ORIGINAL ARTICLE

Testing the equivalence of the aversive core of personality and a blend of agreeableness(-related) items

Luisa K. Horsten[1] | Isabel Thielmann[2] | Morten Moshagen[3] | Ingo Zettler[4] |
David Scholz[1] | Benjamin E. Hilbig[1]

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

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

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

[4]Institut for Psykologi, University of Copenhagen, Copenhagen, Denmark

Correspondence
Luisa K. Horsten, Experimental Psychology & Personality, RPTU Kaiserslautern-Landau, Fortstr. 7, Landau 76829, Germany.
Email: horsten@uni-landau.de

Funding information
Deutsche Forschungsgemeinschaft

Abstract

Objective: There is an ongoing debate in personality research whether the common core of aversive ("dark") traits can be approximated by or even considered equivalent to one of the constructs that have been labeled "Agreeableness". In particular, it has been suggested that the low pole of (what we term) AG+, a broad blend of Big Five Agreeableness and the HEXACO factors Honesty-Humility, Agreeableness, and Altruism, is essentially equivalent to the Dark Factor of Personality (D). Based on theoretical differences, we herein test empirically whether D and AG+ are isomorphic.

Methods: Self-report data on D, AG+, and eight criterion measures reflecting justifying beliefs, inflicting disutility on others, and affiliative tendencies were collected in a pre-registered study ($N = 1156$) and analyzed via confirmatory factor modeling.

Results: Results speak against unity of D and AG+ (35% shared variance) and support the notion that D subsumes a broader range of aversive content (i.e., justifying beliefs and inflicting disutility on others) than AG+, which, in turn, subsumes a slightly broader range of non-aversive, affiliative tendencies.

Conclusion: We conclude that AG+ is non-equivalent to the common core of aversive traits, D.

KEYWORDS
agreeableness, aversive traits, dark core of personality

1 | INTRODUCTION

People differ in their tendencies to behave in unethical and socially harmful ways. These differences are commonly ascribed to ethically and socially aversive (often called "dark") personality traits. Prominent examples of such aversive traits are the Dark Triad components—Narcissism, Machiavellianism, and Psychopathy (Paulhus & Williams, 2002)—but many more aversive traits have been proposed (e.g., Sadism, O'Meara et al., 2011; Spitefulness, Marcus et al., 2014). Importantly, aversive traits are conceptually similar and highly inter-correlated (e.g., Egan et al., 2015; Jonason et al., 2017; Muris et al., 2017), and there is now strong consensus that they share a common core (Jonason

et al., 2017; Muris et al., 2017; Schreiber & Marcus, 2020; Vize & Lynam, 2020). This common core has been 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) and termed the Dark Factor of Personality (D).

Recent research has attempted to approximate the common core of aversive traits via established personality constructs, one of which is (low) Agreeableness as traditionally conceptualized within the Big Five framework (Paulhus & Williams, 2002; Stead & Fekken, 2014; Vize, Collison, et al., 2020). By this logic, features of low Agreeableness—such as a lack of empathy and compassion, as well as a tendency for manipulation and antagonism (Graziano & Tobin, 2017)—also represent the shared features of aversive traits. Indeed, Vize, Miller, et al. (2020) explicitly suggested that, on the construct level, low Big Five Agreeableness is equivalent to the common core of aversive traits, D.

Despite the obvious conceptual overlap between D and low Big Five Agreeableness, the two have been dissociated both theoretically and empirically (Moshagen, Zettler, Horsten, et al., 2020). In particular, a broad measure of Agreeableness—modeled as the commonalities between five well-established Agreeableness scales (i.e., Big Five Aspects Scale [BFAS], Big Five Inventory 2 [BFI2], International Personality Item Pool Big-Five Factor Markers [IPIP-50], NEO Five-Factor Inventory [NEO-FFI], and HEXACO-Agreeableness vs. Anger[1])—shared only about 40% variance with a broad measure of D (Moshagen, Zettler, Horsten, et al., 2020), thus suggesting that the two constructs overlap substantially, but are not identical. Moreover, the two constructs explained unique variance components (above each other) in a host of deductively derived criteria. Specifically, D predicted incremental variance over Agreeableness (mean $\Delta R^2 = 0.13$—a medium-sized effect according to Cohen, 1988) in behavioral dishonesty, competitive and dangerous worldviews, guilt proneness, internet trolling, and stereotypical sexualized behaviors, whereas Agreeableness predicted incremental variance beyond D in empathy ($\Delta R^2 = 0.39$ and thus above a large effect size; Moshagen, Zettler, Horsten, et al., 2020). These results imply that D and Agreeableness are nomologically and functionally distinct in the sense that they comprise unique, behaviorally relevant trait variance (Moshagen, Zettler, Horsten, et al., 2020).

Questioning the validity of this conclusion, Vize, Miller, et al. (2020) argued that Moshagen, Zettler, Horsten et al.'s (2020) "coverage of AG [Agreeableness] may have missed important aspects, particularly related to modesty (Hex-HH) [HEXACO Honesty-Humility] and straightforwardness (NEO-PI-R)" (p. 597) and may thus have

underestimated the similarity of D and Agreeableness. In line with this argument, Vize, Miller, et al. (2020) noted that the overlap between Agreeableness and D is considerably higher when a particularly broad operationalization of Agreeableness, which we denote AG+ in what follows, is used.[2] Moreover, they found that D did not predict incremental variance beyond AG+ in (self-reported) reactive/proactive aggression and crime and analogous behavior. Even though these are only two criteria, the findings appear to support the notion by Vize, Miller, et al. (2020) that D and Agreeableness are equivalent constructs.

However, to avoid jingle-fallacies, it is necessary to clarify that AG+ is not Agreeableness in the sense of a largely orthogonal dimension within the Big Five framework. Specifically, AG+ is represented by an item set compiled of 104 items from diverse scales (Crowe et al., 2018), including Big Five Agreeableness measures (BFAS, BFI, Faceted Inventory of the Five-Factor Model [FI-FFM], IPIP NEO Personality Inventory Revised [IPIP-NEO-PI-R]), items measuring HEXACO Agreeableness vs. Anger (which is not equivalent to Big Five Agreeableness; Ashton et al., 2014; Thielmann et al., 2021; Endnote 1), items measuring HEXACO Honesty-Humility, and items measuring the interstitial Altruism facet of the HEXACO P-IR (a blend of HEXACO Honesty-Humility, Emotionality, and Agreeableness vs. Anger, representing a "tendency to be sympathetic and soft-hearted toward others"; Ashton et al., 2014, p. 142). In turn, the inclusion of content clearly pertaining to basic traits other than Big Five Agreeableness is arguably the reason why AG+ and the remaining Big Five dimensions are associated considerably more strongly than what is normally found (median $r = .35$ as compared to a median of $\bar{r} = .20$ in the meta-analysis by Park et al., 2020), with a particularly striking correlation of $r = .57$ with Conscientiousness. As such, AG+ cannot be included in a model of basic personality structure involving approximately orthogonal dimensions (Goldberg, 1992; Saucier, 2002) and thus does not represent Big Five Agreeableness. It may well represent some variant of a higher-order "stability" factor comprising aspects of Big Five Agreeableness and Conscientiousness (e.g., DeYoung, 2006), but this, too, logically implies that AG+ cannot be Big Five Agreeableness (Hilbig et al., 2021).

Aside from the fact that AG+ does not represent Agreeableness as conceptualized in models of basic personality structure (Big Five, FFM, HEXACO), the findings of Vize, Miller, et al. (2020) do indicate that a construct subsuming content from several basic personality dimensions, such as AG+, may closely approximate D. In fact, this is fully compatible with the previous conclusion that D, at least to a notable extent, can be understood as "a blend of basic traits" (Moshagen et al., 2018, p. 682), especially HEXACO Honesty-Humility, Agreeableness,

and Conscientiousness. The remaining question, then, is whether the particular blend of Agreeableness-related traits put forward by Crowe et al. (2018) and used in Vize, Miller, et al. (2020), AG+, is indeed equivalent to D both theoretically and empirically.

Ultimately, it is impossible to prove that two constructs are equivalent and relate similarly to *all* conceivable criteria. Nonetheless, it is possible to take a falsificationist approach and theoretically derive criteria that are a priori particularly likely to yield *dis*similarity. Then, if the empirical evidence fails to support the hypothesis of dissimilarity, one may retain the view that the two constructs are functionally equivalent (Gonzalez et al., 2020). With regard to the case at hand, this means that if D and AG+ are non-equivalent, it must be possible to theoretically derive aspects and, ultimately, criteria on which the two can be shown to differ to a non-trivial extent.

In selecting such criteria that may differentiate between D and AG+, we relied on evidence and notions on the conceptual differences between D and Big Five Agreeableness (which constitutes the largest part of AG+). Moshagen, Zettler, Horsten, et al. (2020) pointed out that, unlike D, Agreeableness—as represented by common verbal definitions (e.g., Graziano & Tobin, 2009) and adjective lists (Goldberg, 1992; John et al., 2008)—does not capture the willingness to impose disutility on others even at some cost, nor individual differences in social cognition, as broadly as D. In line with this reasoning, D has been shown to explain notable variance beyond Agreeableness—operationalized as the general factor across four common Big Five Agreeableness measures (BFAS, BFI2, IPIP-50, NEO-FFI)—in 11 different justifying beliefs (mean $\Delta R^2 = .10$; Hilbig et al., 2022) and six antagonistic traits (including Grandiosity and Suspiciousness, as well as Deceitfulness and Manipulativeness) from the DSM-5 model (mean $\Delta R^2 = .07$; Scholz et al., 2022). In conclusion, Agreeableness does not comparably subsume the defining aspects of D that aversive tendencies are "accompanied by beliefs that serve as justifications" (Moshagen et al., 2018, p. 657) and involve inflicting disutility on others.

Moreover, D and basic personality dimensions (which are blended in AG+) conceptually differ in that D was deductively derived to represent the "aversive essence" of aversive traits. D therefore represents (only) aversive content whereas non-aversive aspects of these traits (e.g., vulnerability in Narcissism Bader et al., 2022) are, by definition, beyond the scope of D. By contrast, the basic personality dimensions blended into AG+ are inductively derived from lexical studies and therefore necessarily summarize whatever characteristics tend to co-occur—independent of whether these characteristics are aversive

or not. As a consequence, Agreeableness typically also subsumes affable behaviors towards others, overlapping with Extraversion in sharing affiliation (i.e., "enjoying and valuing close interpersonal bonds and being warm and affectionate"; Depue & Morrone-Strupinsky, 2005, p. 314; see also DeYoung et al., 2013). In leaning heavily on Big Five Agreeableness and in borrowing items from the Altruism and Sympathy facets of the IPIP-NEO (e.g., "Make people feel welcome", "Am not interested in other people's problems"), AG+ also subsumes these affiliative tendencies. By contrast, (lack of) affiliation is not ethically or socially aversive per se and thus theoretically beyond the scope of D.

Given the above, a strict test of the (non-)equivalence of D and AG+ ought to involve those aspects subsumed more strongly in D than in AG+—that is, criteria representing the tendency to inflict disutility on others as well as criteria representing justifying beliefs—and, vice versa, aspects subsumed more strongly in AG+ than in D—that is, criteria representing affiliative[3] tendencies. To this end, the following pre-registered study involved nine corresponding criteria, listed in Table 1: Competitive Jungle Social World View, Normlessness, and Social Dominance Orientation, representing *justifying beliefs*, and Pathological Selfishness and Exploitativeness, representing *inflicting disutility on others*, were expected to have incremental variance explained by D above AG+. Extraversion, Withdrawal, (fast) Life History Strategy, and Horizontal Collectivism, representing *affiliative tendencies*, were, in turn, expected to have incremental variance explained by AG+ above D.

2 | METHOD

The study and analysis plan were preregistered (https://aspredicted.org/uy9ms.pdf) before the start of data collection. Raw data, and analysis scripts are available on the Open Science Framework (OSF; https://osf.io/83sae/). The study was run based on approval by the university's local ethics committee.

2.1 | Measures

The study was conducted in Germany, hence German translations of all measures were used (measures without an existing German translation prior to the study, i.e., FI-FFM and IPIP-NEO-PI-R, were translated under our coordination via the translation-retranslation method; Brislin, 1970; see OSF). D was assessed using the D70, a 70-item scale specifically designed to measure the latent factor underlying all aversive traits (Moshagen, Zettler, &

TABLE 1 Overview of criterion measures and corresponding inventories.

Category	Criterion	Pre-registered hypothesis	Scale (original source)	Number of items	Definition	Sample item
Justifying beliefs	Competitive Jungle Social World View	$\Delta R^2_{\mathrm{D}} > 0$	Two World View Scale (Sibley & Duckitt, 2009)	6	"belief that the social world is a dangerous and threatening place in which good, decent people's values and way of life are threatened by bad people" (Duckitt et al., 2002, p. 78)	"It's a dog-eat-dog world where you have to be ruthless at times."
	Normlessness[a]	$\Delta R^2_{\mathrm{D}} > 0$	Normlessness Scale (Kohn & Schooler, 1983)	4	"the belief that socially unapproved behaviors are required to achieve certain goals" (Ulleberg & Rundmo, 2003, p. 430)	"If something works, it does not matter if it is right or wrong."
	Social Dominance Orientation[a]	$\Delta R^2_{\mathrm{D}} > 0$	SDO₇ (Ho et al., 2015)	8	"Individual differences in the preference for group based hierarchy and inequality" (Ho et al., 2015, p. 1003)	"An ideal society requires some groups to be on top and others to be on the bottom."
Disutility on others	Pathological Selfishness	$\Delta R^2_{\mathrm{D}} > 0$	Selfishness Questionnaire (Raine & Uh, 2018)	8	"An inordinate focus on one's own welfare, regardless of the well-being of others." (Raine & Uh, 2018, p. 503)	"I know I love rewards in life, even if there is a cost to others."
	Exploitativeness[a]	$\Delta R^2_{\mathrm{D}} > 0$	Interpersonal Exploitativeness Scale (Brunell et al., 2013)	6	"the state, condition, quality, or degree of unfairly or cynically using another person or group for profit or advantage" (Brunell et al., 2013, p. 2)	"I'm perfectly willing to profit at the expense of others."
Affiliative tendencies	Extraversion	$\Delta R^2_{\mathrm{AG+}} > 0$	HEXACO-60 (Ashton & Lee, 2009)	10	"[Individual differences in] engagement in social endeavors (such as socializing, leading, or entertaining)" (Ashton & Lee, 2007, p. 156)"	"In social situations, I'm usually the one who makes the first move."
	Withdrawal	$\Delta R^2_{\mathrm{AG+}} > 0$	PID-5 (Krueger et al., 2012)	10	"Reticence in social situations; avoidance of social contacts and activity; lack of initiation of social contact" (APA, 2013, p. 766)	"I keep to myself."
	Life History Strategy	$\Delta R^2_{\mathrm{AG+}} > 0$	Mini-K (Figueredo et al., 2006)	20	"A K-selected life-history strategy allocates an individual's bioenergetic and material resources preferentially to Somatic Effort over Reproductive Effort, and to Parental and Nepotistic Effort over Mating Effort, emphasizing the survival of individual organisms (whether self, offspring, or kin) over the production of new ones" (Figueredo et al., 2006, p. 245)	
	Horizontal Collectivism[a]	$\Delta R^2_{\mathrm{AG+}} > 0$	Individualism and Collectivism Scale-Revised (Singelis et al., 1995)	8	"the individual sees the self as an aspect of an in-group [...] all of whom are extremely similar to each other." (Singelis et al., 1995, p. 244)	"My happiness depends very much on the happiness of those around me."

[a]An ad-hoc translation was used, see OSF.

Hilbig, 2020; Bader, Horsten, et al., 2021). AG+ was assessed via the 104 items identified by Crowe et al. (2018) and used by Vize, Miller, et al. (2020). An overview of the nine criterion measures (along with the corresponding hypotheses) is provided in Table 1. The Withdrawal items were rated on a 4-point Likert scale (1 = "very false" to 4 = "very true"). All other items were rated on a 5-point Likert scale (1 = "strongly disagree" to 5 = "strongly agree").

2.2 | Procedure & participants

Participants were recruited and compensated by a professional panel provider. The study comprised two measurement occasions. At T1, we assessed demographics (including sex and age), the D70, and AG+. At T2 (on average 14 days after T1), we again assessed sex and age, along with the nine criterion measures (see Table 1): Competitive Jungle Social World View, Normlessness, Social Dominance Orientation, Pathological Selfishness, Exploitativeness, HEXACO Extraversion, PID-5 Withdrawal, Mini-K, and Horizontal Collectivism. Within each measurement occasion, the order of scales was randomized and two attention checks were embedded within the scales (e.g., "Please select "Agree". This serves to check your attention."). Participants gave informed consent prior to each measurement occasion and they were fully debriefed about the study background at the end of T2.

In order to be able to model D and AG+ in a bifactor structure, we aimed at a final sample of approximately $N = 1,000$. Anticipating a certain dropout between T1 and T2, a total of 1,331 participants completed T1. In line with our preregistered exclusion criteria, 105 participants were excluded for failing at least one attention check and an additional 70 participants were excluded for suspected inattentive response behavior (based on response times of less than 2 seconds per item on average in more than 50% of the questionnaires and/or no variation in responses, i.e., $SD = 0$, on those scales having at least 25% reverse-keyed items[4]). Thus, 1,156 participants were invited to T2, of whom 940 completed the survey. Of these, we had to exclude one participant for failing at least one attention check, and 12 participants for suspected inattentive response behavior. Additionally, we conservatively excluded 16 participants whose demographics did not match across the two measurement occasions (i.e., they reported to be younger or more than one year older at T2 than at T1 or indicated a different sex at T2 than at T1). Thus, a total of $N = 911$ participants (49% female), aged between 18 and 65 ($M = 40.6$, $SD = 13.0$) years, successfully completed both measurement occasions.

2.3 | Analysis

We tested all hypotheses by estimating confirmatory factor models with the *lavaan* package (Rosseel et al., 2019) in R (R Core Team, 2020). Non-normality in the data was accounted for by maximum likelihood estimation with robust standard errors and scaled test statistics (as provided by the *lavaan* package when specifying "MLR" estimation; Yuan & Bentler, 2000; see also Savalei & Rosseel, 2022). Assuming data are missing at random, we addressed missing cases at the second measurement occasion by employing full information maximum likelihood estimation. Due to the high power of the chi-square model test (Moshagen & Erdfelder, 2016),[5] we primarily considered the robust root mean square error of approximation (RMSEA) and the standardized root mean residual (SRMR) to assess model fit. RMSEA around .05 and SRMR around .06 are commonly considered to be indicative of adequate model fit (Browne & Cudeck, 1992). For the sake of transparency, we also report the robust comparative fit index (CFI; Brosseau-Liard & Savalei, 2014). However, given its stronger dependency on loading magnitude than on model misfit (Moshagen & Auerswald, 2018) and its limited value in evaluating absolute model fit (van Laar & Braeken, 2022), we did not rely on this index to assess model fit.

The D70 was modeled specifying a bifactor model, such that all items loaded both on a general factor and on one of five themes (Bader, Hartung, et al., 2021).[6] The general factor in a bifactor model captures the shared variance among all items. In this case, it captures the aversive content of the indicators and thus represents the underlying aversive disposition, D. The themes capture the remaining shared variance between subsets of items, i.e., those aspects that are beyond the general disposition but characterize an individual's specific pattern of aversive attributes in more detail (Bader, Hartung, et al., 2021). Given that specific factors capture what is beyond the common core (and thus D), they are not relevant when investigating the equivalence of D and AG+. The general factor and specific factors were constrained to mutual orthogonality and identified by setting one unstandardized loading each to 1. According to conventional guidelines (Browne & Cudeck, 1992), this model structure fit the data well, $\chi^2(2,275) = 6,590$, $p < .001$, RMSEA = .045, 90%CI [.044, .047], SRMR = .051, CFI = .799.[7] Importantly, there was a very strong general factor (representing D) that accounted for the majority of reliable variance in the themes (.60 ≤ ECV ≤ .82, median .74). Consequently, the specific factors exhibited very low reliabilities (.07 ≤ ω_{HS} ≤ .21, median .18), further supporting that they should not be interpreted substantively (Sellbom & Tellegen, 2019).

Analogously, AG+ was modeled specifying a bifactor structure with five themes representing the five factors extracted by Crowe et al. (2018). This model also fit the data well, $\chi^2(5,148) = 15,020$, $p < .001$, RMSEA = .044, 90%CI [.043, .045], SRMR = .061, CFI = .741.[8] In contrast to D, the general factor representing AG+ was a less strong (.31 ≤ ECV ≤ .83, median .45). In other words, AG+ captured less variance shared by all items and, consequently, the specific factors exhibited higher reliabilities than the D specific factors (.04 ≤ ω_H ≤ .63, median .46). To ensure that results are not dependent on the choice of bifactor models for D and AG+, we replicated all analyses using single-factor specifications for D and AG+, respectively, which yielded essentially equivalent conclusions.[9]

To test our hypotheses, we then ran separate sequential regression analyses for each criterion. That is, we first regressed each criterion on D and AG+, respectively, and then on both simultaneously to calculate the incremental portion of variance (ΔR^2) explained by either AG+ or D, respectively. To gauge the strength of evidence (beyond the effect size ΔR^2), we considered normalized evidence ratios (ERs) computed from weighted BICs (Wagenmakers & Farrell, 2004; Wu et al., 2020). ERs quantify the support in favor of the less parsimonious model (in this case the model in which the criterion is regressed on both predictors) and range from 0 (no evidence for the less parsimonious model, i.e., the second predictor's contribution to the explained variance is negligible) to 1 (perfect evidence for the less parsimonious model, i.e., the second predictor contributes meaningfully to the explained incremental variance).

3 | RESULTS

Descriptive statistics (of observed scales), reliabilities of the factors, and zero-order inter-correlations of latent factors are summarized in Table 2. The latent correlation between D and AG+ ($r = -.59$) implied 35% of shared variance and was significantly different from unity, $\Delta\chi^2(1) = 5.56$, $p = .02$, ER > .999.

As can be seen in the upper part of Table 3, and supporting the hypotheses that justifying beliefs and inflicting disutility on others are more strongly related to D than to AG+, D explained incremental variance over AG+ for all hypothesized outcomes (i.e., for Competitive Jungle World View, Normlessness, Social Dominance Orientation, Pathological Selfishness, and Exploitativeness). Overall, the proportion of incremental variance explained in these criteria by D was in the range of medium-sized to large effects (Cohen, 1988) with a median $\Delta R^2 = .30$.

Moreover, as presented in the lower part of Table 3 and supporting the hypothesis that affiliative tendencies are

TABLE 2 Descriptive statistics of observed scales, factor reliabilities and latent correlations among all measures.

	Variable	n	M	SD	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
1.	D	1,102	2.1	0.5	.80										
2.	AG+	1,140	3.6	0.4	−.58	.74									
3.	Competitive Jungle Social World View	895	2.1	0.6	.85	−.54	.74								
4.	Normlessness	910	3.0	0.6	.69	−.44	.67	.60							
5.	Social Dominance Orientation	889	2.2	0.7	.55	−.36	.71	.35	.81						
6.	Pathological Selfishness	910	2.0	0.8	.72	−.62	.72	.62	.48	.88					
7.	Exploitativeness	899	1.8	0.7	.74	−.55	.78	.64	.51	.83	.85				
8.	Extraversion	915	3.3	0.6	−.15	.17	−.15	−.10*	−.06*	−.12	−.17	.82			
9.	Withdrawal	903	2.0	0.7	.30	−.26	.29	.21	.21	.27	.35	−.60	.93		
10.	Life History Strategy	906	3.6	0.4	−.22	.23	−.28	−.16	−.20	−.18	−.30	.59	−.59	.74	
11.	Horizontal Collectivism	913	3.7	0.6	−.31	.41	−.36	−.31	−.27	−.33	−.44	.62	−.65	.72	.80

Note: D = Dark Factor of Personality; AG+ = Broad blend of Big Five Agreeableness as per Crowe et al. (2018). *n* = number of observations per scale. Reliabilities (hierarchical ω for D and AG+; unidimensional ω for all criterion scales) in italics on the diagonal.

*$p > .05$.

TABLE 3 Latent regression results predicting the outcomes by AG+ and D.

Category	Criterion	$R^2_{\text{D, AG+}}$	ΔR^2_{D}	$\Delta R^2_{\text{AG+}}$	Unique D	Unique AG+	ER
Justifying beliefs	Competitive Jungle Social World View	.74	.45	.05	61%	6%	>.99
	Normlessness	.52	.33	.04	63%	7%	>.99
	Social Dominance Orientation	.31	.17	.03	56%	11%	>.99
Disutility on others	Pathological Selfishness	.63	.24	.10	39%	15%	>.99
	Exploitativeness	.60	.30	.08	50%	13%	>.99
Affiliative tendencies	Extraversion	.05	.02	.03	42%	58%	.99
	Withdrawal	.10	.04	.02	34%	18%	.90
	Life History Strategy	.07	.02	.03	25%	49%	.89
	Horizontal Collectivism	.23	.05	.14	21%	61%	>.99

Note: N = 1,156. D = Dark Factor of Personality; AG+ = Broad blend of Big Five Agreeableness as per Crowe et al. (2018). $\Delta R^2_{\text{D, AG+}}$: variance explained in the full model. ΔR^2_{D}: increase in R^2 after adding D to the model. $\Delta R^2_{\text{AG+}}$: increase in R^2 after adding AG+ to the model. Unique D/Unique AG+: relation of each construct's unique contributions relative to the total variance explained in the criteria. ER: evidence in favor of the full model over the model not including the predictor hypothesized to explain incremental variance over the other. For justifying beliefs and disutility on others, the hypothesis was $\Delta R^2_{\text{D}} > 0$, for affiliative tendencies the hypothesis was $\Delta R^2_{\text{AG+}} > 0$.

more strongly related to AG+ than to D, AG+ predicted incremental variance over D for all hypothesized outcomes in this domain (i.e., Extraversion, Withdrawal, Life History Strategy, and Horizontal Collectivism). However, the proportion of incremental variance explained in these criteria by AG+ was only in the range of small to medium-sized effects for all criteria, with a median of $\Delta R^2 = .03$. Despite varying proportions of (absolute and) incremental variances explained, evidence ratios consistently (and typically very strongly) supported all hypotheses.

In sum, the results show that D and AG+ account for different variance components in all outcome criteria and, more specifically, each construct accounts for incremental variance in exactly those criteria that it ought to based on a priori theoretical considerations.[10]

As an additional exploratory analysis, we followed the recommendation of an anonymous reviewer and applied Multidimensional Scaling (MDS) to the D70 and AG+ items. MDS positions the items in an *n*-dimensional space based on their distances, which allows evaluating which items are closer to and further away from each other, respectively. The analysis and results are described in more detail in the additional materials on the OSF. In brief, we first recoded the items so that all pointed in the same direction (to be interpreted as high D). Levels of stress (a badness-of-fit measure which indicates how many axes are required to minimize the dissimilarity of the distances between the items in the map from their distances in the underlying data; Kruskal, 1964) suggested to consider three axes contrasting (1) low empathy versus dominance, (2) vengefulness versus low humility, and (3) utility at or from the cost of others versus disagreeableness. Whereas the centroid of the AG+ items was in the octant (−, +, −) spanned by the three axes, the centroid of the D items

was in the exact opposite octant (+, −, +). Notably, these octants were exclusively occupied by AG+ or D70 items, respectively. The average Euclidean distance was 0.26 among the D70 items, 0.18 among the AG+ items, and 0.45 across the two constructs. Thus, the items within each construct were notably closer (and thus similar) to each other than the items across constructs. In sum, the MDS results corroborate our conclusion that D and AG+ are not equivalent.

4 | DISCUSSION

Recent research has worked towards specifying and determining the most suitable representation of the common disposition underlying aversive personality traits. This quest has spurred some controversy whether Big Five Agreeableness, per se, represents this common disposition. In a nutshell, Moshagen, Zettler, Horsten, et al. (2020) dissociated the Dark Factor of Personality (D) and low Agreeableness theoretically and empirically (see also Hilbig et al., 2021; Scholz et al., 2022), whereas Vize and Lynam (2020) and Vize, Miller, et al. (2020) argued that the low pole of Agreeableness is essentially equivalent to D and that Moshagen et al.'s dissociations were problematic because their operationalization of Agreeableness was incomplete. As a remedy, Vize, Miller, et al. (2020) pointed to a particularly broad operationalization of "Agreeableness", which we herein term AG+. AG+ comprises content from, and is substantially related to, other Big Five and HEXACO personality dimensions beyond traditional Agreeableness in the sense of a largely orthogonal dimension within the Big Five (which is why the label "Agreeableness" is conducive to the jingle-fallacy and

thus why we prefer to label it "AG+"; Hilbig et al., 2021). Nevertheless, AG+ may be an adequate approximation of the common core of aversive traits, that is, D.

Accordingly, we herein tested whether AG+ is essentially equivalent to D, sketching theoretical differences between the two constructs and thereby deriving criteria in which the constructs ought to account for unique variance components. Specifically, we hypothesized that D should account for unique variance in attitudes and beliefs that may serve as justifications for malevolent behaviors as well as the tendency to inflict disutility on others, whereas AG+ should account for unique variance in affiliative tendencies.

In a preregistered study, we found that D and AG+ shared only about 35% of variance. At first sight, this is incompatible with the 80% reported by Vize, Miller, et al. (2020). Importantly, though, the latter was based on a modeling approach yielding poor model fit and decreased to about 40% when estimating D and AG+ (more appropriately) as bifactor structures—thus mirroring the approach reported herein. As such, given equivalent modeling approaches, the findings are largely comparable. In substantive terms, although this magnitude of shared variance indicates a sizable overlap, it nonetheless speaks against unity of the two constructs. This conclusion was further corroborated by exploratory MDS analyses showing that each constructs' centroid, around which the respective items cluster, are located in different octants in the MDS space.

More importantly, each construct captured variance beyond the other in every one of the nine criteria tested herein. Specifically, as hypothesized, D accounted for unique variance in justifying beliefs, such as that one has to be ruthless to survive (Competitive Jungle World View) or that one is superior to others (Social Dominance Orientation), and in the willingness to inflict disutility on others, as expressed in the disregard for others' needs (Pathological Selfishness) or the active exploitation of others (Exploitativeness) representing notable effect sizes throughout (between approximately 20% and 50% of additional variance explained beyond AG+). In turn, and again as hypothesized, AG+ accounted for unique variance in affiliative tendencies, such as the interest in (being with) people (Extraversion, low Withdrawal), forming emotional bonds (Life History Strategy) and feeling connected with others (Horizontal Collectivism). These conclusions were robust across alternative modeling approaches. In other words, D subsumed a broader range of aversive content than AG+, which, in turn, accounted for a somewhat broader range of non-aversive content related to affiliation. Notably, the portion of unique variance explained by D in aversive criteria (i.e., justifying beliefs and inflicting disutility on others) was much larger than the portion of

unique variance explained by AG+ in affiliative criteria (median $\Delta R^2 = .28$ and .02, respectively). In substantive terms, D is much more strongly linked to justifying beliefs and inflicting disutility on others (than AG+), but AG+ is only marginally more saturated with affiliative tendencies than D.

The present results complement previous research in which D was dissociated from basic personality dimensions that had been proposed to represent the common core of aversive traits, especially Big Five Agreeableness and HEXACO Honesty-Humility (Hilbig et al., 2020; Horsten et al., 2021; Moshagen, Zettler, Horsten, et al., 2020; Scholz et al., 2022). Hence, neither Agreeableness, nor Honesty-Humility—as defined within their respective models of personality structure—nor their (extended) combination AG+ are equivalent to D.

A limitation of our study is that data were collected during a phase of the Covid-19 pandemic when it was generally imperative to restrict in-person meetings with others. This might particularly have affected participants' responses on the Withdrawal scale used (e.g., "I prefer not to get too close to people", or "I avoid social events"), potentially limiting the interpretability of the respective scale score. Given that correlations are invariant to additive shifts, however, this would only have affected the scale score, but arguably not the magnitude of correlations with D and AG+. A further limitation is that we did not include a criterion representing consequential behavior. Although our results show that D and AG+ are non-equivalent with respect to self-report measures, replicating such differences on the level of actual behavior would seem prudent.

5 | CONCLUSION

Although D and AG+ show noticeable overlap, the two constructs are not isomorphic. Ultimately, enriching Big Five Agreeableness with content from HEXACO Honesty-Humility, Agreeableness, and Altruism to become AG+ (Crowe et al., 2018) results in a construct incompatible with the Big Five framework (Hilbig et al., 2021), which—despite its added breadth—is non-equivalent to the common core of aversive traits, D. Although there may well be other advantages of AG+ as a construct and/or operationalization, it does not represent D.

collection, manuscript writing: David Scholz. Idea, conceptualization, data analysis, manuscript writing: Morten Moshagen.

CONFLICT OF INTEREST STATEMENT
The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS STATEMENT
The study reported herein was based on the approval of the local ethics committee.

ORCID
Luisa K. Horsten https://orcid.org/0000-0002-3023-8808
Isabel Thielmann https://orcid.org/0000-0002-9071-5709
Morten Moshagen https://orcid.org/0000-0002-2929-7288
Ingo Zettler https://orcid.org/0000-0001-6140-7160
David Scholz https://orcid.org/0000-0003-1491-5247
Benjamin E. Hilbig https://orcid.org/0000-0002-1470-1882

ENDNOTES
[1] Note that HEXACO Agreeableness vs. Anger and Big Five Agreeableness are rotated variants of each other (Ashton et al., 2014). Consequently, a recent meta-analysis found a correlation between these two of $\rho = .69$ only (Thielmann et al., 2021), suggesting that the two constructs are not equivalent. In Moshagen, Zettler, Horsten, et al. (2020), HEXACO Agreeableness vs. Anger was nonetheless included to represent Agreeableness as broadly as possible.

[2] Vize, Miller, et al. (2020) reported around 80% shared variance when estimating D as a single factor from the observed scores on the 12 scales from which the D70 was compiled and, analogously, AG+ from the observed scores on the respective 20 subscales. However, in the presence of a covariate, a bifactor structure generally yields the most accurate results as it is robust to secondary loadings and covariances between parts making up the hierarchical construct and the covariate (Moshagen, 2023). When relying on the bifactor approach to model D (as is recommended, Bader, Hartung, et al., 2021) and, analogously, modeling AG+ as a bifactor, the shared variance of AG+ and D notably decreases to around 40% in Vize et al.'s dataset, thus fully conforming to prior findings.

[3] In our preregistration, we divided these tendencies into "interest in (being with) people" and "feelings of connectedness with others and an emotional warm glow", but for simplicity will subsume them as *affiliative tendencies* herein.

[4] If any of these criteria were met at the level of individual scales, responses on the respective scale were set to NA, thus treated as missings.

[5] According to a post-hoc power analysis for a global hypothesis test (Jobst et al., 2021) of both bi-factor models—i.e., of D ($df = 2,275$) and of AG+ ($df = 5,148$)—the power to detect even negligible model misspecification (i.e., RMSEA = .010) at an α-level of .05 and given the current N was extremely high with more than 99.9% in both cases.

[6] A bifactor model is preferred to the closely related higher-order model in the current investigation both for substantive and methodological reasons. For one, the bifactor model most closely mirrors the theoretical conceptualization of the common core of aversive traits because it assumes that the general factor directly accounts for the variance shared among all indicators. By contrast, the higher-order model assumes that the higher-order factor only indirectly explains the items' shared variance through the lower-order factors. Additionally, the bifactor model provides unbiased estimates of the correlations between the general factor and covariates (Moshagen, 2023). For a more detailed reasoning, see Bader et al. (2021) and Moshagen et al. (2018).

[7] Additionally, we ran our analyses with modified models in which those parameters yielding the largest modification indices were freed. This alternative approach led to the same conclusions. Corresponding analysis scripts and detailed results are provided in the additional materials on the OSF.

[8] Note that we preregistered to model AG+ according to Vize, Miller, et al. (2020) who estimated AG+ from 20 manifest facet scores representing the facets of the different measures from which the item set was compiled. In doing so, they deviated from their preregistered analysis plan which was to estimate AG+ from the five factors extracted by Crowe et al. (2018). Given that this would have actually been the more appropriate modeling strategy (Moshagen, 2023), for which Vize, Miller, et al. (2020) also reported the results on the OSF, we herein modeled AG+ following this latter approach. As the item mapping to the factors was not reported by Crowe et al. (2018), we extracted five factors through a principal factor exploratory factor analysis with promax rotation (thereby reproducing the approach chosen by Vize, Miller et al., 2020). Content-wise, the emerging factors aligned with the factors identified by Crowe et al. (2018). Nevertheless, we verified our results replicating the approach Vize, Miller, et al. (2020) reported in their manuscript (see Footnote 10).

[9] Model fits of the single factor models, as well as the corresponding scripts and analyses, can be found in the additional materials on the OSF. In summary, as indicated both by likelihood-ratio tests and evidence ratios, the bifactor model yielded the best fit to the data, both for D and for AG+ (see additional materials on the OSF).

[10] Modeling AG+ from the manifest scores on the 20 Agreeableness (facet) scales (see Vize, Miller, et al., 2020) yielded poor fit to the data, $\chi^2(170) = 3,617$, $p < .001$, RMSEA = .143, 90% CI [.139, .147], SRMR = .090. The latent correlation between D and AG+ was $r = -.78$, which was significantly smaller than unity ($\Delta\chi^2(1) = 14.52$, $p < .001$, ER > .999). The pattern of results was largely consistent with the one presented herein, that is, both D and AG+ predicted incremental variance beyond the other in the hypothesized criteria. As such, the analytical approach reported by Vize, Miller, et al. (2020) confirms the present conclusions. The corresponding analysis script and results are provided in the additional materials on the OSF.

REFERENCES

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

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

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

Ashton, M. C., Lee, K., & De Vries, R. E. (2014). The HEXACO honesty-humility, agreeableness, and emotionality factors: A review of research and theory. *Personality and Social Psychology Review*, *18*(2), 139–152. https://doi.org/10.1177/1088868314523838

Bader, M., Hartung, J., Hilbig, B. E., Zettler, I., Moshagen, M., & Wilhelm, O. (2021). Themes of the dark core of personality. *Psychological Assessment*, *37*, 511–525. https://doi.org/10.1037/pas0001006

Bader, M., Hilbig, B. E., Zettler, I., & Moshagen, M. (2022). Rethinking aversive personality: Decomposing the dark triad traits into their common core and unique flavors. *Journal of Personality*. Advance online publication. https://doi.org/10.1111/jopy.12785

Bader, M., Horsten, L. K., Hilbig, B. E., Zettler, I. , & Moshagen, M. (2021). Measuring the dark core of personality in German: Psychometric properties, measurement invariance, predictive validity, and self-other agreement. *Journal of Personality Assessment*, *104*(5), 660–673. https://doi.org/10.1080/00223891.2021.1984931

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, *1*(3), 185–216. https://doi.org/10.1177/135910457000100301

Brosseau-Liard, P. E., & Savalei, V. (2014). Adjusting incremental fit indices for nonnormality. *Multivariate Behavioral Research*, *49*(5), 460–470. https://doi.org/10.1080/00273171.2014.933697

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

Brunell, A., Davis, M., Schley, D., Eng, A., van Dulmen, M., Wester, K., & Flannery, D. (2013). A new measure of interpersonal exploitativeness. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00299

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). L. Erlbaum Associates.

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

Depue, R. A., & Morrone-Strupinsky, J. V. (2005). A neurobehavioral model of affiliative bonding: Implications for conceptualizing a human trait of affiliation. *Behavioral and Brain Sciences*, *28*(3), 313–350. https://doi.org/10.1017/S0140525X05000063

DeYoung, C. G. (2006). Higher-order factors of the Big Five in a multi-informant sample. *Journal of Personality and Social Psychology*, *91*(6), 1138–1151.

DeYoung, C. G., Weisberg, Y. J., Quilty, L. C., & Peterson, J. B. (2013). Unifying the aspects of the Big Five, the interpersonal Circumplex, and trait affiliation: Big Five and IPC. *Journal of Personality*, *81*(5), 465–475. https://doi.org/10.1111/jopy.12020

Duckitt, J., Wagner, C., du Plessis, I., & Birum, I. (2002). The psychological bases of ideology and prejudice: Testing a dual process model. *Journal of Personality and Social Psychology*, *83*(1), 75–93. https://doi.org/10.1037/0022-3514.83.1.75

Egan, V., Hughes, N., & Palmer, E. J. (2015). Moral disengagement, the dark triad, and unethical consumer attitudes. *Personality and Individual Differences*, *76*, 123–128. https://doi.org/10.1016/j.paid.2014.11.054

Figueredo, A., Vasquez, G., Brumbach, B., Schneider, S., Sefcek, J., Tal, I., Hill, D., Wenner, C., & Jacobs, W. (2006). Consilience and life history theory: From genes to brain to reproductive strategy. *Developmental Review*, *26*(2), 243–275. https://doi.org/10.1016/j.dr.2006.02.002

Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment*, *4*(1), 26–42. https://doi.org/10.1037/1040-3590.4.1.26

Gonzalez, O., MacKinnon, D. P., & Muniz, F. B. (2020). Extrinsic convergent validity evidence to prevent jingle and jangle fallacies. *Multivariate Behavioral Research*, *56*(1), 1–17. https://doi.org/10.1080/00273171.2019.1707061

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

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

Hilbig, B., Moshagen, M., Thielmann, I., & Zettler, I. (2022). Making rights from wrongs: The crucial role of beliefs and justifications for the expression of aversive personality. *Journal of Experimental Psychology General*. Advance online publication. https://doi.org/10.1037/met0000423

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

Hilbig, B. E., Thielmann, I., Klein , S. A., Moshagen, M., & Zettler, I. (2020). The dark core of personality and socially aversive psychopathology. *Journal of Personality*, *89*(2), 216–227. https://doi.org/10.1111/jopy.12577

Ho, A. K., Sidanius, J., Kteily, N., Sheehy-Skeffington, J., Pratto, F., Henkel, K. E., Foels, R., & Stewart, A. L. (2015). The nature of social dominance orientation: Theorizing and measuring preferences for intergroup inequality using the new SDO7 scale. *Journal of Personality and Social Psychology*, *109*(6), 1003–1028. https://doi.org/10.1037/pspi0000033

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

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

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research*. The Guilford Press.

Jonason, P. K., Zeigler-Hill, V., & Okan, C. (2017). Good v. evil: Predicting sinning with dark personality traits and moral foundations. *Personality and Individual Differences*, *104*, 180–185. https://doi.org/10.1016/j.paid.2016.08.002

Kohn, M., & Schooler, C. (1983). *Work and personality: An inquiry into the impact of social stratification*. Ablex.

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

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, *29*(1), 1–27. https://doi.org/10.1007/BF02289565

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

Moshagen, M. (2023). When a truly positive correlation turns negative: How different approaches to model hierarchically structured constructs affect estimated correlations to covariates. *European Journal of Personality*, *37*(1), 95–112. https://doi.org/10.1177/08902070211050170

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

Moshagen, M., & Erdfelder, E. (2016). A new strategy for testing structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(1), 54–60. https://doi.org/10.1080/10705511.2014.950896

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

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

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

Muris, P., Merckelbach, H., Otgaar, H., & Meijer, E. (2017). The malevolent side of human nature: 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. https://doi.org/10.1177/1745691616666070

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

Park, H. (. H.)., Wiernik, B. M., Oh, I.-S., Gonzalez-Mulé, E., Ones, D. S., & Lee, Y. (2020). Meta-analytic five-factor model personality intercorrelations: Eeny, meeny, miney, Moe, how, which, why, and where to go. *Journal of Applied Psychology*, *105*(12), 1490–1529. https://doi.org/10.1037/apl0000476

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

R Core Team. (2020). *R: A language and environment for statistical computing*. (3.6.3) [Computer software]. R Foundation for Statistical Computing. https://www.r-project.org/

Raine, A., & Uh, S. (2018). The selfishness questionnaire: Egocentric, adaptive, and pathological forms of selfishness. *Journal of Personality Assessment*, *101*(5), 503–514. https://doi.org/10.1080/00223891.2018.1455692

Rosseel, Y., Jorgensen, T. D., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, E., Hallquist, M., Rhemtulla, M., Katsikatsou, M., Barendse, M., & Scharf, F. (2019). *Lavaan: Latent variable analysis* (0.6-5) [computer software]. https://CRAN.R-project.org/package=lavaan

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

Savalei, V., & Rosseel, Y. (2022). Computational options for standard errors and test statistics with incomplete normal and nonnormal data in SEM. *Structural Equation Modeling: A Multidisciplinary Journal*, *29*(2), 163–181. https://doi.org/10.1080/10705511.2021.1877548

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

Schreiber, A., & Marcus, B. (2020). The pace 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

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

Sibley, C. G., & Duckitt, J. (2009). Big-Five personality, social worldviews, and ideological attitudes: Further tests of a dual process cognitive-motivational model. *The Journal of Social Psychology*, *149*(5), 545–561.

Singelis, T. M., Triandis, H. C., Bhawuk, D. P. S., & Gelfand, M. J. (1995). Horizontal and vertical dimensions of individualism and collectivism: A theoretical and measurement refinement. *Cross-Cultural Research*, *29*(3), 240–275. https://doi.org/10.1177/106939719502900302

Stead, R., & Fekken, G. C. (2014). Agreeableness at the core of the dark triad of personality. *Individual Differences Research*, *12*(4-A), 131–141.

Thielmann, I., Moshagen, M., Hilbig, B. E., & Zettler, I. (2021). On the comparability of basic personality models: Meta-analytic correspondence, scope, and orthogonality of the Big Five and HEXACO dimensions. *European Journal of Personality*, *36*(6), 870–900. https://doi.org/10.1177/08902070211026793

Ulleberg, P., & Rundmo, T. (2003). Personality, attitudes and risk perception as predictors of risky driving behaviour among young drivers. *Safety Science*, *41*(5), 427–443. https://doi.org/10.1016/S0925-7535(01)00077-7

van Laar, S., & Braeken, J. (2022). Caught off base: A note on the interpretation of incremental fit indices. *Structural Equation Modeling: A Multidisciplinary Journal*, *29*(6), 935–943. https://doi.org/10.1080/10705511.2022.2050730

Vize, C., & Lynam, D. (2020). On the importance of the assessment and conceptualization of agreeableness: A commentary on "agreeableness and the common core of dark traits are functionally different constructs." *Journal of Research in Personality*, *90*, 104059. https://doi.org/10.1016/j.jrp.2020.104059

Vize, C. E., 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.

Vize, C. E., Miller, J. D., & Lynam, D. R. (2020). Examining the conceptual and empirical distinctiveness of agreeableness and "dark" personality items. *Journal of Personality*, *89*(3), 594–612. https://doi.org/10.1111/jopy.12601

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

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

Yuan, K.-H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology*, *30*, 165–200. https://doi.org/10.1111/0081-1750.00078