between the themes as well as their second-order loadings on the D factor were not only stable across gender, but also highly stable across all age groups. Therefore, the results indicated that (self-reported) socially aversive personality characteristics do not become more specialized as individuals grow older, but instead individuals differ in the degree of their general tendency to maximize one's individual utility across the age range. This tendency decreases with age while qualitatively staying the same.

More generally, the strong and uniform correlations among the themes, the highly similar age trends, and the strong and uniform second-order loadings lend support for a common factor underlying all socially aversive traits and a common cause for age-associated decreases in these traits. Note that the correlated factors model and the higher-order model both emphasize the importance and usefulness of the concept of the D factor as common underlying disposition. In particular, the present results suggest that the shared variance between the themes remains essentially identical across the whole age range as well as the proportion of variance explained by D.

Decreasing Mean Levels with Increasing Age

As predicted with our second hypothesis, the mean level of all themes uniformly decreased with increasing age, which is also in line with previous results (e.g., Ali & Chamorro-Premuzic, 2010; Barlett & Barlett, 2015; Blonigen et al., 2006; Grosz et al., 2019; Spurk et al., 2016; Ingo Zettler et al., 2020). However, adding to previous findings, our results rule out structural differences by showing that socially aversive personality characteristics do

not decrease in isolation, but rather decrease in concert. By implication, the age-related decline can also be succinctly described by resorting to the D factor as general underlying disposition. Results from the MIMIC model indicated that with each decade that the sample is older, participants reach on average values that are about .2 standard deviations lower. Whereas Zettler et al. (2020) reported a similar decline over a 4-year period, further longitudinal designs are needed to investigate the determinants of individual differences in the development of D across the life course.

The current pattern of results is in line with the maturity principle (Roberts et al., 2006). Less transgressive and norm-violating behavior can be theorized to be more rewarded as one ages due to changes in culture and social roles (Roberts & Wood, 2006). An increase in disadvantageous effects of socially aversive personality dispositions in the workforce (O'Boyle et al., 2012; Spurk et al., 2016) as well as romantic relationships (Ali & Chamorro-Premuzic, 2010; Jonason et al., 2012) could drive personality development. From a conceptual perspective of the life-history theory, lower levels in aversive traits could be the adaptation to a perceptually more predictable environment. From a biological perspective, endocrinological development could also be responsible (Fabbri et al., 2016; Handelsman et al., 2016). Lower testosterone levels in older adults could form the biological basis for the lower tendency for aggressive, antisocial behavior. It is most likely that differences in personality are driven by a complex interaction of biological and social factors.

Higher Levels of Socially Aversive Personality Dispositions in Men than Women

Gaining more knowledge about gender differences and disparities in age differences across gender in the first step, will help to gain a deeper understanding of the causes of such sex differences in a second step. Before mean-level differences can be discussed, measurement invariance should be established. Otherwise, conclusions regarding mean differences are limited at best and misleading at worst. Our results indicated strict invariance of the D70 questionnaire in a second-order model across gender. Indeed, previous research

suggests that items measuring socially aversive traits seem to operate in a comparable manner in adult females and males (e.g., Dirty Dozen: Chiorri et al., 2019; Machiavellianism: Collison et al., 2020; Pathological Narcissism Inventory: Wright et al., 2010). However, measurement invariance can be scale and sample specific, which is why it should be a routine part of investigations of group differences. In addition to the equally functioning items, second-order loadings and residual variances (or factor variances and covariances in the correlated factors model) could be constrained to be equal across gender groups in the current analysis. Thus, the shared variance between the five themes did not differ by gender. To conclude, the results showed that the measurement model was identical for women and men in the current investigation, indicating that the general tendency towards ethically, morally, and socially questionable behavior neither differs conceptually nor operationally between women and men.

Further, in line with previous findings (Muris et al., 2017), men exhibited higher scores on each theme in both the overall sample as well as in each single age group. Research on gender differences in FFM-agreeableness and honesty-humility has consistently shown that females score higher than males with about d = .40 (Chiorri et al., 2016; Lee & Ashton, 2020; Lippa, 2010; Moshagen et al., 2019; Samuel et al., 2015). Importantly, the decrease in socially aversive personality as a function of age (unstandardized factor means) was similar in shape (linear) and strength (albeit with a slightly stronger decrease for females) for both gender groups, which largely corresponds to the age-related difference observed in dimensions of basic personality (Ashton & Lee, 2016; Klimstra et al., 2020; Moshagen et al., 2019) .

From an evolutionary perspective, socially aversive personality dispositions have been associated with sexual strategies, such as lack of concern to abandon a partner and short-term relationships, that are considered as rather “masculine” (Jonason et al., 2012). The present data suggests that despite theoretically expected differences, D items functioned identically

for women and for men, although female participants did have consistently lower levels of socially aversive personality dispositions. Based on changes of gender-specific social roles and endocrinological developments, one could expect gender specific age trends in personality. However, there is no evidence for gender differences in personality change in general (Roberts et al., 2006) and the current analysis also revealed rather parallel slopes. An exception is narcissistic entitlement for which more pronounced gender differences were evident among older participants. Thus, evidence for an interaction of age and gender was found for this theme and it can be speculated that this effect might be more pronounced in the absence of floor effects. Contrarily, in a meta-analytic investigation of mean-level gender differences in narcissism, gender differences were found to be stable across age with men exhibiting a higher narcissism level (Grijalva et al., 2015). However, age ranges of the primary studies included in this analysis are not reported. Potential age differences could be covered by aggregating samples with large age ranges. Klimstra et al. (2020) found varying mean-level gender differences across age ranging from no significant gender differences to gender differences favoring men, as well as different invariance levels across gender in different age groups in their analysis of the Dirty Dozen. Similarly, empirical results for mean-level gender differences of FFM-agreeableness across age are inconsistent (Klimstra et al., 2020; Marsh et al., 2013). A Japanese study including 30 to 50 year old participants found main effects for age and gender on HEXACO honesty-humility (women and older participants scoring higher), but no interaction effect. More research focusing on gender differences in socially aversive personality traits across the age range is needed to accumulate evidence whether gender differences in these traits differ across the age range.

Discussion of LSEM

LSEM allows for continuous age moderation of measurement model parameters. Due to the weighting of participants on many focal points in LSEM, artificial grouping is not necessary and estimates for comparatively old participants can be examined. By contrast, in a

multiple-group analysis, participants (especially in older age) would need to be allocated to groups with a wide age range or, if groups for every age year would be build, such analysis would only be possible for the younger half of the sample. As there has been no previous research on cross-sectional age effects on the themes investigated in the current study, it was rather exploratory without specific expectations about the shape of the age effect. LSEM showed that mean value effects (intercepts and factor means) are linear effects, which was confirmed with MIMIC analysis. However, as the current analysis is the first to investigate age-associated differences in the five personality themes on a latent level, no hypotheses about the function of age trends would have been possible. LSEM is superior to other data-analytical tools in this case – although no direct inference statistical test of model parameter comparisons was conducted – because an *a priori* specification of the age function of the parameter estimates is not necessary and age is treated as a continuous variable.

Limitations and future research

In the current analysis, we go beyond previous research by investigating cross-sectional differences using the integrative framework of the D-factor with adequate statistical methods in a large sample and thus provide an important contribution to the cumulative knowledge about age and gender-related personality differences. Future research should attempt to replicate the recent findings and overcome some of the limitations of the current investigation. Klimstra et al. (2020) investigated a wider age range and showed that gender differences in raw data varied across the age range when adolescence and older adults are added to the sample. The youngest participants in our data were in early adulthood and the sample was also quite sparse on the older side of the age distribution, which is probably due to recruitment mechanisms and online testing. To obtain a whole life course perspective, we recommend a replication with a larger number of older participants to examine the development of socially aversive personality across the whole life span. From a personality developmental perspective, it would also be interesting to relate the trajectories in socially

aversive personality traits with other personality traits and to study the uniformity of biological and environmental determinants of age-associated differences, which might give important insights on the underlying mechanisms of age-associated differences as well as the unity and diversity of personality constructs. Of course, this should ideally be done using a longitudinal design that entangles cross-sectional and longitudinal differences.

To integrate the current results in the literature on age-related differences in pro-social personality, it is necessary to consider how D relates to dimensions of basic personality. In particular, FFM-agreeableness and HEXACO honesty-humility have been consistently found to be closely related to the commonality of instruments covering anti-socially labeled constructs (e.g., Hodson et al., 2018; Howard & Van Zandt, 2020). The exact nature of this relationship – to what extend the core of socially aversive personality traits overlaps with basic personality traits – is topic of ongoing debate (Ashton & Lee, 2020; Moshagen et al., 2018; Moshagen, Zettler, Horsten, et al., 2020; Schreiber & Marcus, 2020; Vize, Collison, et al., 2020; Vize, Miller, et al., 2020). As such, it would be interesting to consider differences in the age-related trends of these traits as well as potential age-specific covariation. Unfortunately, the present study cannot tackle these issues of trait overlap and comparability of age trends given the data at hand. Future research should pursue these questions to contribute to the understanding of the similarities and differences of these traits.

In the current analysis of age-associated trends, we used separate samples for women and men. Using multivariate LSEM (Hartung et al., 2018) to investigate measurement invariance for gender as categorical variable and age as continuous variable simultaneously might reveal more insight in gender differences in developmental changes. In the current analysis, some loadings in the female and male subsample differed with factor loadings estimated as the optimal weighting of indicators in each subsample. Therefore, different specific items might work better in one subsample than the other (e.g., for young females compared with old males). Unfortunately, there is currently no methodological solution for

such a mixture of variable types in the LSEM context. As such, it would be desirable to use such an implementation in the future. Nonetheless, the measurement model was invariant across gender and the same age trends for most themes were found in both subsamples.

Although the current study yields evidence that the level of socially aversive traits decreases with increasing age, one limitation is that age trends are probably underestimated. The floor effects on items evident in older participants show that the manifestations of socially aversive traits that are covered with the 70 items are not prevalent in older participants. Theoretically, it would be possible that socially aversive traits manifest differently in old age (Olaru et al., 2018). However, unlike other personality questionnaires, none of the items seems obviously age dependent (e.g., "I keep my workplace tidy" as an example personality item that could be biased by retirement of the respondents). None of the D70 items is tied to a specific situation that only specific age-groups encounter (e.g., "Taking credit for someone else's ideas is a no-go."). Nevertheless, it is well possible that the decline with increasing age was underestimated.

Furthermore, there is evidence that socially desirable response behavior increases with increasing age (Soubelet & Salthouse, 2011; Vigil-Colet, Morales-Vives, & Lorenzo-Seva, 2013),[6] thereby providing an alternative explanation for decreasing factor means. Arguably, social desirability is an important effect to consider when measuring socially aversive traits that represent malevolent behavior (e.g., Ray et al., 2013). Therefore, it is to be expected that the floor effects could turn out weaker if social desirability was considered. Nevertheless, in the current analysis, the distribution of responses on the D items especially in older age seems to be bounded by the lower end of the scale. Most of the items of the D70 questionnaire drawn from the item universe of D items stratify well on relatively high D levels, but less well in less socially aversive but more social areas. Taken together, steeper age trends in the

[6] This interpretation rests on the assumption that common measures of socially desirable responding actually measure a motivated response distortion rather than signaling actual honesty (Müller & Moshagen, 2019; Uziel, 2010; Zettler et al., 2015).

socially aversive personality themes would be expected if the distribution in old age was not winsorized by floor effects. Importantly, the factor structure did not differ substantially and the reliability of factors remained quite stable.

Finally, on a related note, the current analyses are based solely on self-report data and thereby interpretation is restricted by all issues related to this data type (e.g., Haeffel & Howard, 2010). Although all types of data come with their problems (Dang et al., 2020; e.g., Gardner, 2001), a broader measurement, in the sense of different data sources, would be beneficial to get a better approximation of the constructs of interest. Differences in tendencies for ethically, morally, and socially questionable behavior could occur when measurement instruments other than self-report would be used. Such data types could include different types of self-report like forced-choice, but also observer or interview-based measures. Deviations in results when using various data types can provide further insight in age and gender differences in personality.

Conclusion

In summary, relying on a large sample and innovative analyses, the present investigation provided evidence for highly stable correlations between socially aversive personality themes and higher-order loadings on a common factor from early to middle adulthood and a uniform decline in latent means as age increases. The strong and stable intercorrelations between or loadings of the themes across the adult age range as well as the parallel mean age trends, in turn, support the conceptualization and usefulness of the common D factor of socially aversive personality traits.

Data Accessibility Statement

The study materials, data and analysis scripts used for this article can be accessed at https://osf.io/4hgub/

References

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

Anusic, I., Lucas, R. E., & Brent Donnellan, M. (2012). Cross-sectional age differences in personality: Evidence from nationally representative samples from Switzerland and the United States. *Journal of Research in Personality*, *46*(1), 116–120. https://doi.org/10.1016/j.jrp.2011.11.002

Ashton, M. C., & Lee, K. (2016). Age trends in HEXACO-PI-R self-reports. *Journal of Research in Personality, 64,* 102–111. https://doi.org/10.1016/j.jrp.2016.08.008

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

Bader, M., Hartung, J., Hilbig, B. E., Zettler, I., Moshagen, M., & Wilhelm, O. (2020). Themes of the dark core of personality. *Manuscript submitted for publication*.

Barlett, C. P., & Barlett, N. D. (2015). The young and the restless: Examining the relationships between age, emerging adulthood variables, and the Dark Triad. *Personality and Individual Differences*, *86,* 20–24.

Berenson, K. R., Ellison, W. D., & Clasing, R. (2017). Age differences in the desirability of narcissism. *Journal of Individual Differences*, *38*(4), 230–240. https://doi.org/10.1027/1614-0001/a000240

Bertl, B., Pietschnig, J., Tran, U. S., Stieger, S., & Voracek, M. (2017). More or less than the sum of its parts? Mapping the Dark Triad of personality onto a single Dark Core. *Personality and Individual Differences*, *114,* 140–144. https://doi.org/10.1016/j.paid.2017.04.002

Blonigen, D. M., Hicks, B. M., Krueger, R. F., Patrick, C. J., & Iacono, W. G. (2006). Continuity and Change in Psychopathic Traits as Measured Via Normal-Range Personality: A Longitudinal–Biometric Study. *Journal of abnormal psychology*, *115*(1), 85–95. https://doi.org/10.1037/0021-843X.115.1.85

Book, A. S., Starzyk, K. B., & Quinsey, V. L. (2001). The relationship between testosterone and aggression: A meta-analysis. *Aggression and Violent Behavior*, *6*(6), 579–599.

Book, A. S., Visser, B. A., Blais, J., Hosker-Field, A., Methot-Jones, T., Gauthier, N. Y., Volk, A., Holden, R. R., & D'Agata, M. T. (2016). Unpacking more "evil": What is at the core of the dark tetrad? *Personality and Individual Differences*, *90*, 269–272.

Borsboom, D. (2006). When does measurement invariance matter? *Medical care*, *44*(11), S176–S181.

Braithwaite, J. (1989). *Crime, shame and reintegration*. Cambridge University Press.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological methods & research*, *21*(2), 230–258.

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A Tutorial on Hierarchically Structured Constructs. *Journal of Personality*, *80*(4), 796–846. https://doi.org/10.1111/j.1467-6494.2011.00749.x

Byrne, B. M., & Stewart, S. M. (2006). Teacher's corner: The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process. *Structural equation modeling*, *13*(2), 287–321.

Carlson, K. S., & Gjerde, P. F. (2009). Preschool personality antecedents of narcissism in adolescence and young adulthood: A 20-year longitudinal study. *Journal of Research in Personality*, *43*(4), 570–578.

Carstensen, L. L., Fung, H. H., & Charles, S. T. (2003). Socioemotional selectivity theory and the regulation of emotion in the second half of life. *Motivation and emotion*, *27*(2), 103–123.

Carter, G. L., Campbell, A. C., Muncer, S., & Carter, K. A. (2015). A Mokken analysis of the Dark Triad 'Dirty Dozen': Sex and age differences in scale structures, and issues with individual items. *Personality and Individual Differences*, *83*, 185–191.

Caspi, A., Roberts, B. W., & Shiner, R. L. (2005). Personality development: Stability and change. *Annu. Rev. Psychol.*, *56*, 453–484.

Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Chen, F. F., Sousa, K. H., & Stephen, G. W. (2005). Testing measurement invariance of second-order factor models. *Structural Equation Modeling*, *12*(3), 471–492.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Chiorri, C., Garofalo, C., & Velotti, P. (2019). Does the Dark Triad Manifest Similarly in men and Women? Measurement Invariance of the Dirty Dozen across sex. *Current Psychology*, *38*(3), 659–675. https://doi.org/10.1007/s12144-017-9641-5

Chiorri, C., Marsh, H. W., Ubbiali, A., & Donati, D. (2016). Testing the factor structure and measurement invariance across gender of the Big Five Inventory through exploratory structural equation modeling. *Journal of Personality Assessment*, *98*(1), 88–99.

Chopik, W. J., & Grimm, K. J. (2019). Longitudinal changes and historic differences in narcissism from adolescence to older adulthood. *Psychology and aging*, *34*(8), 1109-1123.

Collison, K. L., South, S., Vize, C. E., Miller, J. D., & Lynam, D. R. (2020). Exploring Gender Differences in Machiavellianism Using a Measurement Invariance Approach. *Journal of Personality Assessment*. https://doi.org/10.1080/00223891.2020.1729773

Craker, N., & March, E. (2016). The dark side of Facebook®: The Dark Tetrad, negative social potency, and trolling behaviours. *Personality and Individual Differences*, *102*, 79–84.

Dang, J., King, K. M., & Inzlicht, M. (2020). *Why Are Self-Report and Behavioral Measures Weakly Correlated?*. PsyArXiv. https://doi.org/10.31234/osf.io/v796c

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*(4), 367–376.

DiStefano, C. (2002). The Impact of Categorization With Confirmatory Factor Analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(3), 327–346. https://doi.org/10.1207/S15328007SEM0903_2

Eisenegger, C., Haushofer, J., & Fehr, E. (2011). The role of testosterone in social interaction. *Trends in cognitive sciences*, *15*(6), 263–271.

Ellis, B. J., Figueredo, A. J., Brumbach, B. H., & Schlomer, G. L. (2009). Fundamental dimensions of environmental risk. *Human Nature*, *20*(2), 204–268.

Fabbri, E., An, Y., Gonzalez-Freire, M., Zoli, M., Maggio, M., Studenski, S. A., Egan, J. M., Chia, C. W., & Ferrucci, L. (2016). Bioavailable Testosterone Linearly Declines Over A Wide Age Spectrum in Men and Women From The Baltimore Longitudinal Study of Aging. *The Journals of Gerontology: Series A*, *71*(9), 1202–1209. https://doi.org/10.1093/gerona/glw021

Farrington, D. P. (1986). Age and crime. *Crime and justice*, *7*, 189–250.

Fox, J., & Rooney, M. C. (2015). The Dark Triad and trait self-objectification as predictors of men's use and self-presentation behaviors on social networking sites. *Personality and Individual Differences*, *76*, 161–165.

Frick, P. J., Kimonis, E. R., Dandreaux, D. M., & Farell, J. M. (2003). The 4 year stability of psychopathic traits in non-referred youth. *Behavioral sciences & the law*, *21*(6), 713–736.

Gardner, G. (2001). Unreliable memories and other contingencies: Problems with biographical knowledge. *Qualitative Research*, *1*(2), 185–204. https://doi.org/10.1177/146879410100100205

Goldberg, L. R. (2006). Doing it all bass-ackwards: The development of hierarchical factor structures from the top down. *Journal of Research in Personality*, *40*(4), 347–358.

Götz, F. M., Bleidorn, W., & Rentfrow, P. J. (2020). Age differences in Machiavellianism across the life span: Evidence from a large-scale cross-sectional study. *Journal of Personality*. https://doi.org/10.1111/jopy.12545

Graham, E. K., Weston, S. J., Gerstorf, D., Yoneda, T. B., Booth, T., Beam, C. R., Petkus, A. J., Drewelies, J., Hall, A. N., Bastarache, E. D., Estabrook, R., Katz, M. J., Turiano, N. A., Lindenberger, U., Smith, J., Wagner, G. G., Pedersen, N. L., Allemand, M., Spiro, A., … Mroczek, D. K. (2020). Trajectories of Big Five Personality Traits: A Coordinated Analysis of 16 Longitudinal Samples. *European Journal of Personality*, *34*(3), 301–321. https://doi.org/10.1002/per.2259

Greitemeyer, T., & Sagioglou, C. (2017). The longitudinal relationship between everyday sadism and the amount of violent video game play. *Personality and Individual Differences*, *104*, 238–242. https://doi.org/10.1016/j.paid.2016.08.021

Grijalva, E., Newman, D. A., Tay, L., Donnellan, M. B., Harms, P. D., Robins, R. W., & Yan, T. (2015). Gender differences in narcissism: A meta-analytic review. *Psychological bulletin*, *141*(2), 261-310.

Grosz, M. P., Göllner, R., Rose, N., Spengler, M., Trautwein, U., Rauthmann, J. F., Wetzel, E., & Roberts, B. W. (2019). The development of narcissistic admiration and

machiavellianism in early adulthood. *Journal of personality and social psychology*, *116*(3), 467-482.

Haeffel, G. J., & Howard, G. S. (2010). Self-Report: Psychology's Four-Letter Word. *The American Journal of Psychology*, *123*(2), 181–188. https://doi.org/10.5406/amerjpsyc.123.2.0181

Handelsman, D. J., Sikaris, K., & Ly, L. P. (2016). Estimating age-specific trends in circulating testosterone and sex hormone-binding globulin in males and females across the lifespan. *Annals of Clinical Biochemistry*, *53*(3), 377–384. https://doi.org/10.1177/0004563215610589

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

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

Hildebrandt, A., Lüdtke, O., Robitzsch, A., Sommer, C., & Wilhelm, O. (2016). Exploring Factor Model Parameters across Continuous Variables with Local Structural Equation Models. *Multivariate Behavioral Research*, *51*(2–3), 257–258. https://doi.org/10.1080/00273171.2016.1142856

Hildebrandt, A., Wilhelm, O., & Robitzsch, A. (2009). Complementary and competing factor analytic approaches for the investigation of measurement invariance. *Review of Psychology*, *16*(2), 87–102.

Hodson, G., Book, A. S., Visser, B. A., Volk, A. A., Ashton, M. C., & Lee, K. (2018). Is the Dark Triad common factor distinct from low Honesty-Humility? *Journal of Research in Personality*, *73*, 123–129.

Holland, J. H. (1992). Genetic algorithms. *Scientific american*, *267*(1), 66–73.

Howard, M. C., & Van Zandt, E. C. (2020). The discriminant validity of honesty-humility: A meta-analysis of the HEXACO, Big Five, and Dark Triad. *Journal of Research in Personality*, *87*, 103982. https://doi.org/10.1016/j.jrp.2020.103982

James, S., Kavanagh, P. S., Jonason, P. K., Chonody, J. M., & Scrutton, H. E. (2014). The Dark Triad, schadenfreude, and sensational interests: Dark personalities, dark emotions, and dark behaviors. *Personality and Individual Differences*, *68*, 211–216. https://doi.org/10.1016/j.paid.2014.04.020

Jonason, P. K., Luevano, V. X., & Adams, H. M. (2012). How the Dark Triad traits predict relationship choices. *Personality and Individual Differences*, *53*(3), 180–184. https://doi.org/10.1016/j.paid.2012.03.007

Jonason, P. K., & Webster, G. D. (2010). The dirty dozen: A concise measure of the dark triad. *Psychological assessment*, *22*(2), 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*(6), 521–531.

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

Klimstra, T. A., Jeronimus, B. F., Sijtsema, J. J., & Denissen, J. J. A. (2020). The unfolding dark side: Age trends in dark personality features. *Journal of Research in Personality*, *85*, 103915. https://doi.org/10.1016/j.jrp.2020.103915

Koh, K. H., & Zumbo, B. D. (2008). Multi-group confirmatory factor analysis for testing measurement invariance in mixed item format data. *Journal of Modern Applied Statistical Methods*, *7*(2), 12. https://doi.org/10.22237/jmasm/1225512660

Kornadt, A. E., Hagemeyer, B., Neyer, F. J., & Kandler, C. (2018). Sound Body, Sound Mind? The Interrelation Between Health Change and Personality Change in Old Age. *European Journal of Personality*, *32*(1), 30–45. https://doi.org/10.1002/per.2135

Kowalski, C. M., Vernon, P. A., & Schermer, J. A. (2017). Vocational interests and dark personality: Are there dark career choices? *Personality and individual differences*, *104*, 43–47.

Lee, K., & Ashton, M. C. (2020). Sex differences in HEXACO personality characteristics across countries and ethnicities. *Journal of Personality*, *88*(6), 1075–1090. https://doi.org/10.1111/jopy.12551

Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, *48*(3), 936–949. https://doi.org/10.3758/s13428-015-0619-7

Lippa, R. A. (2010). Gender Differences in Personality and Interests: When, Where, and Why? *Social and Personality Psychology Compass*, *4*(11), 1098–1110. https://doi.org/10.1111/j.1751-9004.2010.00320.x

Little, T. D., Slegers, D. W., & Card, N. A. (2006). A Non-arbitrary Method of Identifying and Scaling Latent Variables in SEM and MACS Models. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(1), 59–72. https://doi.org/10.1207/s15328007sem1301_3

Lucas, R. E., & Donnellan, M. B. (2011). Personality development across the life span: Longitudinal analyses with a national sample from Germany. *Journal of Personality and Social Psychology*, *101*(4), 847–861. https://doi.org/10.1037/a0024298

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological methods*, *7*(1), 19–40.

Maples, J. L., Lamkin, J., & Miller, J. D. (2014). A test of two brief measures of the dark triad: The dirty dozen and short dark triad. *Psychological assessment*, *26*(1), 326-331.

Marcus, D. K., & Zeigler-Hill, V. (2015). A big tent of dark personality traits. *Social and Personality Psychology Compass*, *9*(8), 434–446.

Marsh, H. W., Nagengast, B., & Morin, A. J. S. (2013). Measurement invariance of big-five factors over the life span: ESEM tests of gender, age, plasticity, maturity, and la dolce vita effects. *Developmental Psychology, 49*(6), 1194–1218. https://doi.org/10.1037/a0026913

Mazur, A., & Booth, A. (1998). Testosterone and dominance in men. *Behavioral and brain sciences, 21*(3), 353–363.

McDonald, R. P. (2013). *Test theory: A unified treatment*. Psychology Press.

Međedović, J., & Petrović, B. (2015). The Dark Tetrad. *Journal of Individual Differences*.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525–543.

Miller, J. D., Few, L. R., Seibert, L. A., Watts, A., Zeichner, A., & Lynam, D. R. (2012). An examination of the Dirty Dozen measure of psychopathy: A cautionary tale about the costs of brief measures. *Psychological Assessment, 24*(4), 1048-1053.

Miller, J. D., Hyatt, C. S., Maples-Keller, J. L., Carter, N. T., & Lynam, D. R. (2017). Psychopathy and Machiavellianism: A Distinction Without a Difference? *Journal of Personality*, *85*(4), 439–453. https://doi.org/10.1111/jopy.12251

Milojev, P., & Sibley, C. G. (2017). Normative personality trait development in adulthood: A 6-year cohort-sequential growth model. *Journal of Personality and Social Psychology, 112*(3), 510–526. https://doi.org/10.1037/pspp0000121

Moshagen, M., & Auerswald, M. (2018). On congruence and incongruence of measures of fit in structural equation modeling. *Psychological methods*, *23*(2), 318-336.

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

Moshagen, M., Thielmann, I., Hilbig, B. E., & Zettler, I. (2019). Meta-Analytic Investigations of the HEXACO Personality Inventory(-Revised). *Zeitschrift für Psychologie*, *227*(3), 186–194. https://doi.org/10.1027/2151-2604/a000377

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

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

Müller, S., & Moshagen, M. (2019). True virtue, self-presentation, or both?: A behavioral test of impression management and overclaiming. *Psychological assessment, 31*(2), 181-191.

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*(2), 183–204.

O'Boyle, E. H., Forsyth, D., 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*(3), 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*(6), 644–664.

Olaru, G., Schroeders, U., Hartung, J., & Wilhelm, O. (2019). A tutorial on novel item and person sampling procedures for personality research. *European Journal of Personality*, *3*(33), 400–419. https://doi.org/10.1002/per.2195

Olaru, G., Schroeders, U., Wilhelm, O., & Ostendorf, F. (2018). A confirmatory examination of age-associated personality differences: Deriving age-related measurement-invariant solutions using ant colony optimization. *Journal of Personality*, *86*(6), 1037–1049. https://doi.org/10.1111/jopy.12373

Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of research in personality*, *36*(6), 556–563.

R Development Core Team. (2001). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Ray, J. V., Hall, J., Rivera-Hudson, N., Poythress, N. G., Lilienfeld, S. O., & Morano, M. (2013). The relation between self-reported psychopathic traits and distorted response styles: A meta-analytic review. *Personality Disorders: Theory, Research, and Treatment*, *4*(1), 1–14. https://doi.org/10.1037/a0026482

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological bulletin*, *132*(1), 1-25. https://doi.org/10.1037/0033-2909.132.1.1

Roberts, B. W., & Wood, D. (2006). *Personality Development in the Context of the Neo-Socioanalytic Model of Personality.* In D. K. Mroczek & T. D. Little (Eds.), *Handbook of personality development* (p. 11–39). Lawrence Erlbaum Associates Publishers.

Robitzsch, A. (2019). *sirt: Supplementary Item Response Theory Models* (R package version 3.6-21) [Computer software]. https://CRAN.R-project.org/package=sirt

Samuel, D. B., South, S. C., & Griffin, S. A. (2015). Factorial invariance of the five-factor model rating form across gender. *Assessment*, *22*(1), 65–75.

Schreiber, A., & Marcus, B. (2020). The Place of the 'Dark Triad'in General Models of Personality: Some Meta-Analytic Clarification. *Psychological Bulletin*, *146*(11), 1021–1041. https://doi.org/10.1037/bul0000299

Schroeders, U., Wilhelm, O., & Olaru, G. (2016). Meta-heuristics in short scale construction: Ant colony optimization and genetic algorithm. *PloS one*, *11*(11), e0167110.

Schwaba, T., & Bleidorn, W. (2018). Individual differences in personality change across the adult life span. *Journal of Personality*, *86*(3), 450–464. https://doi.org/10.1111/jopy.12327

Sijtsema, J. J., Garofalo, C., Jansen, K., & Klimstra, T. A. (2019). Disengaging from Evil: Longitudinal Associations Between the Dark Triad, Moral Disengagement, and Antisocial Behavior in Adolescence. *Journal of Abnormal Child Psychology*, *47*(8), 1351–1365. https://doi.org/10.1007/s10802-019-00519-4

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2011). Age differences in personality traits from 10 to 65: Big Five domains and facets in a large cross-sectional sample. *Journal of personality and social psychology*, *100*(2), 330-348. https://doi.org/10.1037/a0021717

Soubelet, A., & Salthouse, T. A. (2011). Influence of Social Desirability on Age Differences in Self-Reports of Mood and Personality. *Journal of Personality*, *79*(4), 741–762. https://doi.org/10.1111/j.1467-6494.2011.00700.x

Specht, J., Bleidorn, W., Denissen, J. J. A., Hennecke, M., Hutteman, R., Kandler, C., Luhmann, M., Orth, U., Reitz, A. K., & Zimmermann, J. (2014). What Drives Adult Personality Development? A Comparison of Theoretical Perspectives and Empirical Evidence. *European Journal of Personality*, *28*(3), 216–230. https://doi.org/10.1002/per.1966

Spurk, D., & Hirschi, A. (2018). The Dark Triad and competitive psychological climate at work: A model of reciprocal relationships in dependence of age and organization change. *European Journal of Work and Organizational Psychology*, *27*(6), 736–751.

Spurk, D., Keller, A. C., & Hirschi, A. (2016). Do Bad Guys Get Ahead or Fall Behind? Relationships of the Dark Triad of Personality With Objective and Subjective Career Success. *Social Psychological and Personality Science*, *7*(2), 113–121. https://doi.org/10.1177/1948550615609735

Srivastava, S., John, O. P., Gosling, S. D., & Potter, J. (2003). Development of personality in early and middle adulthood: Set like plaster or persistent change? *Journal of personality and social psychology*, *84*(5), 1041-1053. https://doi.org/10.1037/0022-3514.84.5.1041

Staudinger, U. M., & Kessler, E.-M. (2009). *Adjustment and growth: Two trajectories of positive personality development across adulthood.* In M. C. Smith & N. DeFrates-Densch (Eds.), *Handbook of research on adult learning and development* (p. 242–268). Routledge/Taylor & Francis Group.

Terracciano, A., McCrae, R. R., Brant, L. J., & Costa Jr, P. T. (2005). Hierarchical linear modeling analyses of the NEO-PI-R scales in the Baltimore Longitudinal Study of Aging. *Psychology and aging*, *20*(3), 493-506. https://doi.org/10.1037/0882-7974.20.3.493

Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of clinical epidemiology*, *60*(1), 34–42.

Tucker-Drob, E. M., Briley, D. A., McAdams, D. P., Shiner, R. L., & Tackett, J. L. (2019). Theoretical concepts in the genetics of personality development. *The handbook of personality development*, 40–58.

Uziel, L. (2010). Rethinking social desirability scales: From impression management to interpersonally oriented self-control. *Perspectives on Psychological Science*, *5*(3), 243–262.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational research methods*, *3*(1), 4–70.

Veselka, L., & Vernon, P. A. (2014). Introduction to the special issue on the Dark Triad and related traits. *Personality and Individual Differences, 67,* 1. https://doi.org/10.1016/j.paid.2014.03.027

Vigil-Colet, A., Morales-Vives, F., & Lorenzo-Seva, U. (2013). How social desirability and acquiescence affect the age-personality relationship. *Psicothema, 25*(3), 342–348.

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

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

Vize, C. E., Miller, J., & Lynam, D. (2020). *Examining the Conceptual and Empirical Distinctiveness of Agreeableness and "Dark" Personality Items* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/kf85p

Volmer, J., Koch, I. K., & Wolff, C. (2019). Illuminating the 'dark core': Mapping global versus specific sources of variance across multiple measures of the dark triad. *Personality and Individual Differences*, *145*, 97–102.

Wagner, J., Lüdtke, O., & Voelkle, M. C. (2019). Using Dynamic Panel Models to Study Age-related Differences and Time-related Changes in Personality. *European Journal of Personality*, *33*(3), 420–434. https://doi.org/10.1002/per.2200

Wetzel, E., Grijalva, E., Robins, R. W., & Roberts, B. W. (2019). You're still so vain: Changes in narcissism from young adulthood to middle age. *Journal of personality and social psychology, 119*(2), 479-496. https://doi.org/10.1037/pspp0000266

Wille, B., Hofmans, J., Lievens, F., Back, M. D., & De Fruyt, F. (2019). Climbing the corporate ladder and within-person changes in narcissism: Reciprocal relationships over two decades. *Journal of Vocational Behavior*, *115*, 103341.

Wilson, M. S., & Sibley, C. G. (2011). „Narcissism creep?“: Evidence for Age-Related Differences in Narcissism in the New Zealand General Population. *New Zealand Journal of Psychology*, *40*(3), 89–95.

Wright, A. G., Lukowitsky, M. R., Pincus, A. L., & Conroy, D. E. (2010). The higher order factor structure and gender invariance of the Pathological Narcissism Inventory. *Assessment*, *17*(4), 467–483.

Yuan, K.-H., & Bentler, P. M. (2000). 5. Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological methodology*, *30*(1), 165–200.

Zettler, I., Schild, C., Lilleholt, L., Kroencke, L., & Utesch, T. (2020). The role of personality in COVID-19 related perceptions, evaluations, and behaviors: Findings across five samples, nine traits, and 17 criteria. *PsyArXiv*. https://doi.org/10.31234/osf.io/pkm2a

Zettler, Ingo, Hilbig, B. E., Moshagen, M., & de Vries, R. E. (2015). Dishonest responding or true virtue? A behavioral test of impression management. *Personality and Individual Differences*, *81*, 107–111.

Zettler, Ingo, Moshagen, M., & Hilbig, B. E. (2020). Stability and Change: The dark factor of personality shapes dark traits. *Social Psychological and Personality Science*. https://doi.org/10.1177/1948550620953288

Table 1

Unstandardized Latent Factor Means and Standard Deviations in the Metric of the Response Scales (1 = strongly disagree, 5 = strongly agree)

	overall sample		female		male	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
D	2.49	0.69	2.35	0.65	2.62	0.70
Callousness	2.31	0.72	2.15	0.66	2.46	0.74
Deceitfulness	2.42	0.76	2.26	0.68	2.57	0.79
Narcissistic Entitlement	2.53	0.71	2.46	0.69	2.59	0.72
Sadism	2.32	0.78	2.18	0.73	2.45	0.79
Vindictiveness	2.88	0.79	2.70	0.77	3.05	0.78

Note. Estimates for the five themes are taken from the correlated factors model. Estimates for the D factor are taken from the higher-order model. N = 12,501, n_{female} =5,982, and n_{male} = 6,519.

Table 2

Tests of Measurement Invariance of the Second-Order Model across Gender

	Model	$\chi^2(df)$	*p*	*RMSEA*	*SRMR*	Model Comparison	$\Delta\chi^2(\Delta df)$	*p*	*ΔCFI*	Δ*RMSEA*
1.	Configural	61,043 (4,680)	< .001	.047	.042					
2.	first-order factor loadings invariant	61,675 (4,745)	< .001	.047	.045	2 vs. 1	592 (65)	< .001	-.002	0
3.	first- and second-order factor loadings invariant	61,769 (4,749)	< .001	.047	.046	3 vs. 2	98 (4)	< .001	0	0
4.	loadings and indicator intercepts invariant	63,689 (4,814)	< .001	.048	.048	4 vs. 3	2,077 (65)	< .001	-.005	.001
5.	loadings, and indicator and first-order intercepts invariant	63,922 (4,818)	< .001	.048	.048	5 vs. 4	257 (4)	< .001	0	0
6.	loadings, intercepts, and residual variances of first-order factors invariant	64,002 (4,823)	< .001	.048	.048	6 vs. 5	79 (5)	< .001	-.001	0
7.	loadings, intercepts, and residual variances of first-order factors and indicators invariant	65,628 (4,893)	< .001	.048	.049	7 vs. 6	1,560 (70)	< .001	-.004	0

Note. df = degrees of freedom; *RMSEA* = robust root mean square error of approximation; *SRMR* = standardized root mean square residual; *CFI* = robust comparative fit index; ΔCFI and ΔRMSEA indices were calculated by subtracting the value of the less constrained from the more constrained model; $\Delta\chi^2$ calculated by the scaled difference between the χ^2 statistics (Yuan & Bentler, 2000). N = 12,051; n_{female} =5,982 and n_{male} = 6,519.

Table 3

Regression Weights from MIMIC Analysis

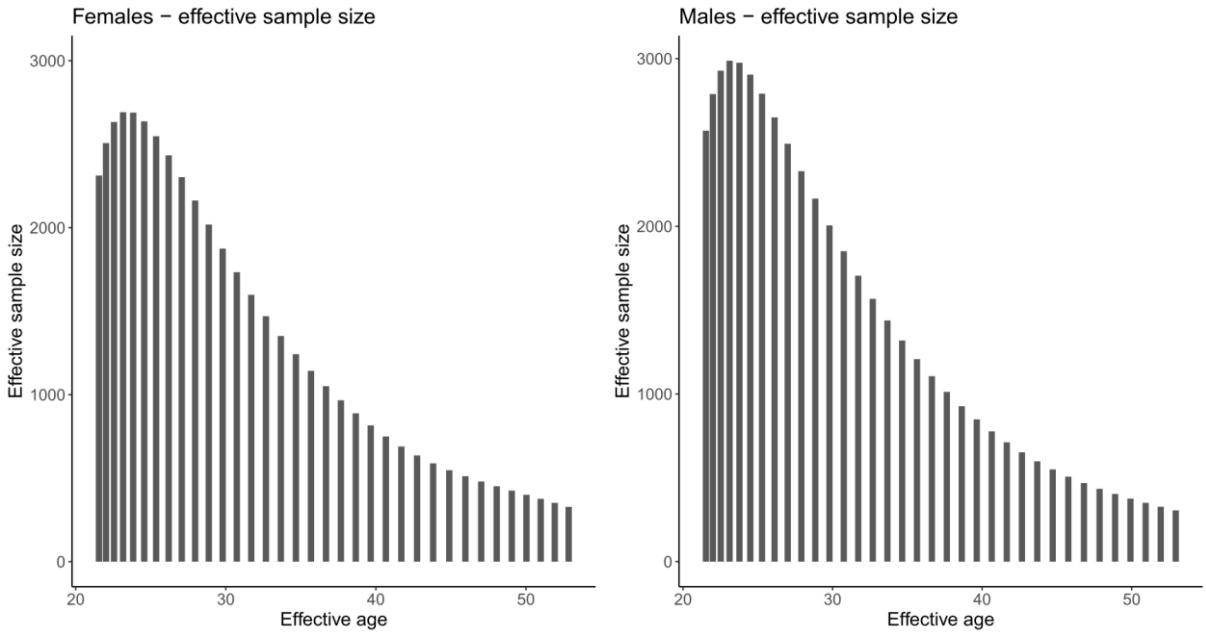
	Age		Gender	Interaction
	unstandardized	standardized	standardized	standardized
D	-.02	-.26	.20	.00 (.41)
Callousness	-.01	-.18	.21	-.00 (.43)
Deceitfulness	-.02	-.26	.19	-.01 (.38)
Narcissistic Entitlement	-.02	-.28	.09	.02 (.01)
Sadism	-.02	-.22	.18	.00 (.36)
Vindictiveness	-.02	-.28	.21	.01 (.43)

Note. All estimates for age and gender are significant with $p < .001$. p values for the interaction effects are reported in brackets. $N = 12{,}501$.

Appendix

Figure A.1

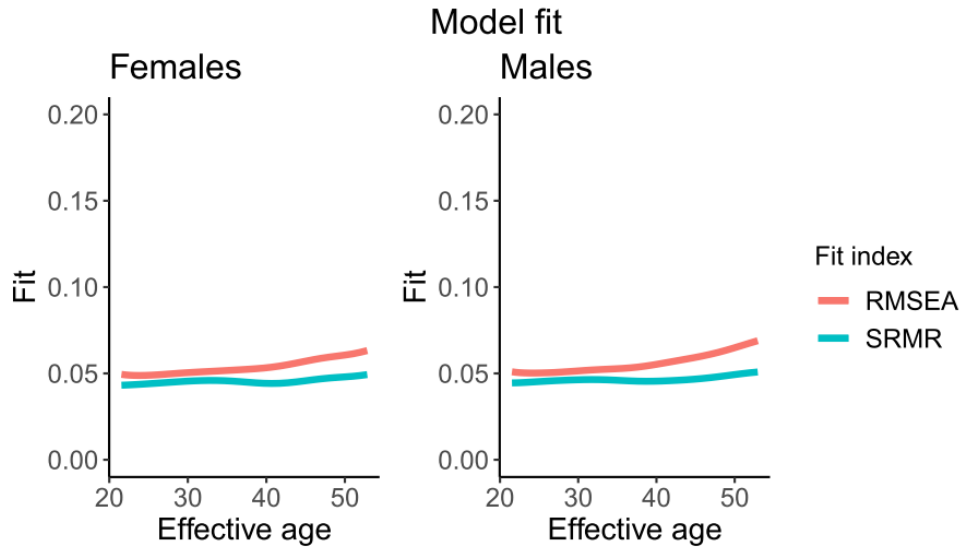
Effective Sample Size) at Every Focal Point Estimated with LSEM



Note. Resulting from the weighting procedure.

Figure A.2

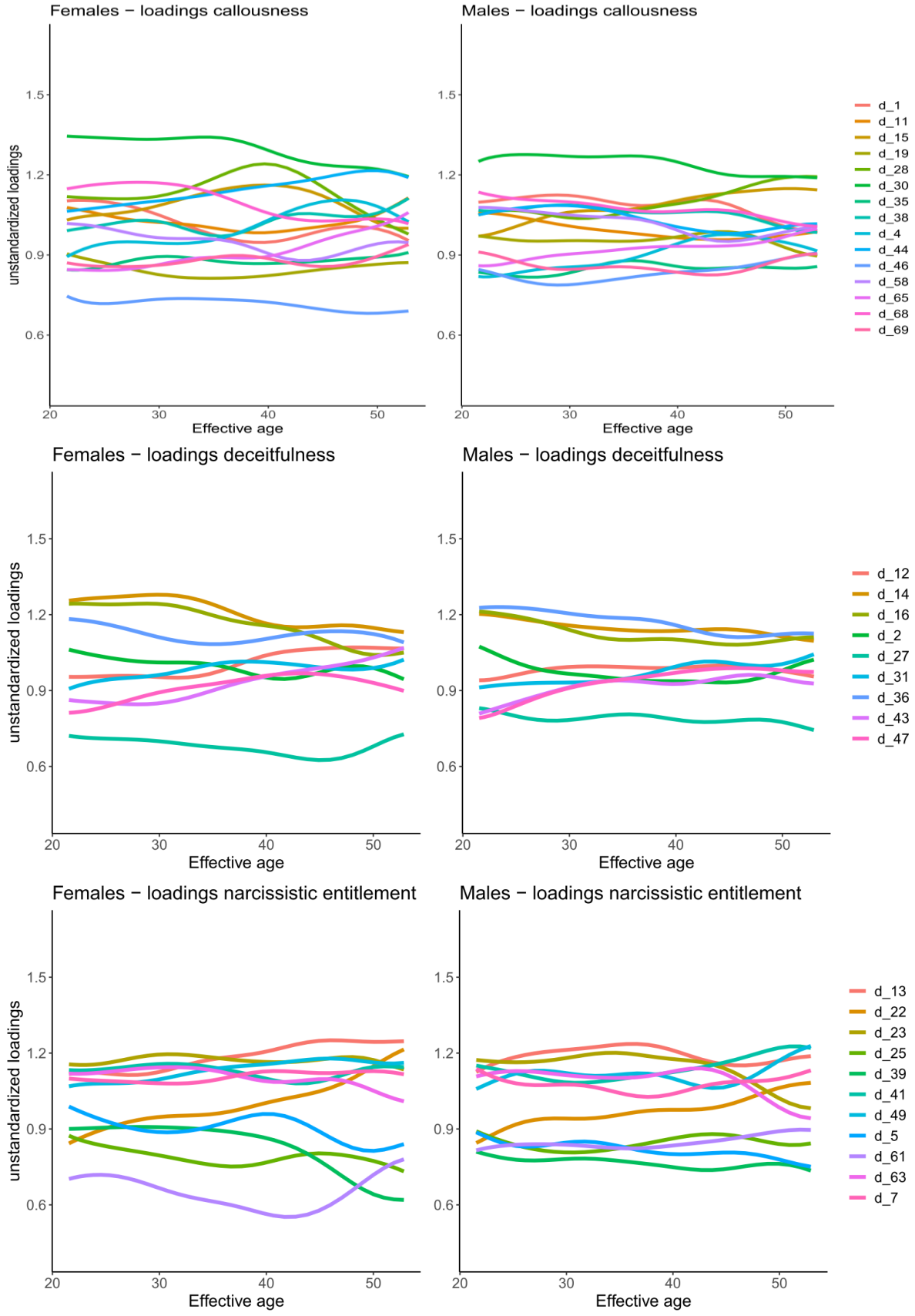
Age Gradients of Robust Root Mean Squared Error of Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR)

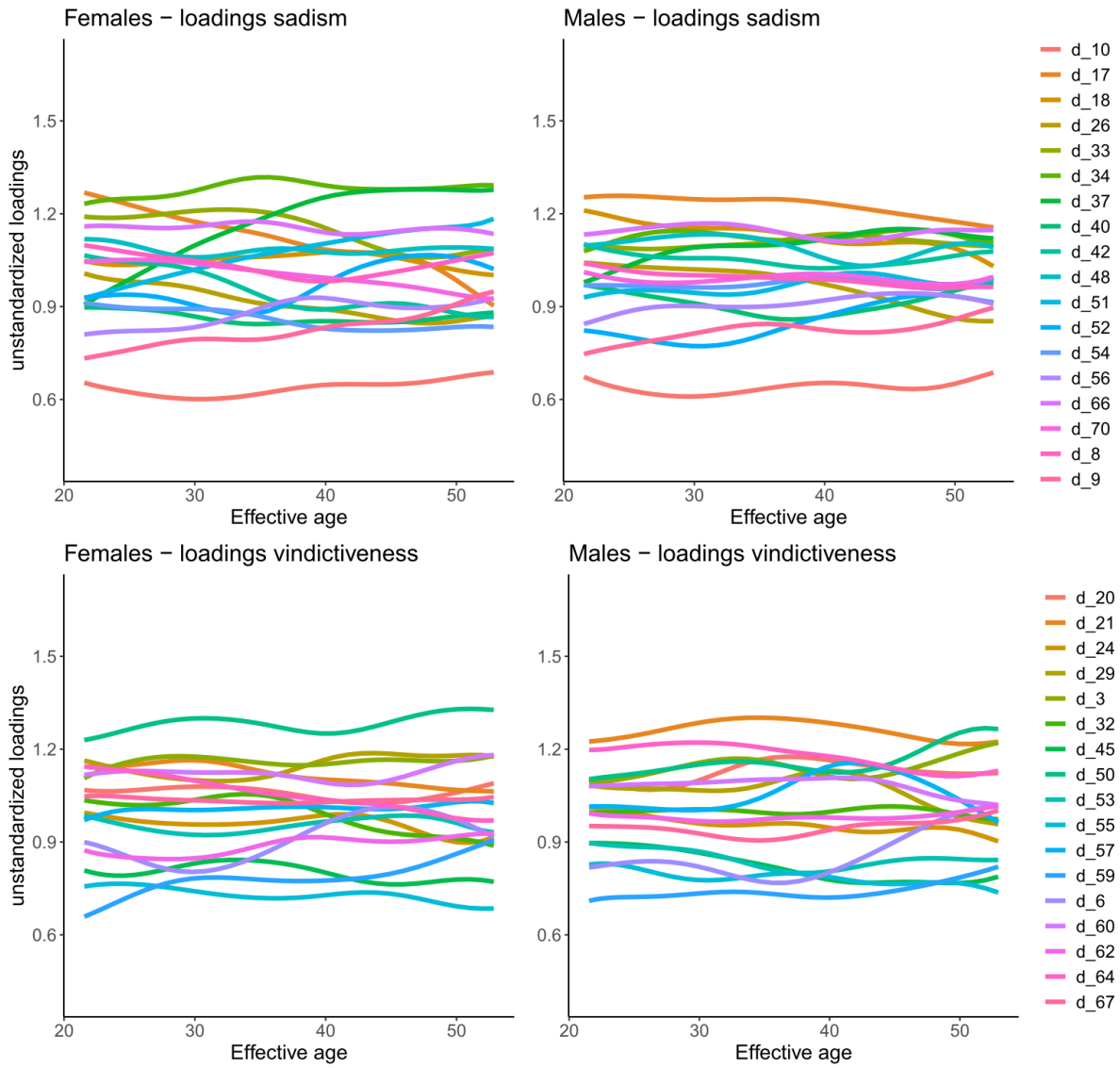


Note. Estimated with local structural equation modeling for a measurement model for the D70 questionnaire with five first-order factors and a second-order factor.

Figure A.3

Age Gradients of First-Order Factor Loadings Estimated with LSEM

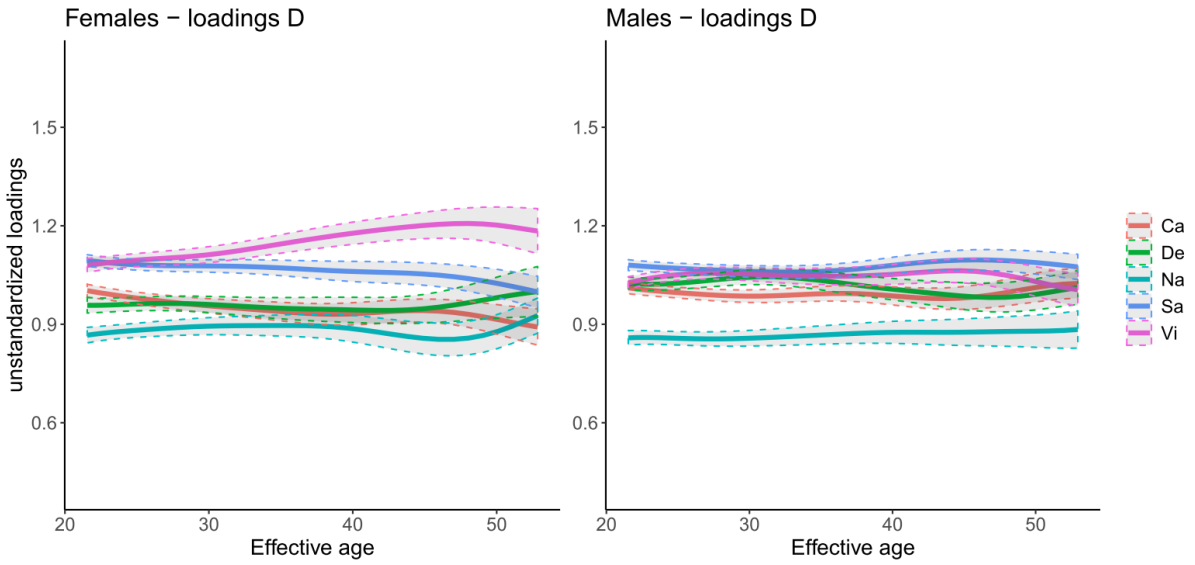




Note. Graphs with confidence intervals can be found on the OSF repository (https://osf.io/4hgub/).

Figure A. 4

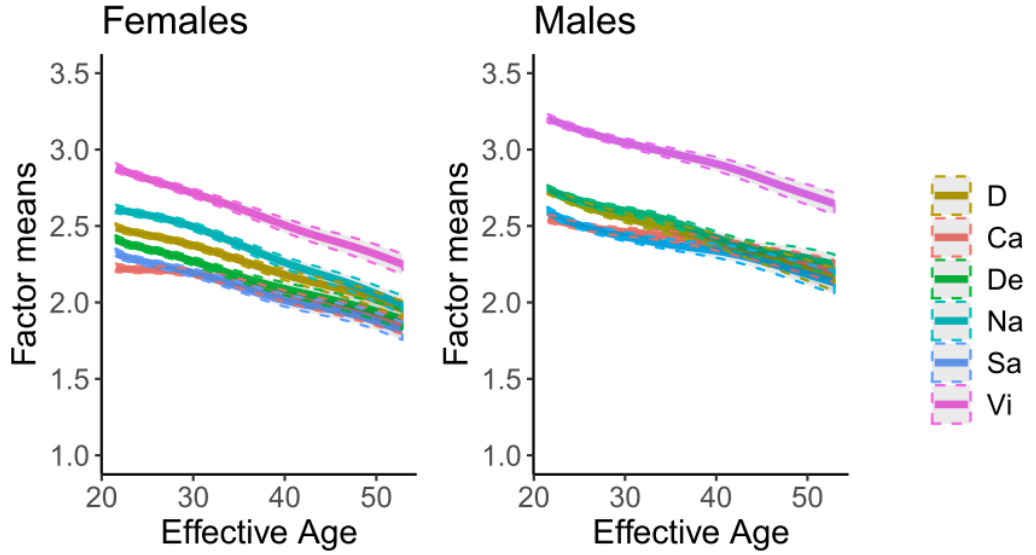
Age Gradients of Second-Order Factor Loadings Estimated with LSEM



Note. Solid lines represent the estimated factor loadings from CFA. Dotted lines represent the 95% confidence intervals. Ca = Callousness, De = Deceitfulness, Na = Narcissistic Entitlement, Sa = Sadism, Vi = Vindictiveness.

Figure A.5

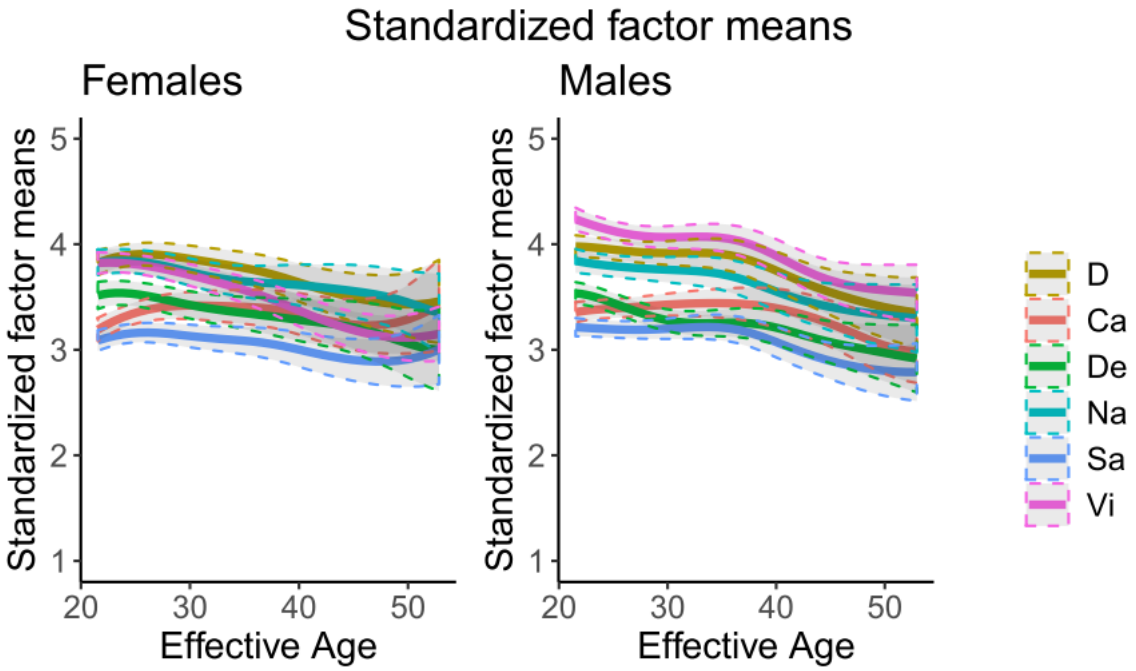
Age Gradients of Unstandardized Factor Means Estimated with LSEM



Note. Solid lines represent the estimated factor loadings from CFA. Dotted lines represent the 95% confidence intervals. Estimates for the five themes are derived from the correlated factors model. Estimates for the D factor are derived from the higher order-model. Ca = Callousness, De = Deceitfulness, Na = Narcissistic Entitlement, Sa = Sadism, Vi = Vindictiveness.

Figure A.6

Age Gradients of Standardized Factor Means Estimated with LSEM



Note. Solid lines represent the estimated factor loadings from CFA. Dotted lines represent the 95% confidence intervals. Estimates for the five themes are derived from the correlated factors model. Estimates for the D factor are derived from the higher order-model. Ca = Callousness, De = Deceitfulness, Na = Narcissistic Entitlement, Sa = Sadism, Vi = Vindictiveness.

Table A.1

Descriptive Statistics of the D70 Items for the Unweighted Overall Female and Male Subsamples

	female subsample						male subsample					
	M	*SD*	median	skew	kurtosis	% 1*	*M*	*SD*	median	skew	kurtosis	% 1*
d_1	1.92	1.09	2	1.18	0.63	45	2.43	1.26	2	0.62	-0.70	27
d_2	1.79	1.00	2	1.33	1.26	50	2.02	1.12	2	1.03	0.25	40
d_3	3.05	1.32	3	-0.13	-1.20	16	3.31	1.29	4	-0.35	-1.03	11
d_4	2.87	1.20	3	0.23	-0.98	11	3.05	1.22	3	0.12	-1.07	8
d_5	2.06	1.05	2	0.97	0.27	34	2.05	1.03	2	1.02	0.51	33
d_6	2.14	1.12	2	0.81	-0.18	34	2.45	1.19	2	0.55	-0.61	24
d_7	2.29	1.25	2	0.73	-0.57	33	2.40	1.32	2	0.65	-0.75	31
d_8	1.99	1.18	2	0.99	-0.19	47	2.11	1.21	2	0.87	-0.38	42
d_9	2.45	1.23	2	0.48	-0.90	27	2.68	1.28	2	0.27	-1.10	21
d_10	2.00	0.96	2	1.14	1.12	32	2.15	1.02	2	1.02	0.69	26
d_11	2.06	1.05	2	0.95	0.25	34	2.33	1.16	2	0.68	-0.42	27
d_12	2.08	1.13	2	0.92	-0.07	38	2.49	1.24	2	0.47	-0.88	25
d_13	2.81	1.25	3	0.15	-1.13	16	2.94	1.27	3	0.07	-1.16	14
d_14	2.29	1.21	2	0.65	-0.67	32	2.69	1.31	2	0.28	-1.12	22
d_15	2.29	1.30	2	0.71	-0.74	35	2.66	1.45	2	0.36	-1.30	29
d_16	2.22	1.19	2	0.82	-0.32	33	2.43	1.27	2	0.61	-0.78	27
d_17	1.90	1.12	2	1.19	0.49	49	2.35	1.27	2	0.63	-0.75	32
d_18	1.91	1.25	1	1.26	0.36	54	2.24	1.40	2	0.81	-0.76	43
d_19	1.64	0.89	1	1.64	2.71	55	1.91	1.03	2	1.19	0.91	43
d_20	2.89	1.29	3	0.11	-1.16	16	3.22	1.31	3	-0.19	-1.18	11
d_21	2.43	1.23	2	0.52	-0.80	27	2.90	1.32	3	0.10	-1.18	17
d_22	2.89	1.22	3	-0.04	-1.16	15	3.06	1.24	3	-0.17	-1.11	13
d_23	2.66	1.26	2	0.23	-1.15	21	2.82	1.29	3	0.14	-1.19	18

	female subsample						male subsample					
	M	*SD*	median	skew	kurtosis	% 1*	*M*	*SD*	median	skew	kurtosis	% 1*
d_24	2.69	1.22	3	0.27	-0.95	19	2.90	1.19	3	0.17	-0.91	12
d_25	1.95	0.96	2	1.19	1.16	35	1.97	1.02	2	1.19	0.96	36
d_26	1.71	1.08	1	1.53	1.36	61	2.03	1.25	2	1.00	-0.24	48
d_27	1.62	0.92	1	1.73	2.76	59	1.90	1.09	2	1.19	0.58	47
d_28	2.19	1.20	2	0.85	-0.30	35	2.63	1.27	2	0.37	-1.00	21
d_29	2.58	1.36	2	0.30	-1.30	29	3.01	1.37	3	-0.10	-1.33	18
d_30	2.36	1.27	2	0.64	-0.76	31	2.62	1.35	2	0.40	-1.09	25
d_31	2.66	1.33	2	0.18	-1.29	26	2.92	1.37	3	-0.01	-1.31	21
d_32	2.58	1.28	2	0.33	-1.09	25	2.96	1.30	3	-0.01	-1.19	16
d_33	2.37	1.29	2	0.58	-0.90	33	2.57	1.32	2	0.42	-1.06	26
d_34	2.43	1.34	2	0.43	-1.18	34	2.78	1.36	3	0.11	-1.31	23
d_35	2.10	0.99	2	0.89	0.40	29	2.56	1.12	2	0.51	-0.51	16
d_36	2.20	1.25	2	0.84	-0.44	37	2.54	1.40	2	0.48	-1.12	30
d_37	2.65	1.26	2	0.30	-1.06	21	2.89	1.29	3	0.12	-1.16	16
d_38	1.95	1.06	2	1.16	0.70	41	2.37	1.25	2	0.70	-0.58	28
d_39	2.05	1.10	2	0.93	0.01	38	2.09	1.10	2	0.87	-0.07	36
d_40	1.96	1.01	2	1.10	0.79	38	2.27	1.12	2	0.79	-0.16	27
d_41	2.56	1.25	2	0.39	-0.96	23	2.84	1.30	3	0.14	-1.13	18
d_42	1.85	0.99	2	1.16	0.84	46	2.26	1.15	2	0.72	-0.30	30
d_43	2.96	1.28	3	0.03	-1.16	15	3.21	1.33	3	-0.17	-1.21	12
d_44	2.58	1.17	2	0.37	-0.92	19	2.88	1.19	3	0.13	-1.04	12
d_45	2.45	1.16	2	0.53	-0.65	23	2.65	1.26	2	0.36	-0.97	20
d_46	1.61	0.78	1	1.46	2.67	53	1.90	0.94	2	1.19	1.42	38
d_47	2.53	1.19	2	0.42	-0.82	21	2.90	1.29	3	0.13	-1.15	15
d_48	1.99	1.26	1	1.09	-0.04	50	2.27	1.39	2	0.76	-0.80	41

	female subsample						male subsample					
	M	*SD*	median	skew	kurtosis	% 1*	*M*	*SD*	median	skew	kurtosis	% 1*
d_49	2.89	1.29	3	0.09	-1.22	16	3.05	1.31	3	-0.06	-1.24	14
d_50	3.10	1.35	3	-0.15	-1.23	16	3.50	1.27	4	-0.52	-0.84	9
d_51	3.26	1.23	4	-0.29	-1.01	9	3.42	1.23	4	-0.39	-0.93	7
d_52	1.93	1.12	2	1.15	0.42	46	2.06	1.14	2	0.98	0.06	39
d_53	3.46	1.24	4	-0.34	-1.07	6	3.74	1.18	4	-0.65	-0.64	4
d_54	1.91	1.03	2	1.13	0.60	43	2.27	1.17	2	0.68	-0.53	31
d_55	2.24	1.16	2	0.69	-0.51	32	2.60	1.22	2	0.34	-0.97	21
d_56	2.97	1.15	3	-0.03	-1.00	10	3.22	1.18	3	-0.20	-0.98	8
d_57	2.27	1.33	2	0.75	-0.74	38	2.66	1.43	2	0.32	-1.31	28
d_58	1.81	1.01	2	1.33	1.28	48	2.16	1.16	2	0.86	-0.16	35
d_59	3.05	1.16	3	-0.20	-1.02	11	3.32	1.17	4	-0.40	-0.84	8
d_60	2.66	1.34	3	0.24	-1.18	26	3.08	1.35	3	-0.11	-1.19	17
d_61	2.00	0.97	2	1.14	1.15	32	2.10	1.05	2	1.02	0.49	31
d_62	2.80	1.15	3	0.14	-0.89	14	2.98	1.18	3	0.03	-0.95	11
d_63	2.89	1.28	3	0.00	-1.18	17	3.20	1.31	3	-0.24	-1.13	13
d_64	2.41	1.24	2	0.54	-0.83	28	2.83	1.32	3	0.18	-1.19	18
d_65	2.75	1.23	2	0.35	-0.97	14	2.80	1.23	3	0.32	-0.99	13
d_66	2.04	1.14	2	0.95	-0.03	41	2.38	1.22	2	0.60	-0.66	28
d_67	3.17	1.23	3	-0.19	-1.04	10	3.69	1.17	4	-0.66	-0.55	5
d_68	1.97	1.07	2	1.12	0.61	.41	2.33	1.19	2	0.74	-0.38	27
d_69	2.13	0.97	2	0.97	0.68	25	2.22	1.01	2	0.98	0.59	22
d_70	1.96	1.13	2	1.04	0.04	46	2.16	1.18	2	0.78	-0.43	36

Note. *percentage of participants, who chose the lowest possible response option. Verbatim items can be downloaded from http://www.darkfactor.org. n_{female} =5,982 and n_{male} = 6,519.