Journal of Research in Personality 96 (2022) 104147

Contents lists available at ScienceDirect

Journal of Research in Personality

journal homepage: www.elsevier.com/locate/jrp

Incorporating prosocial vs. antisocial trait content in Big Five measurement: Lessons from the Big Five Inventory-2 (BFI-2)

Jaap J.A. Denissen [a,*], Christopher J. Soto [b], Rinie Geenen [a], Oliver P. John [c], Marcel A. G. van Aken [a]

[a] Utrecht University, the Netherlands
[b] Colby College, United States
[c] University of California, Berkeley, United States

ARTICLE INFO

Keywords:
Agreeableness
Honesty-humility
Dark personality
Prosocial and antisocial behavior
Big Five

ABSTRACT

In three studies, we tried to create a novel scale within the existing Big Five Inventory 2 (BFI-2) structure to measure prosocial vs. antisocial personality traits, like the Dark Triad and honesty-humility. While our new scale converged strongly with dark personality and honesty-humility scales, we failed to establish sufficient discriminant validity vis-a-vis the existing BFI-2 agreeableness domain. Instead, we found that dark traits and honesty-humility were best measured as a facet of agreeableness that correlated strongly with other agreeableness facets and established measures of dark traits and honesty-humility. These findings suggest that honesty-humility and dark personality traits can be measured as opposite facets of a broader agreeableness-antagonism continuum when adopting the BFI-2 (Big Five) domain structure.

1. Introduction

The Big Five trait paradigm —with the domains of extraversion, agreeableness, conscientiousness, neuroticism (or negative emotionality), and openness to experience (or open-mindedness)— has emerged as the most widely used paradigm in personality psychology (John, Naumann, & Soto, 2008). This emergence has been accompanied by the development of personality inventories that provide reliable and valid assessment of the Big Five. For example, the recently published Big Five Inventory–2 (BFI-2; Soto & John, 2017a) uses 60 descriptive phrases to efficiently measure the Big Five domains, as well as 15 facet traits. Although still relatively new, the BFI-2 has become widely used due to its conceptual coherence, brevity, psychometric soundness, and free availability, and has already been translated into dozens of languages (e.g., Denissen, Geenen, Soto, John, & van Aken, 2019; Gardiner, Sauerberger, & Funder, 2019).

Some researcher traditions have also identified and investigated personality traits beyond the Big Five. One major tradition focuses on traits that predict antisocial outcomes, which are often studied in terms of the Dark Triad: machiavellianism, narcissism, and psychopathy (Paulhus & Williams, 2002). Some of these dark traits appear to mix variance associated with different Big Five dimensions, such as low agreeableness and extraversion for narcissism (Paulhus, 2001) or low agreeableness, low conscientiousness, and high neuroticism for the social deviance facet of psychopathy (Lynam et al., 2005). Another tradition is the HEXACO model, which is based on lexical studies that have distinguished a sixth major trait domain, which is labeled honesty-humility. This tradition also rotated the Big Five agreeableness and neuroticism domains (Ashton & Lee, 2007), resulting in revised definitions. Specifically, one rotated dimension combines high levels of agreeableness and low neuroticism (vs. low levels of agreeableness and high neuroticism), yet it rather confusingly retained the label of "Agreeableness" despite this meaningfully altered definition.[1] Furthermore, the other rotated dimension combines high levels of agreeableness and high neuroticism (vs. low levels of agreeableness and low neuroticism), which is labeled as Emotionality to reflect the difference in content from neuroticism. Both honesty-humility and the Dark Triad capture individual differences in prosocial vs. antisocial tendencies, which raises questions about whether and how this content can be represented within the BFI-2. Indeed, the BFI-2 has been criticized for

* Corresponding author at: Department of Developmental Psychology, Utrecht University, Heidelberglaan 1, 3584 CS Utrecht, the Netherlands.
E-mail address: jjadenissen@gmail.com (J.J.A. Denissen).

[1] When discussing underlying dimensions, we will refer to the dimension as a A+/N- blend, while we reserve agreeableness for the Big Five dimension. Only when capitalized, we refer to HEXACO "Agreeableness" as the corresponding scale that measures these blended Big Five domains.

https://doi.org/10.1016/j.jrp.2021.104147
Received 29 March 2021; Received in revised form 8 August 2021; Accepted 21 September 2021
Available online 25 September 2021

not directly assessing honesty-humility or dark personality content (Ashton & Lee, 2019; Ashton, Lee, & Visser, 2019; Vize, Collison, Miller, & Lynam, in press).

At present, researchers and practitioners who find the BFI-2 appealing as a widespread measure of the Big Five but who also wish to investigate individual differences in prosocial vs. antisocial personality traits face a dilemma about which measure (or battery of measures) to use. Ideally, they could use an additional scale that preserves the design and psychometric properties of the BFI-2 while also making sure that the additional content is sufficiently distinct from the original scales. We aimed to do so without changing (e.g., rotating) the existing BFI-2 content because it has already been widely adopted and embedded in the theoretical framework of the Big Five. We therefore conducted the present research to construct and validate an optional, supplemental scale for the existing BFI-2 to measure prosocial vs. antisocial personality traits. We were cautiously optimistic that the effort could succeed because the BFI has been identified as including less H content than other Big Five measures (Howard & Van Zandt, 2020), suggesting there might be some free space to explore. The successful development of such a scale would allow the BFI-2 to serve as a "one-stop shop" for assessing both the Big Five and dark personality traits/honesty-humility. While pursuing this scale development goal, we also hoped to gain insights into how dark traits and honesty-humility are best understood, incorporated, and/or represented within a framework that focuses on conceptually and empirically coherent traits like the Big Five.

2. Conceptualizing prosocial vs. antisocial traits: agreeableness, the Dark Triad, and honesty-humility

There are both similarities and differences in how the Big Five, Dark Triad frameworks, and HEXACO conceptualize individual differences in prosocial and antisocial behaviors. Within the Big Five tradition, definitions of agreeableness emphasize how high levels of this trait domain foster close relationships and promote cooperative behavior (Crowe, Lynam, & Miller, 2018; Denissen & Penke, 2008), whereas low levels underlie antagonistic behavior as well as aggressive and antisocial outcomes (Lynam & Miller, 2019). Supporting this conceptualization, empirical research examining the hierarchical structure of agreeableness indicates that this broad domain comprises 5 facets: compassion (vs. callousness), affability (vs. combativeness), trust (vs. distrust), morality (vs. immorality), and modesty (vs. arrogance) (Crowe et al., 2018). The BFI-2 focuses on the first three of these facets, which it labels Compassion, Respectfulness, and Trust (Soto & John, 2017a). Longer Big Five measures, such as both commercial and public domain versions of the NEO Personality Inventory, assess all five (Costa & McCrae, 2008).[2] Of note, the HEXACO "Agreeableness" facets correlated primarily with the agreeableness facets of trust and affability of Crowe et al. (2018), whereas the HEXACO Honesty-Humility facets correlated strongly with the agreeableness facets morality and modesty. This suggests that higher-order conceptualizations of agreeableness might capture variance underlying both the "Agreeableness" (i.e., A+/N−) and Honesty-Humility scales of the HEXACO model. This has also been pointed out by the authors of the HEXACO model: "researchers who favor perfectly orthogonal factors may prefer to view Honesty as a large and peripheral element of a very broad Agreeableness factor" Ashton & Lee, 2001, p. 333).

The HEXACO model and accompanying HEXACO Personality Inventory were established in a series of lexical and questionnaire-based studies across multiple cultures by Ashton, Lee, and colleagues (Ashton & Lee, 2001; for a review, see Ashton, Lee, & de Vries, 2014). Compared with the Big Five, the HEXACO framework blends agreeableness with neuroticism to exclude compassion while including anger proneness. It furthermore adds a honesty-humility (H) domain, which is most commonly defined and measured in terms of four lower-level facets: sincerity (vs. deceptiveness), modesty (vs. arrogance), greed avoidance (vs. greed), and fairness (vs. immorality) (Lee & Ashton, 2004). Both HEXACO's blend of A+/N- and honesty-humility have been theorized as relevant for explaining individual differences in cooperation, but with an important difference: "High levels of [honesty-humility] represent a tendency to cooperate with another person even when one could successfully exploit that individual, whereas [high levels of "Agreeableness"] represent a tendency to cooperate with another person even when that individual appears to be somewhat exploitive (or, equivalently, not fully cooperative)" (Ashton et al., 2014, p. 144).

Consistent with this conceptualization, honesty-humility has been found to predict prosocial behavior in one-shot dictator games, in which participants can allocate resources (e.g., points or money) to another person without repercussions for low allocations (e.g., Hilbig, Thielmann, Hepp, Klein, & Zettler, 2015). That said, studies have also found that the affability, morality, and modesty facets of Big Five agreeableness predict prosocial behavior in such paradigms (Zhao, Ferguson, & Smillie, 2017). Consistent with this, these same facets of Big Five agreeableness have been found to correlate most strongly with HEXACO honesty-humility (Ashton & Lee, 2005). Thus, the ability of honesty-humility to predict prosocial behavior might overlap with aspects of a broader agreeableness trait domain.

Finally, the Dark Triad consists of three traits thought to capture individual differences in selfish and antisocial behavior: narcissism, machiavellianism, and psychopathy (Paulhus & Williams, 2002). Recent research has found that these dark traits share a strong underlying core, which represents the "basic tendency to maximize one's own utility at the expense of others" (Moshagen, Hilbig, & Zettler, 2018). Conceptually, dark personality traits seem inverse to Big Five agreeableness and (especially) HEXACO honesty-humility, and indeed research has found strong, negative associations between the Dark Triad on the one hand, and agreeableness and honesty-humility on the other (Lee & Ashton, 2005; Muris, Merckelbach, Otgaar, & Meijer, 2017; Vize et al., in press). This research also suggests that the Dark Triad explain various aggressive and norm violating outcomes that are costly to society (Muris et al., 2017). One nuance is that narcissism has an admiration component that taps into self-enhancement and is positively correlated with extraversion and a rivalry component that taps into derogation of others and is negatively correlated with agreeableness (Back et al., 2013). Of these two, the rivalry component is most closely related to dark personality traits (Moshagen, Zettler, & Hilbig, 2020).

3. Measuring prosocial vs. antisocial traits with the BFI-2

As noted above, the BFI-2 directly assesses the three most conceptually and empirically central facets of Big Five agreeableness: compassion, affability, and trust (Crowe et al., 2018; Soto & John, 2017a). This makes the BFI-2 a short but sufficiently extensive measure of the agreeableness trait domain. However, it may also limit the BFI-2's capacity to assess prosocial vs. antisocial traits, such as captured by the other two facets identified by Crowe et al. (2018) that are critical to the Dark Triad framework and included in the HEXACO: morality (vs. immorality) and modesty (vs. arrogance). This conceptual concern is borne out by recent empirical research. For example, a recent analysis demonstrated that BFI-2 Agreeableness accounted for substantial variance among Dark Triad constructs, but less so than did a broader representation of the agreeableness domain or HEXACO honesty-humility (Vize et al., in press). Furthermore, a direct comparison of the BFI-2 and the HEXACO Personality Inventory found that the HEXACO Honesty-Humility scale could not be subsumed by the BFI-2 domain scales and loaded on a different factor than the BFI-2 Agreeableness facets (Ashton

[2] Specifically, compassion maps onto NEO-PI-R Altruism; affability maps onto NEO-PI-R Compliance, trust maps onto NEO-PI-R Trust, morality onto Straightforwardness, and modesty with Modesty=same (the more attitudinal Tender-Mindedness was thus not represented in the facet solution).

et al., 2019).

In order to address the limited representation of honesty-humility vs. dark personality traits in the BFI-2, we conducted the present research pursuing two overarching goals. First, we aimed to develop and validate an optional, supplemental Honesty-Humility scale for the BFI-2. Successful construction of this "BFI-2-H" scale would allow interested researchers to broadly assess moral vs. immoral or antisocial personality traits, while still measuring the most central facets of each Big Five domain as operationalized by the 15 facets that were published by Soto and John (2017a).[3] It would thereby allow the BFI-2 to cover more areas of overlap between the Big Five, Dark Triad, and HEXACO personality models. Second, during the process of developing and evaluating this scale, we also hoped to build on recent research (e.g., Crowe et al., 2018; Vize et al., in press) by gaining further insights about how prosocial vs. antisocial personality traits can be conceptualized within the Big Five framework.

4. Study 1: Effort to create and validate BFI-2-H items

Our main goal for Study 1 was to develop and validate supplemental honesty-humility domain and facet scales for the BFI-2. We wanted these BFI-2-H scales to parallel the standard BFI-2 in terms of their item content and format and yield reliable scores. Our main analysis was subsequently focused on establishing a) whether the H facet scales would be distinguishable from the established BFI-2 facet scales in analyses of multidimensional structure, and b) whether the H facet and domain scales converge with established measures of dark personality traits and honesty-humility.

4.1. Method of Study 1

4.1.1. Procedure and sample

Participants in Study 1 were 800 adolescents and adults who volunteered to complete an online survey in exchange for automatically generated personality feedback. The survey was hosted at http://www.personalitylab.org/, an advertisement-free, non-commercial survey website maintained by the second author. Potential participants could find the survey using online search engines, links from other websites, or word of mouth.

Consistent with similar research with participants who volunteered to complete an online survey in exchange for feedback (e.g., Soto & John, 2017a; Soto, John, Gosling, & Potter, 2011), participants were excluded from the study if (a) they completed fewer than 90% of the survey items (to minimize missing data), (b) their responses to the BFI-2 items had a within-person standard deviation less than 0.5 (to prevent straight-line responding), (c) they completed the survey multiple times (indicated by a question about multiple responding at the end of the survey, and/or responses with combinations of IP address and demographic information), (d) they reported that they did not speak English fluently (to ensure comprehension), (e) they reported inconsistent information about their country and state of residence, (f) they reported an age less than 15 or greater than 85 years old, or (g) they did not report gender (to ensure accurate demographic information).

A total of 1,004 cases met these inclusion criteria, and a final sample of 800 participants was randomly selected from blocks of these cases to balance for gender (50% female, 50% male) and survey version (described below). This selection procedure would help control for gender effects on item and scale characteristics at minimal cost to statistical precision (given the final sample size). This final sample was diverse in terms of its age range (15 to 83; $M = 28.03$, $SD = 13.11$) and ethnicity: 70% identified as White/Caucasian, 16% as Asian/Asian American, 9% as Hispanic/Latino, 7% as Black/African American, 2% as American Indian/Native American, 1% as Native Hawaiian/Pacific Islander, and 5% as another ethnicity, with 10% identifying with multiple ethnicities and 3% not reporting ethnicity. Regarding nationality, 73% resided in the United States, 3% each in Canada, Australia/New Zealand, or the United Kingdom/Ireland, and 10% in another country, with 8% not reporting nationality.

4.1.2. Measures

The Big Five Inventory–2. All participants completed the Big Five Inventory–2 (BFI-2), a 60-item measure of the Big Five personality domains and 15 facet traits (Soto & John, 2017a). The BFI-2 items are short, descriptive phrases that share a common item stem ("I am someone who…"), with each item rated on a five-point Likert scale ranging from 1 = "disagree strongly" to 5 = "agree strongly". Each BFI-2 scale is balanced, indicating that it has an equal number of positively and negatively keyed items, which limits the distorting influence of acquiescence bias (Soto & John, 2019). In the present sample, alpha reliabilities for the five domain scales were 0.86 for Extraversion, 0.79 for Agreeableness, 0.87 for Conscientiousness, 0.91 for Negative Emotionality, and 0.84 for Open-Mindedness. Alphas for the 15 facet scales averaged 0.74, with a range from 0.57 to 0.85. These reliabilities are comparable to previous studies using the BFI-2, indicating good overall data quality.

Candidate BFI-2-H items. All participants also completed 24 candidate BFI-2 Honesty-Humility items, which were interspersed with the 60 standard BFI-2 items. These candidate items were developed by the first, second, and fourth authors to incorporate content from the Dark Triad (Jonason & Webster, 2010) as well as lexical (Ashton, Lee, Perugini, et al., 2004) and questionnaire measures of honesty-humility (Lee & Ashton, 2018). The items were written to parallel the format and hierarchical structure of the BFI-2. Specifically, we wrote items to represent three of the four facets of honesty-humility selected from the HEXACO Personality Inventory–Revised (HEXACO-PI-R; Lee & Ashton, 2018): Sincerity (vs. deception and machiavellianism), Modesty (vs. arrogance and narcissism), and Greed Avoidance (vs. material selfishness). Examples of positively and negatively keyed items were "Is cocky, likes to show off (R)" vs. "Never shows off" for Modesty; "Is cunning, tricks people" vs. "Follows moral and ethical principles" for Sincerity; and "Is greedy, materialistic" (R) vs. "Has inexpensive tastes" for Greed Avoidance. The full set of candidate items is presented in the Appendix. We did not generate items tapping into the fourth Sincerity facet because these were mostly focused on responses to hypothetical situations (e.g., "If I knew that I could never get caught, I would be willing to steal a million dollars") that seemed less conducive to a lexical approach. We revisit this issue in Study 2 when taking a closer look at adjective markers of fairness.

Selection of BFI-2-H items. To create our new H scales, we first computed each candidate BFI-2-H item's corrected item-total correlation with its intended domain and facet scale, as well as its correlations with the standard BFI-2 domain scales. Based on these results as well as item content, we selected three sets of 4 items to maximize (a) internal consistency of the BFI-2 H domain and facet scales, (b) discrimination from the other five standard BFI-2 domain scales, and (c) discrimination between the three BFI-2-H facet scales, while also (d) avoiding highly redundant item pairs. These item selection criteria were selected to yield coherent, differentiated, and content valid domain and facet scales. The alpha reliability of the Honesty-Humility domain scale was 0.75, and alphas for the facet scales were 0.71 for Sincerity, 0.68 for Modesty, and 0.61 for Greed Avoidance, averaging 0.67. These values are somewhat lower than the standard BFI-2 scales (mean facet alpha = 0.74), but comparable to the HEXACO-PI-R (mean alpha in this sample = 0.65).

4.1.3. Validation measures

In addition to the 60 standard BFI-2 items and 24 candidate BFI-2-H items, participants were randomly assigned to complete one of two sets of validation measures: (a) the 16-item Honesty-Humility scale from the

[3] For this reason, we did not consider modifying BFI-2 Agreeableness and Neuroticism content to correspond more with the rotated HEXACO dimensions.

HEXACO-PI-R and the 12-item Dirty Dozen scale, or (b) the 27-item Short Dark Triad scale. All of these measures were administered using the same 5-point Likert scale as the BFI-2.

The HEXACO-PI-R (Lee & Ashton, 2018) includes 16 statement-based items assessing four facet scales of honesty-humility: Sincerity, Modesty, Greed Avoidance, and Fairness (vs. theft and corruption). Of the 16 items, 10 were in the direction of low honesty-humility (e.g., "If I want something from someone, I will laugh at that person's worst jokes"), and 5 of the remaining items were negations (e.g., "I wouldn't want people to treat me as though I were superior to them"). The alpha reliability of the Honesty-Humility domain scale was 0.74, with alphas for the facet scales ranging from 0.54 to 0.75 and averaging 0.65.

The Dirty Dozen (D12; Jonason & Webster, 2010) is a 12-item measure that uses short statements to assess the three Dark Triad trait scales of Machiavellianism (e.g., "I tend to exploit others towards my own end"), Narcissism (e.g.," I tend to want others to admire me"), and Psychopathy (e.g., "I tend to lack remorse"). Whereas the Machiavellianism scale was expected to (inversely) map HEXACO Sincerity and Fairness, Narcissism was expected to (inversely) map onto HEXACO Modesty (associations with Greed Avoidance were investigated in a more exploratory fashion). All items are keyed in the direction of higher values indicating higher dark trait levels. The alpha reliabilities of the three scales were 0.71 for Machiavellianism, 0.68 for Narcissism, and 0.62 for Psychopathy. Scoring all 12-items as a composite measure of dark personality traits yielded an alpha of 0.78.

The Short Dark Triad (SD3; Jones & Paulhus, 2014) uses 27 statement-based items to assess the three Dark Triad traits. Only 3 Narcissism items and 2 Psychoticism items (and 0 Machiavellianism items) are reverse scored, and 3 of these latter items are negations (e.g., "I have never gotten into trouble with the law"). Alpha reliabilities of the three scales were 0.77 for Machiavellianism (e.g., "I like to use clever manipulation to get my way"), 0.74 for Narcissism (e.g., "People see me as a natural leader"), and 0.73 for Psychopathy (e.g., "I like to get revenge on authorities"); scoring a 27-item composite measure of dark traits yielded an alpha of 0.85.

4.2. Results of Study 1

In Study 1, we examined the structure and the convergent and discriminant validity of the preliminary BFI-2-H scales.

4.2.1. Structural validity of the preliminary BFI-2-H scales

Our first analytic goal was to establish whether the BFI-2-H domain is clearly distinguishable from and supplementary to the five standard BFI-2 domains. We thus conducted exploratory factor analyses (EFAs) to examine the multidimensional structure of the expanded instrument. Because we were interested in uncovering orthogonal dimensions, we applied varimax rotation. When we subjected the expanded set of 18 BFI-2 facet scales (15 standard facets plus 3 BFI-2-H facets) to an EFA, the pattern of eigenvalues indicated the presence of five, rather than six, broad dimensions. Specifically, the eigenvalues of the first five unrotated dimensions were 3.98, 2.74, 2.07, 1.75, and 1.48, followed by 0.92, 0.65, 0.59, 0.54, and 0.49. A supplementary CFA analysis (using lavaan, Rosseel, 2012; Version 0.6–5) indicated that adding a sixth factor to a five-factor model improved model fit, $\Delta\chi^2(5) = 102.64$, $p < .001$. That said, even when we extracted and varimax-rotated six dimensions, the sixth component had only one distinctive loading: It was only clearly marked by Greed Avoidance (see Table 1). Modesty also loaded on Extraversion (negatively) and on Agreeableness (positively), and Sincerity barely loaded on the sixth factor, but instead on Agreeableness. When we applied an oblique rotation (oblimin, see Table S1 of the supplement), the latter secondary loading was attenuated in the six-factor solution, but no clear primary loading of the theoretically central Sincerity facet was found. Furthermore, the agreeableness and honesty-humility factors intercorrelated substantially (r = 0.39).

Submitting the expanded set of 72 BFI-2 items (60 standard items plus 12 BFI-2-H items) to an EFA yielded a similar pattern of results. The eigenvalues of the first five unrotated dimensions were 9.70, 6.33, 4.92, 4.42, and 3.64, followed by 2.40, 1.92, 1.66, 1.59, and 1.41, again showing that the 6th factor was much smaller than the first five. However, extracting and varimax-rotating six dimensions did not yield a recognizable Honesty-Humility dimension. Instead, it yielded one dimension defined by positively keyed (i.e., prosocially oriented) items from Agreeableness and Honesty-Humility, and a separate dimension defined by negatively keyed (i.e., antisocially oriented) items from these domains (see https://osf.io/xynf4/?view_only=d157e22c39a4483fa12cb0c6532e94f8 for detailed results).

Table 1
Factor Loadings of Existing and Experimental BFI-2 Facets in Five- and Six-Factor Solutions (Study 1).

Domain and facets	5-factor solution loadings					6-factor solution loadings					
	EXT	AGR	CON	NEG	OPE	EXT	AGR	CON	NEG	OPE	F6
Extraversion domain											
E-Sociability	**.77**	.01	−.01	−.07	.04	**.78**	.06	−.02	−.06	.03	−.06
E-Assertiveness	**.59**	−.23	.29	−.23	.21	**.60**	−.20	.27	−.22	.20	−.10
E-Energy Level	**.65**	.11	.15	−.16	.07	**.66**	.14	.14	−.15	.07	−.01
Agreeableness domain											
A-Compassion	.19	**.72**	.07	.13	.10	.16	**.71**	.08	.11	.10	.13
A-Respectfulness	−.11	**.66**	.26	−.05	−.06	−.18	**.69**	.28	−.09	−.04	.02
A-Trust	.17	**.65**	−.05	−.24	−.03	.13	**.65**	−.04	−.26	−.02	.10
Conscientiousness domain											
C-Organization	.03	.07	**.68**	−.02	−.08	.03	.07	**.68**	−.03	−.08	−.03
C-Productiveness	.26	.07	**.75**	−.18	.07	.27	.07	**.75**	−.18	.07	−.03
C-Responsibility	.02	.25	**.75**	−.11	.01	.03	.21	**.75**	−.12	.01	.07
Negative Emotionality domain											
N-Anxiety	−.11	−.06	−.01	**.84**	−.01	−.12	−.02	−.01	**.83**	−.01	−.12
N-Depression	−.35	−.11	−.16	**.76**	.00	−.34	−.10	−.16	**.77**	.00	−.02
N-Emotional Volatility	.01	−.12	−.18	**.81**	−.04	.01	−.09	−.19	**.82**	−.04	−.06
Open-Mindedness domain											
O-Intellectual Curiosity	.06	−.02	.03	−.08	**.76**	.08	−.04	.03	−.07	**.75**	.05
O-Aesthetic Sensitivity	−.02	.14	−.11	.16	**.69**	−.05	.16	−.10	.15	**.71**	−.05
O-Creative Imagination	.17	−.02	.05	−.11	**.74**	.18	−.04	.05	−.10	**.73**	.04
Honesty-Humility domain											
H-Sincerity	−.09	**.53**	.35	−.03	.06	−.07	**.46**	.36	−.02	.05	.26
H-Modesty	**−.40**	**.44**	.23	−.12	−.06	−.36	.33	.26	−.09	−.07	**.40**
H-Greed avoidance	−.14	.27	−.03	−.15	.07	−.05	.14	−.03	−.10	.05	**.56**

Note. Factor loadings of |0.40| and greater are displayed in bold.

4.2.2. Convergent and discriminant validity of the preliminary BFI-H scales

Our second main analytic goal was to establish the convergent and discriminant validity of the preliminary BFI-2-H scales. For this purpose, we correlated the BFI-2-H domain and facet scales with our selected validity measures: the HEXACO-PI-R, Dirty Dozen, and Short Dark Triad.

We first checked whether each BFI-2-H facet had its strongest correlation with the corresponding dark triad and HEXACO scales. Because of the low reliability of the H facet scales of both instruments, we report disattenuated correlations (displayed below the diagonal in Table 2. The BFI-2-H Sincerity facet indeed correlated strongly with low Machiavellianism as measured by both the Dirty Dozen ($r = -0.90$) and Short Dark Triad ($r = -0.68$). Consistent with the strong overlap between machiavellianism and psychopathy (Muris et al., 2017), a similar (though somewhat weaker) pattern of associations was found with psychopathy. Furthermore, the BFI-2-H Modesty facet was correlated with Narcissism as measured by both the Dirty Dozen ($r = -0.68$) and the Short Dark Triad ($r = -0.47$). Convergent validity was thus especially strong for the Dirty Dozen conceptualization of machiavellianism and narcissism. Furthermore, the BFI-2 facet scales always correlated most strongly with their direct HEXACO equivalent. Specifically, the disattenuated correlation was 0.50 for Sincerity, 0.52 for Modesty, and 0.75 for Greed Avoidance.

We then turned our attention to the preliminary H domain scale. As shown in Table 2, the BFI-2-H domain scale correlated strongly with the dark triad scales, especially Dirty Dozen Machiavellianism (disattenuated $r = -0.80$). The convergent correlations with the other dark triad scales ranged between -0.42 (Short Dark Triad Narcissism) and -0.65 (Dirty Dozen Narcissism). Second, the BFI-2-H domain scale correlated 0.70 with the corresponding HEXACO-PI-R Honesty-Humility domain scale, indicating that instruments scales shared close to half of their reliable variance.

Whereas the pattern of correlations indicated convergent validity with the dark personality/honesty-humility construct, there was a general absence of discriminant validity vis-à-vis the BFI-2 Agreeableness scales. Specifically, especially the BFI-2 Respectfulness facet was strongly associated with the Sincerity ($r = 0.59$) and Modesty ($r = 0.63$) BFI-2-H facets. Also, the BFI-2-H domain scale correlated 0.59 with the BFI-2 Agreeableness domain scale. Only the BFI-2-H Greed Avoidance facet demonstrated a relative lack of associations with BFI-2 Agreeableness, but this was also the scale with the weakest convergent validity with the dark triad scales, casting doubt on its ability to represent content that is shared between the dark triad and honesty-humility.

4.3. Discussion of Study 1

Taken together, the results of Study 1 indicated that we succeeded in constructing supplemental Honesty-Humility scales for the BFI-2 that were face valid and internally consistent. Importantly, these scales showed strong convergent relations with theoretically relevant measures (dark triad measures and the HEXACO-PI-R). However, the results also indicated that these BFI-2-H scales were not clearly differentiated from the Big Five domains as operationalized by the BFI-2. In particular, our H facet and domain scales correlated particularly strongly with the Respectfulness facet of BFI-2 Agreeableness. One possible explanation for the lack of discriminant validity vis-à-vis BFI-2 Agreeableness is that our pool of candidate BFI-2-H items was too narrow, and did not adequately capture the lexical breadth or depth of the honesty-humility construct. Another problem is that that our validity measures consisted of unbalanced criteria (e.g., the entire Dirty Dozen and most of the Short Dark Triad and HEXACO H scales consisted of negatively formulated items), which might have distorted validity coefficients. We therefore conducted additional research to redouble our scale construction efforts while casting a broader conceptual net.

5. Study 2: Lexical correlates of preliminary BFI-2-H facet scales

In Study 1, we tried to establish a preliminary BFI-2 trait domain consisting of three balanced H facet scales measuring Honesty, Humility, and Greed Avoidance. We intended this domain to correlate strongly with HEXACO H but to be (relatively) independent from agreeableness. We did not achieve this goal because, even though the BFI-2-H scales converged quite well with HEXACO H, they overlapped substantially with existing BFI-2 agreeableness facets. As we have noted, the HEXACO not only includes a separate H domain but also blends of agreeableness and neuroticism. Perhaps these modifications have caused its H scale to deviate somewhat from the original lexical H domain (for an alternative operationalization of this domain, see Thalmayer & Saucier, 2014). If that is the case, then perhaps there might exist other H markers that are more independent from Big Five agreeableness. Our first goal was therefore to investigate convergent associations of the experimental BFI-2-H scale with lexical and questionnaire measures of dark personality and H. We expected to replicate the high convergent validity of Study 1 using these alternative validation measures. Our second goal was to demonstrate discriminant validity but given the results of Study 1 and the findings of Crowe et al. (2018) that honesty-humility can be understood as a facet of agreeableness, it was possible that the discriminant validity of these additional H items vis-à-vis BFI-2 Agreeableness would again be low. Therefore, our third goal was to identify additional markers that are more closely mapping onto lexical H while also being relatively independent of BFI-2 Agreeableness facets.

5.1. Method of Study 2

5.1.1. Procedure and sample

We created a study using Qualtrics. All participants answered a total of 298 questionnaire items and adjectives (see below), which were divided into 11 blocks of 27–28 items each to improve user-friendliness. The order of the blocks was randomized, and the order of the questions was also randomized within blocks.

We recruited participants via the Prolific Academic portal, which is similar to Amazon Mturk but based in the European Union. As remuneration, we offered 3 Pounds for an a-priori estimated 35 min of questionnaire time. Post-hoc, on average participants completed the survey a bit faster, in about 25 min. We specified that we wanted to sample a total of 400 participants, because this is well above the sample size at which correlations stabilize (Schönbrodt & Perugini, 2013). A total of 429 entries were registered by Qualtrics, with 404 participants finishing the entire questionnaire. All questions were answered on a 1–5 scale. Due to a clerical error, the answering format of Block 1 was as intended, ranging from 1 = "Completely disagree" to 5 = "Completely agree" but the other blocks had an opposite and slightly differently worded response format, ranging from 1 = "Strongly agree" to 5 = "Strongly disagree". The middle options were identically formulated: "Somewhat agree", "Neither agree nor disagree", and "Somewhat disagree". One participant indicated to have distorted responses, so we excluded this person. Fortunately, we also built in attention check items so we could remove participants who might have been thrown off by the

Table 2
Reliabilities, Raw Correlations, and Disattenuated Correlations of BF2-H Domain and Facet Scales With Selected Validation Measures (Study 1).

Domain and facets	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.
1.BFI-2 H domain	*.75*	.73	.76	.68	.46	.35	.43	.32	-.58/-.43	-.47/-.31	-.37/-.43	**.52**	.31	.40	.35	.30
2.H-Sincerity	1.00	*.71*	.55	.20	.46	.39	.40	.32	**-.64/-.50**	-.28/-.18	-.43/-.44	.43	**.31**	.28	.13	.39
3.H-Modesty	1.00	.38	*.68*	.29	.34	.22	.42	.19	-.37/-.27	**-.46/-.33**	-.24/-.34	.31	.21	**.34**	.13	.13
4.H-Greed Avoidance	1.00	.30	.45	*.61*	.19	.15	.11	.18	-.22/-.18	-.26/-.18	-.11/-.17	.38	.13	.23	**.51**	.11
5.BFI-2 A domain	.59	.62	.47	.27	*.79*	.80	.78	.81	-.39/-.51	-.27/-.14	-.65/-.52	.29	.12	.32	.09	.24
6.A-Compassion	.54	.61	.36	.25	1.00	*.57*	.48	.45	-.29/-.46	-.20/-.08	-.61/-.39	.26	.11	.27	.11	.19
7.A-Respectfulness	.61	.59	.63	.17	1.00	.78	*.65*	.45	-.37/-.37	-.26/-.19	-.50/-.54	.25	.07	.28	.02	.28
8.A-Trust	.46	.47	.29	.29	1.00	.73	.68	*.66*	-.28/-.39	-.19/-.08	-.46/-.32	.19	.10	.22	.07	.12
D12/SD3 dark triad																
9.Machiavellianism	-.80/-.57	**-.90/-.68**	-.54/-.37	-.34/-.26	-.52/-.66	-.46/-.70	-.54/-.52	-.41/-.54	*.71/.77*	.41/.35	.43/.55	-.80	-.77	-.50	-.40	-.57
10.Narcissism	-.65/-.42	-.40/-.25	**-.68/-.47**	-.40/-.26	-.36/-.19	-.32/-.13	-.40/-.27	-.28/-.11	.59/.46	*.68/.74*	.29/.43	-.76	-.50	-.79	-.59	-.25
11.Psychopathy	-.55/-.58	-.65/-.61	-.37/-.48	-.18/-.26	-.92/-.68	-1.00/-.60	-.78/-.79	-.71/-.46	.64/.74	.44/.59	*.62/.73*	-.47	-.22	-.41	-.11	-.56
12.Hexaco H domain	**.70**	.59	.44	.57	.38	.41	.36	.27	-.58	-.54	-.32	*.74*	.62	.63	.72	.61
13.H-Sincerity	.48	**.50**	.35	.22	.18	.20	.13	.17	-.48	-.30	-.13	.99	*.54*	.45	.24	.17
14.H-Modesty	.57	.40	**.52**	.37	.44	.45	.43	.33	-.34	-.53	-.26	.91	.26	*.65*	.35	.09
15.H-Greed Avoidance	.47	.17	.18	**.75**	.11	.17	.03	.11	-.29	-.42	-.07	.96	.37	.50	*.75*	.24
16.H-Fairness	.42	.57	.20	.17	.33	.31	.43	.17	-.40	-.17	-.36	.85	.28	.14	.33	*.68*

Note. Alpha reliability coefficients are displayed on the diagonal in italics. Below the diagonal, disattenuated correlations are displayed. Raw correlations are displayed above the diagonal. Squares depict correlations between constructs of the same content domain. For the dark triad, values for the Dirty Dozen (D12) are displayed left of the slash, whereas values for the Short Dark Triad (SD3) are displayed on the right. Bold values represent the most relevant convergent validity coefficients for the BFI-2H scales.

change in response format, as evidenced by a failed attention check. In total, we thus excluded 54 participants (who still received reimbursement) and ended up with 349 participants.

We included a number of demographic questions to describe our sample. The age of our sample ranged between 18 and 65 years and was 30.0 on average (SD = 9.4), thus skewed towards younger ages. Regarding highest completed education, the most prevalent self-endorsed categories were "High school graduate" (18.6%), "Some college but no degree" (19.2%), "Bachelor's degree" (34.1%) and "Master's degree" (16.9%). Most participants self-identified as "white" (85.7%) or "Asian" (4.6%). In terms of country of origin, many participants resided in the UK (28.9%), Portugal (15.5%), Poland (15.2%), Greece (8.0%), Italy (5.1%), and other countries (mostly from the EU). The sample was gender-balanced, with 50.4% male and 49.6% female.

5.1.2. Measures

We included the 60 standard BFI-2 items to measure the Big Five personality trait domains and facets as outlined by Soto and John (2017a). We also included the 12 preliminary BFI-2-H items selected from Study 1 (see Appendix). The alpha reliability of the BFI-2-H domain scale in Study 2 (0.74) was nearly identical to Study 1. At the facet level, the alpha of the Greed Avoidance was higher in Study 2 (0.69), whereas the alphas of the Sincerity (0.48) and Modesty (0.58) facets were lower. Overall, however, the reliabilities of the existing BFI-2 facet scales were comparable to the values reported by Soto and John (2017a), indicating that most participants indeed answered the questionnaire faithfully.

To measure the variance associated with dark traits, we took a somewhat different approach than in Study 1 by including the newly developed 16-item Dark Factor Scale (DFS; Moshagen et al., 2020) to tap into common variance shared by most dark personality scales. The DFS was developed to create a balanced dark scale consisting of 8 positively and 8 negatively keyed items from established instruments, including 4 items from the Short Dark Triad Scale (Jones & Paulhus, 2014), 4 items from a Serbian dark triad measure (Knežević, 2003), and the other items from various other established instruments (alpha = 0.85). Sample items include "People who mess with me always regret it" and "I feel sorry if things I do upset people". We also included the H scale from the HEXACO questionnaire (alpha = 0.69).

To select *lexical* markers for honesty-humility, we included original lexical markers of H but we also took care to create a more balanced lexical scale by generating antonyms of the lexical markers. This last change appeared necessary because existing instruments of the dark triad and H consist largely of negative (dark) item content, as described in Study 1. First, the adjectives listed in Ashton, Lee, and Goldberg (2004) were included; in the publication, 32 marker items are listed in Table 2. In four cases (wise; uneducated; unlearned; unwise), the markers were deemed less representative of the target dimension, so these terms were excluded. In five cases, the negation of the adjective (e.g., "uncrafty" vs. "crafty") was excluded because we preferred affirmative descriptors. The list of adjectives was expanded by additionally generating antonyms but avoiding the prefix "un-" as much as possible except when the unnegated adjective could be confused with a verb (e.g., "unpretending" cannot be confused with a verb but "pretending" can) or otherwise was ambiguous (e.g., "unmercenary" is less ambiguous than "mercenary" when used with the item stem "I am someone who is…"). In total, this resulted in 39 adjectives. In a second publication by Ashton et al. (2006), a total number of 47 honesty-humility adjectives were reported for Italian, Dutch and English.[4] Of all items, 17 were redundant with Ashton, Lee, and Goldberg (2004b), Ashton, Lee, Perugini, et al. (2004a), either as a primary adjective or antonym. Furthermore, one item could refer to both an adjective and verb ("lying") and

[4] An anonymous reviewer of an earlier version of this paper pointed out that we missed the study by Lee and Ashton, who reported lists of Honesty-Humility items in 14 samples. Subsequent inspection of that study revealed that its two English item lists overlapped 92% and 71% with the adjectives used in the current study, indicating strong representation.

was therefore excluded. Of the remaining 28 adjectives, 3 were negations ("unwily"), so they were used only in its affirmative form ("wily"). In total, this resulted in 55 adjectives. In total, we included 94 possible lexical markers for honesty-humility. We report these items on the project's OSF page: https://osf.io/xynf4/?view_only=d157e22c39a4483fa12cb0c6532e94f8. Our OSF page also includes the raw data and R scripts that were used to produce the reported results.

Our second strategy to generate possible lexical markers was to scan the following narcissism, psychopathy, and machiavellianism instruments with the goal of extracting dark adjectives. For machiavellianism: the Mach IV (Christie & Geis, 1970) and the Machiavellian Personality Scale (Dahling, Whitaker, & Levy, 2009). For narcissism: the Five-Factor Narcissism Inventory (Glover, Miller, Lynam, Crego, & Widiger, 2012), the Pathological Narcissism Inventory (Pincus et al., 2009), the Narcissistic Personality Inventory (Raskin & Hall, 1979), the Narcissistic Admiration and Rivalry Questionnaire (Back et al., 2013), and the Narcissistic Grandiosity Scale (Rosenthal et al., 2020). Finally, for psychopathy: the Levenson Self-Report Psychopathy Inventory (Levenson, Kiehl, & Fitzpatrick, 1995), the Antisocial Features Scale from the Personality Assessment Inventory (Morey & Boggs, 2004), the short form of the Psychopathic Personality Inventory (Lilienfeld & Andrews, 1996), and the Self-Report Psychopathy Scale (Paulhus, Neumann, & Hare, 2009). Consistent with the findings of Moshagen et al. (2018), we found that item content not clearly differentiated between scales focusing on (ostensibly) different dark constructs, so we abandoned our efforts to classify adjectives as belonging to narcissism, psychopathy, or machiavellianism and instead aggregated them into a global index of lexical dark traits. This strategy resulted in 106 additional adjectives (not including adjectives that were already contained in the Ashton publications; these adjectives were excluded to avoid redundancy).

Together, both strategies produced 200 adjectives. We factor analyzed responses (using oblimin rotation) to explore the dimensionality of these markers and found evidence for multiple dimensions, but most items loaded on two correlated factors ($r = -0.49$, $p < .001$, between corresponding scale scores) tapping into positively vs. negatively valenced "light" and "dark" content, respectively. To separately investigate convergent validity of the BFI-2 scales vis-a-vis dark personality vs. honesty-humility content, we proceeded with calculating two lexical marker scales. First, we aggregated all 106 dark adjectives (including 48 reverse coded items), which resulted in a highly reliable scale (alpha = 0.92). Still, 17 of these items were (slightly) negatively correlated with the total scale, so we excluded these items and computed the scale score again (alpha = 0.94). Second, our *Lexical H Total scale* consisted of all 94 lexical markers reported by Ashton and colleagues as well as generated antonyms (49 items were keyed in a "dark" direction, indicating that the total scale was relatively balanced). This scale was highly reliable, with an alpha of 0.89, but 17 adjectives had negative item-total correlations (e.g., crude, boorish, and uncultivated). We excluded these items, which increased the alpha level to 0.92. To further explore convergence with the original results by Ashton et al., we also computed a *Lexical HH Ashton scale* consisting only of the 54 original adjectives that were reported in their publication (i.e., excluding the additionally generated antonyms). Note that 42 of these adjectives were keyed in a "dark" direction and needed to be recoded so that higher levels indicated more honesty-humility (alpha = 0.90). There were two items with a slightly negative item-total correlation ("unpolished"; "unmercenary"), which resulted in a composite scale with a reliability of 0.91.

5.2. Results of Study 2

Study 2 had three main goals: Establishing a) the convergent validity of our preliminary H scales vis-à-vis dark personality and H content, b) the discriminant validity of the preliminary scales vis-à-vis BFI-2 Agreeableness, and c) if this discriminant validity was found to be lacking, whether additional markers can be identified that are more closely mapping onto lexical H while also being relatively independent of BFI-2 Agreeableness facets.

5.2.1. Structural validity of the preliminary BFI-2-H scales

As in Study 1, we began by jointly factor analyzing the 15 standard BFI-2 facets plus the three preliminary BFI-2-H facets. Replicating the results of Study 1, an inspection of the scree plot as well as parallel analysis suggested five factors. The eigenvalues of the first five unrotated dimensions were 4.62, 2.68, 2.06, 1.25, and 1.12, followed by 0.86, 0.79, 0.64, 0.56, and 0.50. An inspection of the factor loadings of the five-factor solution indicated that the proposed HH facets loaded most strongly on the Agreeableness factor (Table 3). Modesty again had a negative loading on Extraversion, whereas Sincerity had a positive (though smaller) loading on Conscientiousness. Applying oblimin rotation to the five-factor solution attenuated the loadings with Agreeableness, but a strongly negative loading of Modesty on Extraversion remained. We also inspected a six-factor solution, and present results in Table 3. There was indeed evidence of a weakly defined sixth factor, dominated by Modesty. Greed Avoidance also had a loading > 0.40 on this factor, whereas Sincerity failed to demonstrate a substantial loading, instead loading on the Agreeableness factor even in the six-factor solution. Conversely, none of the A facets, like compassion, loaded on the 6th factor with the H facets. Applying an oblimin rotation to the six-factor solution did not result in an increased factor loading of Sincerity with the sixth factor (see Table S2).

5.2.2. Convergent and discriminant validity of the preliminary BFI-2-H scales

A first goal of Study 2 was to check whether the preliminary BFI-2 honesty-humility domain and facet scales developed in Study 1 tap into lexical and questionnaire measures of dark personality and honesty-humility (convergent validity). The raw and disattenuated correlations are displayed in Table 4. Focusing on the disattenuated correlations, the BFI-2-H domain scale correlated somewhat less than -0.70 with the dark personality indicators, around 0.80 with the lexical honesty-humility indicators, and around 0.90 with HEXACO Honesty-Humility. These figures indicated an overlap of at least 50% with the reliable variance in these target constructs, especially with the questionnaire operationalization of H. The strong disattenuated correlation of our BFI-2-H domain scale with lexical markers of honesty-humility was replicated by two of our facets, Sincerity (0.95 with the full adjective scale and 0.88 with the Ashton subset) and Modesty (0.73 with both lexical scales) but not by Greed Avoidance (disattenuated correlation of 0.54 and 0.53 with the respective lexical scales). Lexical honesty humility was more strongly overlapping with our total BFI-2-H domain scale (disattenuated correlation of 0.80 with the original Ashton adjectives) than with the HEXACO H scale (disattenuated correlation of 0.71), $t = 3.03$, $p < .01$. This was to our surprise because the latter scale was derived from Ashton's own lexical studies. The dark personality indicators also converged more strongly with lexical honesty-humility than the HEXACO scale, with convergent (disattenuated) correlations of up to -0.89 (between lexical dark personality and the full lexical H scale).

Table 3
Factor Loadings of Existing and Experimental BFI-2 Facets in Five- and Six-Factor Solutions (Study 2).

Domain and facets	5-factor solution loadings					6-factor solution loadings					
	EXT	AGR	CON	NEG	OPE	EXT	AGR	CON	NEG	OPE	F6
Extraversion domain											
E-Sociability	**.61**	.00	.03	−.16	.02	**.47**	.11	.03	−.21	.02	−.28
E-Assertiveness	**.60**	−.19	.20	−.13	.05	**.82**	−.17	.12	−.12	.06	−.11
E-Energy Level	**.55**	.21	.31	−.20	.25	**.40**	.32	.34	−.26	.25	−.22
Agreeableness domain											
A-Compassion	.19	**.76**	.12	.14	.11	.13	**.76**	.15	.11	.11	.14
A-Respectfulness	−.09	**.65**	.23	−.15	.11	−.09	**.58**	.25	−.15	.10	.28
A-Trust	.13	**.56**	.01	−.26	.18	.01	**.57**	.05	−.29	.17	.06
Conscientiousness domain											
C-Organization	.02	.08	**.58**	−.08	.21	−.03	.05	**.65**	−.09	.20	−.01
C-Productiveness	.27	.20	**.73**	−.19	.07	.32	.18	**.68**	−.21	.07	.08
C-Responsibility	.12	.37	**.58**	−.15	−.01	.15	.32	**.57**	−.16	−.01	.16
Negative Emotionality domain											
N-Anxiety	−.21	.03	−.04	**.83**	.14	−.16	.03	−.03	**.84**	.13	.04
N-Depression	−.36	−.23	−.23	**.68**	−.03	−.27	−.25	−.24	**.71**	−.03	.04
N-Emotional Volatility	.03	−.13	−.17	**.70**	.02	−.02	−.06	−.16	**.68**	.02	−.17
Open-Mindedness domain											
O-Intellectual Curiosity	.01	.02	.09	.09	**.77**	−.06	.05	.12	.07	**.77**	−.09
O-Aesthetic Sensitivity	.02	.14	.05	.05	**.69**	.06	.12	.04	.06	**.70**	.09
O-Creative Imagination	.29	.19	.20	−.05	**.50**	.33	.20	.16	−.06	**.51**	.05
Honesty-Humility domain											
H-Sincerity	−.14	**.55**	.35	.00	.09	−.11	**.46**	.36	.01	.09	.29
H-Modesty	**−.45**	**.47**	.19	−.13	−.07	−.16	.21	.12	−.07	−.05	**.79**
H-Greed avoidance	−.26	**.44**	.06	−.10	.07	−.11	.29	.03	−.06	.08	**.45**

Note. Factor loadings of |0.40| and greater are displayed in bold.

All in all, these results suggest that the preliminary BFI-2-H domain scale captured the lexical honesty-humility factor at least as well as do lexical and questionnaire measures of dark personality traits and questionnaire measures of honesty-humility. In terms of discriminant validity, however, we again were not able to establish our preliminary H scales as independent of BFI-2 Agreeableness. Specifically, both the BFI-2-H domain scale ($r = 0.66$) and its H facets (especially Sincerity, $r = 0.85$) were very strongly associated with the BFI-2 Agreeableness domain scale. The (disattenuated) correlation between the H facet of Sincerity and the A facet of Respectfulness was even as high as 0.89, suggesting almost perfect overlap of reliable variance. Of note, the BFI-2 Agreeableness domain was very strongly correlated with lexical and questionnaire indicators of dark personality as well as lexical indicators of honesty-humility, capturing the bulk of their respective reliable variance (e.g., BFI-2 Agreeableness captured $-.93^2 = 86\%$ of reliable variance in lexical indicators of dark personality).

5.2.3. Exploration of individual adjectives

The above results indicate that the preliminary BFI-2-H domain and facet scales (a) clearly captured the content of dark personality and honesty-humility dimensions (convergent validity), but (b) were insufficiently differentiated from the BFI-2 Agreeableness domain (lack of discriminant validity). To clarify whether the lack of differentiation of the full BFI-2-H domain simply reflects an issue with the content of the specific items that we developed and selected for the BFI-2-H scales, our final set of analyses for Study 2 aimed to cast a broader net for potential item content. Specifically, we aimed to identify personality descriptive adjectives that both (a) converge strongly with established measures of

Table 4
Reliabilities, Raw Correlations, and Disattenuated Correlations of BF2-H, BFI2-A, Dark Personality Factors, and Honesty-Humility.

Domain and facets	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
1.BFI-2 H domain	*.75*	.69	.80	.81	.51	.41	.48	.33	**-.58**	**-.54**	**.66**	**.68**	**.66**
2.H-Sincerity	1.00	*.48*	.38	.28	.52	.43	.47	.34	-.63	-.53	.58	.63	.46
3.H-Modesty	1.00	.72	*.58*	.49	.33	.26	.39	.17	-.39	-.37	.53	.53	.48
4.H-Greed Avoidance	1.00	.50	.77	*.69*	.34	.28	.27	.27	-.36	-.35	.42	.43	.57
5.BFI-2 A domain	.66	.85	.50	.47	*.78*	.81	.79	.78	-.80	-.69	.65	.69	.47
6.A-Compassion	.61	.79	.44	.44	1.00	*.62*	.50	.43	-.69	-.64	.58	.63	.40
7.A-Respectfulness	.71	.89	.66	.42	1.00	.83	*.59*	.41	-.66	-.58	.59	.62	.40
8.A-Trust	.49	.63	.29	.41	1.00	.71	.70	*.60*	-.55	-.44	.38	.41	.33
Dark personality traits													
9. Lexical dark traits	**-.69**	-.94	-.52	-.45	-.93	-.91	-.88	-.73	*.94*	.72	-.75	-.83	-.53
10. Dark factor of personality	**-.67**	-.82	-.52	-.46	-.85	-.88	-.82	-.62	.80	*.85*	-.67	-.72	-.56
Honesty-humility													
11. Lexical honesty-humility (Ashton)	**.80**	.88	.73	.53	.77	.77	.80	.52	-.81	-.75	*.91*	.96	.57
12. Lexical honesty-humility (all)	**.81**	.95	.73	.54	.82	.83	.84	.55	-.89	-.81	1.00	*.92*	.58
13. Hexaco honesty-humility	**.92**	.80	.76	.83	.63	.60	.63	.52	-.66	-.73	.71	.72	*.69*

Note. Alpha reliability coefficients are displayed on the diagonal in italics. Below the diagonal, disattenuated correlations are displayed. Raw correlations are displayed above the diagonal. Squares depict correlations between constructs of the same content domain. Bold values represent the most relevant convergent validity coefficients for the BFI-2H scales.

dark personality traits/honesty-humility, and (b) are clearly differentiated from the BFI-2 domains, especially Agreeableness. We therefore correlated all 200 dark personality adjectives with both BFI-2-Agreeableness and the full lexical H scale, and then used the resulting 2 × 200 matrix as input for subsequent analysis. Notably, the two columns of this matrix were very strongly correlated, $r = 0.96$, indicating that adjectives with a strong association with BFI-2-Agreeableness were also strongly associated with H. We plotted the corresponding correlations in Fig. 1 (for an interactive version of this graph that allows the user to sort through the individual adjectives, see https://osf.io/xynf4/?view_only=d157e22c39a4483fa12cb0c6532e94f8) and found that the scatter plot confirmed the strong convergence between adjectives' associations with Agreeableness and with H. Thus, there were only limited opportunities to sample H markers that were relatively independent from BFI-2 Agreeableness.

We still proceeded and applied a very lenient criterion of $r < 0.30$ with Agreeableness and $r > 0.40$ with Lexical HH Total. This resulted in 12 possible adjectives: "ostentatious", "entitled", "superior", "artificial", "haughty", "snobbish", "presumptuous", "slick", "foxy", "swaggering", "grandiose", and "wily". These adjectives are reminiscent of grandiose narcissism, which has been shown to be associated less strongly (compared to other narcissism dimensions) with negative or antisocial outcomes (Back et al., 2013). We thus proceeded by creating a composite scale with these items (alpha = 0.82). The convergent validity of this experimental scale was moderate, as it was correlated -0.57 with the preliminary BFI-2-H domain scale and -0.51 with the HEXACO H scale. As can be expected based on the item content, the adjectives correlated most strongly with the Modesty facet of BFI-2-H, $r = -0.55$. However, we again ran into problems with discriminant validity, as the correlation across all items was -0.45 with the BFI-2 Agreeableness domain scale. Accordingly, even these items did not succeed in being sufficiently independent from BFI-2 Agreeableness to form the basis of a broad and independent BFI-2-H domain scale.

We repeated the above analyses with the HEXACO H scale as benchmark and found a similarly strong vector correlation of $r = 0.94$, indicating that adjectives that correlated highly with BFI-2 Agreeableness also did so with HEXACO H. Only a single adjective correlated $r < 0.30$ with Agreeableness and $r > 0.40$ with HEXACO H: "superior," again reminiscent of grandiose narcissism. With the D scale as benchmark, the vector correlation was -0.97, and the only distinguishing item was again "superior". Thus, regardless of the benchmark for honesty-humility, we did not identify markers that were sufficiently independent from BFI-2 Agreeableness.

Finally, we revisited the decision to forego the construction of a fairness facet while exploring an expanded BFI-2. Might this decision have hampered our ability to find honesty-humility content that is independent of agreeableness? Inspection of the adjectives analyzed in Study 2 suggests that the answer is "no." The adjective list included several terms representing fairness. Specifically, a content analysis produced 20 items (e.g., "exploitative", "corrupt", "fair", vs. "unfair")

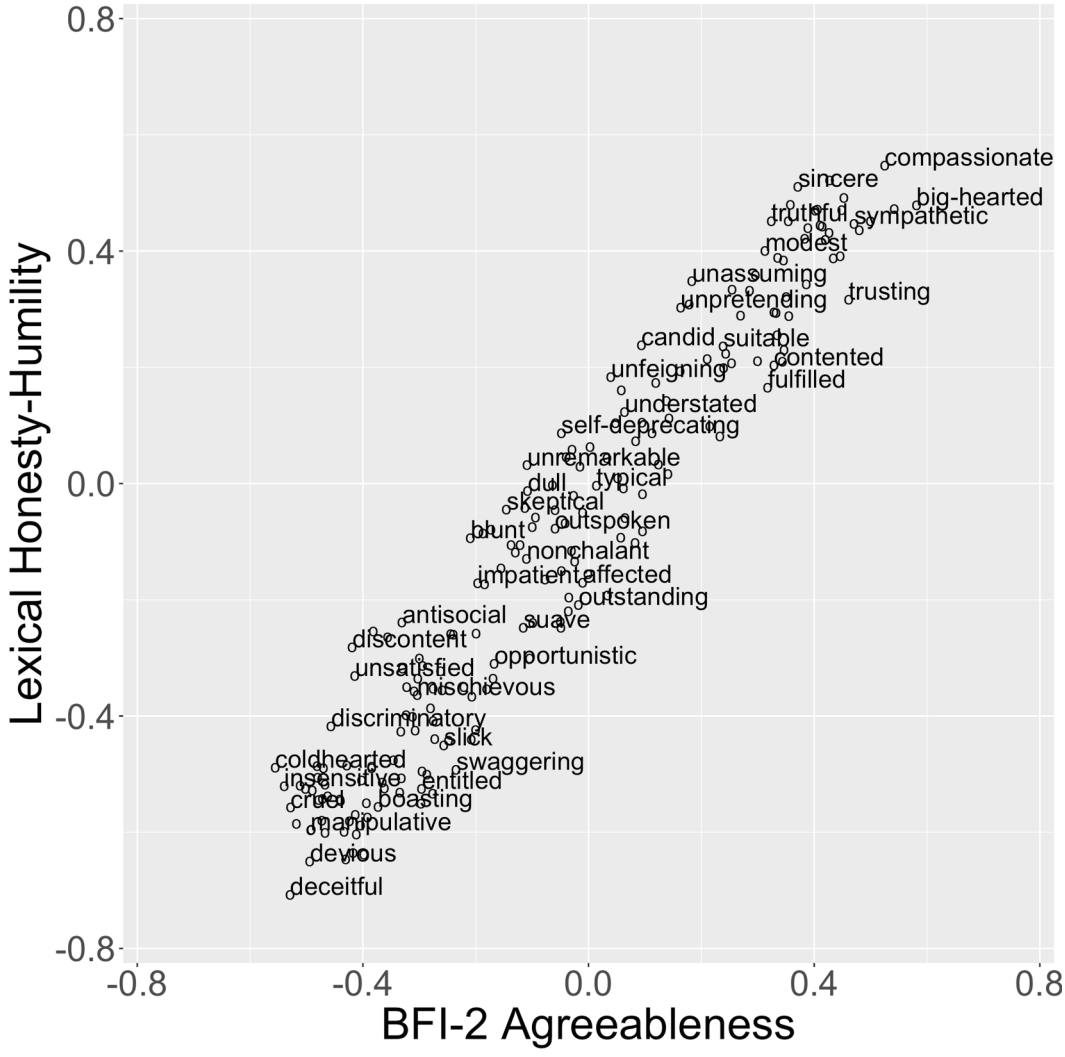

Fig. 1. Graphical Depiction of Correlations of Adjectives With BFI-2 Agreeableness Domain Scale and Lexical Honesty-Humility Scale in Study 3.

that were deemed highly representative of the fairness construct. When we constructed a corresponding scale from these 20 adjectives (alpha = 0.83), it correlated strongly with HEXACO Honesty-Humility (r = −0.44), but even more strongly with BFI-2 Agreeableness (r = −0.66).

5.3. Discussion of Study 2

Study 2 had three main goals. First, we wanted to test whether the preliminary BFI-2-H domain and facet scales developed in Study 1 adequately captured the lexical dark personality/honesty-humility domain. Our results confirmed that the BFI-2-H scales represent this lexical content at least as well as do established lexical and questionnaire measures of dark personality and honesty-humility. Second, we aimed to further test whether the preliminary BFI-2-H scales could be clearly distinguished from the rest of the Big Five domains as operationalized by the BFI-2. However, replicating the results of Study 1, we found that the results using the preliminary BFI-2-H scales were not strong enough to define a separate honesty-humility dimension. Our third goal was to cast a broader net for potential BFI-2-H content, by identifying personality-descriptive adjectives that capture dark personality/honesty-humility content distinct from BFI-2 Agreeableness. However, our analyses of 200 adjectives drawn from previous lexical research and established questionnaire measures did not yield enough content to fulfill this aim. Thus, we reluctantly concluded that constructing an independent, lexically based Honesty-Humility domain scale for the BFI-2 may not be possible. However, we still believed that it should be possible to supplement the representation of dark personality/honesty-humility content in the BFI-2. This has precedent in the literature, as the NEO-PI-R Agreeableness domain also contains facets that are associated with honesty-humility: Straightforwardness and Modesty (Ashton & Lee, 2005). We therefore conducted a final study with the more modest goal of developing and validating a supplemental Honesty-Humility facet scale for the BFI-2 within the Big Five Agreeableness domain.

6. Study 3: Creating and validating a new H scale as a facet of agreeableness

The aim of Study 3 was to develop a supplemental BFI-2 Agreeableness facet scale to assess individual differences in behavior related to dark personality traits/honesty-humility. As stated in the introduction, several studies have identified facets of agreeableness-antagonism that are closely linked to both dark personality and honesty-humility. Specifically, in the Crowe et al. (2018) analysis, two facets of agreeableness were identified that would meet this goal: morality (i.e., the honesty aspect of honesty-humility) and modesty (i.e., the humility aspect). These correspond to the BFI-2-H facets of Sincerity and Modesty that were most strongly associated with dark personality traits and lexical honesty-humility, in contrast to the Greed Avoidance that was less strongly associated with these benchmarks (see Table 4). Such a situation would somewhat resemble the structure of the NEO-PI-R Agreeableness domain, which features two facets (Straightforwardness and Modesty) that are more strongly linked to H (Ashton & Lee, 2005). These two facets are conceptually related to low levels of machiavellianism and narcissism, respectively. Furthermore, Morizot (2014) previously expanded the BFI-1 with some additional content, resulting in an Agreeableness scale that included a machiavellianism item, with good psychometric properties.

Our criteria of convergent validity differed from Study 1 and 2: Not only did we aim to establish convergent correlations with dark personality traits and honesty-humility, but now also with the other BFI-2 Agreeableness facets. However, to establish discriminant validity we did expect that the BFI-2-H-facet of Agreeableness would be more strongly correlated with "darker" content (e.g., negatively with the Straightforwardness and Modesty facets of NEO-PI-R Agreeableness) than the other BFI-2 Agreeableness facets. Conversely, we expected the existing BFI-2 Agreeableness facets of Compassion, Respectfulness, and Trust to display stronger correlations with corresponding "lighter" agreeableness content. All expected facet associations were preregistered on the OSF: https://osf.io/gb6wu/?view_only=2a438fcaae004e3e83335fb695d90ec8. Finally, we aimed to test the ability of the new H facet subscale to predict prosocial behavior in a behavioral economic paradigm that resembles the dictator game, which has been used to demonstrate the predictive validity of H in earlier studies (Hilbig et al., 2015; Zhao et al., 2017).

6.1. Method of Study 3

6.1.1. Procedure and sample

We again recruited participants via Prolific Academic, specifying a similar target sample size of 400. The questionnaire was somewhat shorter than in Study 2, so we offered 2.10 Pounds for an estimated 25 min of questionnaire time. A total number of 408 completed entries were registered by Qualtrics. Questions were implemented in 5 blocks of 48 questions each, in randomized order. We excluded the seven participants who failed three or more of our inattentive responding checks. Of the remaining participants, age ranged between 18 and 70 years and was 32.76 on average (SD = 10.61). Regarding education, the most prevalent self-endorsed categories were "High school graduate" (16.7%), "Some college but no degree" (21.3%), "Bachelor's degree" (28.9%) and "Master's degree" (20.3%). Like in Study 2, most participants self-identified as "white" (89.4%). In terms of country of origin, many participants resided in the UK (36.8%) and various other countries (mostly from the EU). The sample was gender-balanced, with 47.4% males and 52.6% females.

6.1.2. Measures

To represent content directly relevant to both honesty and humility, we used the lexical dark personality/honesty-humility content from Study 2 to create 4 additional honesty and 4 additional humility items. These items were added to all preliminary Sincerity and Modesty items that we previously developed in Study 1 and used in Study 2 (8 items per construct, see Appendix). For the Modesty items, we made sure that the additional items would tap into more rivalrous/entitled aspects of narcissism because the results of Study 2 had indicated that admiration aspects might fail to load high on a continuum of antagonism-agreeableness. For the full list of candidate items, see https://osf.io/xynf4/?view_only=d157e22c39a4483fa12cb0c6532e94f8.

For convergent validation purposes, we administered both the 10-item Honesty-Humility scale and the 10-item Agreeableness scale from the HEXACO. Furthermore, we additionally measured all six NEO-PI-R Agreeableness facets (Costa & McCrae, 2008), as well as the Agreeableness scale of the Five Individual Reaction Norms Inventory (FIRNI; Denissen & Penke, 2008), which taps into individuals' tendency to act selfishly when competing with others for resources. Furthermore, like in Study 2 we again included the 16-item Dark-Factor scale (Moshagen et al., 2020), as well as the Narcissism, Machiavellianism, and Psychopathy facets of the Short Dark Triad Scale (Jones & Paulhus, 2014) and the Dirty Dozen Scale (Jonason & Webster, 2010).

Finally, to establish the predictive validity of the BFI-2 H-facet of

Agreeableness, we implemented the Social Value Orientation (SVO) sliders (Murphy, Ackermann, & Handgraaf, 2011), which are conceptually similar to the Dictator Game. Participants were told that they were "randomly paired with another person, [who is] someone you do not know and will remain mutually anonymous." They were then presented with six outcome matrices with 9 response options each. Each of these response options represented distinction distributions of 100 points between the participant and the "other person". They were told to "imagine that your decisions will yield money for both yourself and the other person" and provided with an example. Points allocated to the self vs. the other were averaged to measure the relative utility of outcomes attached to the self vs. other people (alpha = 0.53 and 0.67, respectively).

6.2. Results of Study 3

6.2.1. Selecting four items for a BFI-2 honesty-humility facet of agreeableness

Our first goal for Study 3 was to select items for a possible Honesty-Humility facet scale within the BFI-2 Agreeableness domain. Similar to Study 1, we correlated each of the 24 candidate BFI-2-H items developed for Study 3 with the standard BFI-2 domain and facet scales, and also computed each item's corrected item-total correlation. Based on these convergent and discriminant correlations, as well as judgments of item content (to avoid pairs of highly redundant items), we iteratively reduced the pool of 24 candidate items to a final, 4-item facet scale. These items are: "is cunning, tricks people", "is arrogant, has a big ego", "is humble, doesn't think they're better than others", and "is sincere, genuine."

The reliability of the final Honesty-Humility facet scale was 0.68, which is on par with the reliability of the other agreeableness facet scales and close to the median reliability of 0.71 found across all BFI-2 facet scales (ranging from 0.61 to 0.82). Furthermore, structural analyses indicated that the new Honesty-Humility facet was clearly located within the Agreeableness domain. When we submitted the 15 standard BFI-2 facets plus the supplemental Honesty-Humility facet to an EFA, the eigenvalues of the first five unrotated dimensions were 4.39, 2.67, 2.02, 1.28, and 1.11, followed by 0.64, 0.59, 0.51, 0.46, and 0.42. Moreover, as shown in Table 5, when we extracted and varimax-rotated five dimensions, the Honesty-Humility facet loaded strongly (0.72) on the agreeableness dimension. Notably, this was only slightly lower than the factor loading of the Compassion facet (0.75), which is the core facet of Agreeableness (Soto & John, 2017a). Secondary loadings of the H facet on the other four dimensions were weak ($\leq$0.21). Finally, forcing a six-factor solution did not result in a clearly defined sixth factor, since no facet loaded $>$ 0.40 on this factor). These results clearly supported the placement of H as facet of agreeableness. This support was also found when rotating the five- and six-factor solutions obliquely (see Table S3).

Item-level structural analyses yielded similar results. Specifically, when we submitted the 60 standard BFI-2 items plus 4 supplemental Honesty-Humility items to an EFA or PCA, the first five eigenvalues were 10.28, 6.21, 4.73, 3.31, and 2.81, followed by 2.16, 1.63, 1.51, 1.35, and 1.27. Because this did not unequivocally suggest a five-factor solution, we first extracted and varimax-rotated six factors. However, separate positively keyed vs. negatively keyed agreeableness factors emerged, mirroring the results of Study 2 and suggesting the impact of acquiescence. Indeed, when we controlled for acquiescence, the first five eigenvalues were 10.67, 5.99, 4.47, 3.05, and 2.91, followed by 1.73, 1.63, 1.43, 1.32, and 1.27, which is consistent with a five-factor solution. When we extracted and varimax-rotated five dimensions, all four Honesty-Humility items had their strongest loading ($\geq$0.42) on the Agreeableness dimension (see https://osf.io/xynf4/?view_only=d157e22c39a4483fa12cb0c6532e94f8. These results indicate that the supplemental Honesty-Humility facet scale was clearly located within the BFI-2 Agreeableness domain.

6.2.2. Convergent and discriminant validity

As stated above, we pre-registered the hypothesized pattern of convergent and discriminant correlations with the new BFI-2 Honesty-Humility facet scale. Table 6 displays the correlations with all assessed outcomes scales, with the hypothesized correlations displayed in bold. Our 25 predictions were confirmed in all but one case. For example, we aimed for (and predicted) a H facet scale that would correlate with the other BFI-2 Agreeableness facets, and indeed this was the case, with a disattenuated correlation 0.87 with Compassion, 0.85 with Respectfulness, and 0.58 with Trust. This suggests a very strong degree of

Table 5
Factor Loadings of Existing and Experimental BFI-2 Facets in Five- and Six-Factor Solutions (Study 3).

Domain and facets	5-factor solution loadings					6-factor solution loadings					
	EXT	AGR	CON	NEG	OPE	EXT	AGR	CON	NEG	OPE	F6
Extraversion domain											
E-Sociability	**.81**	.04	.03	−.10	−.01	**.80**	.05	.04	−.12	.00	.07
E-Assertiveness	**.61**	−.21	.14	−.15	.18	**.68**	−.20	.16	−.15	.19	−.11
E-Energy Level	**.61**	.18	.23	−.23	.22	**.53**	.17	.25	−.25	.24	.39
Agreeableness domain											
A-Compassion	.06	**.75**	.12	.13	.13	.05	**.75**	.12	.13	.13	.02
A-Respectfulness	−.07	**.68**	.26	−.15	.13	−.08	**.68**	.26	−.14	.12	−.01
A-Trust	.18	**.59**	.09	−.25	−.01	.15	**.58**	.10	−.25	−.01	.11
A-Honesty-Humility	−.16	**.72**	.21	.05	.12	−.16	**.72**	.20	.05	.12	−.05
Conscientiousness domain											
C-Organization	.08	.16	**.69**	−.10	−.03	.06	.16	**.69**	−.09	−.03	.02
C-Productiveness	.34	.19	**.76**	−.17	.11	.28	.19	**.77**	−.18	.12	.13
C-Responsibility	.01	**.41**	**.67**	−.11	.10	.02	**.42**	**.67**	−.10	.09	−.10
Negative Emotionality domain											
N-Anxiety	−.23	.05	−.04	**.76**	.18	−.20	.05	−.05	**.74**	.17	−.14
N-Depression	−.38	−.11	−.19	**.71**	−.04	−.32	−.11	−.20	**.70**	−.05	−.21
N-Emotional Volatility	.01	−.09	−.16	**.84**	−.04	.01	−.10	−.15	**.91**	−.05	.24
Open-Mindedness domain											
O-Intellectual Curiosity	.07	.08	−.02	.05	**.71**	.10	.09	−.02	.06	**.72**	−.15
O-Aesthetic Sensitivity	−.02	.16	.03	.08	**.66**	−.05	.16	.03	.08	**.66**	.04
O-Creative Imagination	.23	.03	.10	−.07	**.66**	.18	.02	.11	−.08	**.67**	.15

Note. Factor loadings of |0.40| and greater are displayed in bold.

convergence but still at least 25% of reliable variance in Honesty-Humility (in the case of Compassion) was not shared with existing facets.

Turning to the convergent and discriminant validity of the existing BFI-2 Agreeableness scale, Table 6 displays the convergent correlations (raw and disattenuated) in bold. In the one case where a discriminant correlation was stronger than this pre-registered convergent correlation, it is underlined. Consistent with predictions, Compassion was most strongly associated with NEO-PI-R Tender Mindedness and Altruism, as well as with FIRNI Agreeableness. This is consistent with the compassionate content of these validation scales, which are focused on helping others. Furthermore, BFI-2 Respectfulness was strongly correlated with NEO-PI-R Compliance and HEXACO Agreeableness, although the latter was somewhat more highly correlated with BFI-2 Trust. Finally, and consistent with our prediction, the BFI-2 Trust facet was most strongly correlated with the corresponding NEO-PI-R scale (the disattenuated correlation was 1.00, suggesting complete overlap).

Most importantly, our pre-registered hypotheses for the experimental Honesty-Humility facet were all confirmed: Compared to the existing facets, this new facet correlated more strongly with NEO-PI-R Straightforwardness and Modesty, the D-Factor scale, and all dark trait scales (especially SDT Psychopathy and all Dirty Dozen scales). Of note, and consistent with its operationalization, the new facet correlated 0.75 (disattenuated) with HEXACO Honesty-Humility, whereas the convergence with H was "only" between 0.42 and 0.57 for the existing BFI-2 Agreeableness facets.

Consistent with the stronger convergent correlation between the new BFI-2 facet and HEXACO Honesty-Humility, the expanded BFI-2 agreeableness subscales (four in total) were better able to predict HEXACO Honesty-Humility than were the three original BFI-2 agreeableness subscales. Specifically, when using the BFI-2 facets as predictors in a multiple regression, adding the BFI-2 Honesty-Humility facet increased multiple R^2 from 0.19 to 0.29, a 53% gain in relative predictive power. Compared with this, the multiple R^2 for the six NEO-PI-R facets was 0.41. Thus, expanding BFI-2 agreeableness brings it into closer alignment with HEXACO Honesty-Humility, though still not quite as close as the six NEO-PI-R facets.

Finally, we investigated what happened if the traditional BFI-2 Agreeableness domain scale (based on Compassion, Respectfulness, and Trust) is expanded with the novel Honesty-Humility facet. As can be seen in Table 6 (Columns 1 and 2), the expanded scale was somewhat more reliable (0.85 vs. 0.80, consistent with the increased number of items). While associations with validation criteria that were linked to Compassion did not decrease, the disattenuated correlations with criteria linked to Respectfulness and Trust decreased slightly ($\Delta|r|$ less than 0.06). Conversely, convergent validity of the new scale with criteria linked to Honesty Humility increased by up to 0.10.

6.2.3. Prediction of pro-social behavior

As stated in the introduction, HEXACO H has been used to predict prosocial behavior in behavioral economic games, most often the Dictator Game. To evaluate the predictive validity of the supplemental BFI-2 Agreeableness H facet, we implemented the Social Value Orientation (SVO) sliders (Murphy et al., 2011). The value attached to outcomes of others is conceptually similar to the Dictator Game. Table 7 shows the correlations of the HEXACO Agreeableness and H scales and the BFI-2 Agreeableness facets (including the new H facet) with this criterion. As can be seen, all scales correlated positively with the value of outcomes for others. We then proceeded with a series of regression models to establish the incremental validity of our new BFI-2 Agreeableness H facet. Model 1 was intended as a baseline model to replicate a dissociation between HEXACO Agreeableness and H in predicting outcomes for others (e.g., Hilbig et al., 2015). Indeed, we found that only HEXACO H but not HEXACO "Agreeableness" uniquely predicted SVO. In Model 2, we added our new H facet scale to the two HEXACO predictors and found that our new facet scale uniquely predicted variance in SVO-others. Model 3 was intended as a baseline model to gauge to performance of the established BFI-2 Agreeableness facets in the prediction of SVO. As can be seen in Table 7, Compassion positively predicted value attached to outcomes for others. In Model 4, however, we added our new H facet scale and found that it was the only BFI-2 Agreeableness facet that uniquely predicted individual differences in SVO. Expanding the BFI-2 with a facet to measure honesty-humility thus succeeded in predicting a key criterion variable that has been used to demonstrate the utility of the HEXACO H scale.

6.3. Discussion of Study 3

Overall, our analyses of convergent, discriminant, and predictive validity supported the convergent and discriminant validity of the supplemental BFI-2 Honesty-Humility facet scale within a nomological network of conceptually relevant personality measures. Perhaps most

Table 6
Reliabilities, Raw Correlations, and Disattenuated Correlations of BF2-Agreeableness (Including the HH Facet) and Various Validation Criteria.

Scale label	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.	22.
1. BFI-2 Agreeableness (excl. H)	*.80*	.97	.82	.80	.79	.64	.60	.50	.76	.52	.41	.55	.63	.59	.43	-.68	-.42	-.17	-.60	-.51	-.30	-.68
2. BFI-2 Agreeableness (incl. H)	1.00	*.85*	.81	.80	.73	.80	.56	.58	.78	.52	.50	.57	.65	.57	.49	-.73	-.45	-.26	-.65	-.59	-.39	-.73
3. BFI-2 A-Compassion	1.00	1.00	*.61*	.52	.45	.56	.35	.46	**.71**	.32	.41	**.53**	**.60**	.32	.38	-.60	-.36	-.23	-.51	-.49	-.30	-.60
4. BFI-2 A-Respectfulness	1.00	1.00	.80	*.69*	.42	.58	.33	.43	.66	**.51**	.38	.39	.43	**.51**	.37	-.57	-.29	-.17	-.59	-.45	-.33	-.59
5. BFI-2 A-Trust	1.00	.97	.70	.61	*.68*	.40	**.74**	.32	.46	.43	.19	.40	.48	.59	.29	-.46	-.36	-.02	-.34	-.30	-.10	-.45
6. BFI-2 A-Honesty-Humility	.86	1.00	.87	.85	.58	*.68*	.32	**.62**	.62	.36	**.60**	.45	.53	.36	.52	**-.66**	**-.41**	**-.41**	**-.62**	**-.62**	**-.50**	**-.66**
7. NEO-PI-R Trust	.74	.68	.50	.45	**1.00**	.43	*.81*	.33	.42	.34	.07	.39	.39	.43	.23	-.44	-.36	.02	-.33	-.23	-.06	-.41
8. NEO-PI-R Straightforwardness	.68	.76	.71	.63	.47	**.90**	.45	*.69*	.48	.37	.47	.39	.52	.31	.55	-.58	-.60	-.43	-.59	-.67	-.42	-.66
9. NEO-PI-R Altruism	.97	.97	**1.00**	.91	.65	.85	.53	.66	*.76*	.33	.37	.53	.59	.35	.30	-.64	-.29	-.10	-.51	-.42	-.20	-.62
10. NEO-PI-R Compliance	.73	.71	.52	**.77**	.65	.55	.47	.56	.47	*.63*	.35	.37	.38	.67	.32	-.45	-.36	-.26	-.50	-.31	-.27	-.38
11. NEO-PI-R Modesty	.51	.61	.59	.52	.27	**.82**	.09	.64	.48	.50	*.77*	.42	.52	.26	.52	-.52	-.38	-.66	-.51	-.51	-.59	-.47
12. NEO-PI-R Tender Mindedness	.79	.79	**.86**	.61	.63	.70	.55	.60	.77	.60	.61	*.61*	.61	.35	.34	-.66	-.34	-.26	-.45	-.36	-.23	-.47
13. FIRNI Agreeableness	.79	.79	**.86**	.58	.65	.72	.48	.69	.76	.54	.66	.87	*.80*	.37	.59	-.70	-.55	-.36	-.54	-.53	-.39	-.57
14. HEXACO Agreeableness	.77	.73	.48	**.73**	.83	.51	.56	.44	.47	.98	.34	.52	.49	*.73*	.24	-.39	-.30	-.12	-.36	-.26	-.18	-.39
15. HEXACO Honesty-Humility	.57	.63	.57	.53	.42	.75	.31	.79	.42	.48	.71	.52	.78	.34	*.71*	-.55	-.51	-.45	-.55	-.62	-.57	-.52
16. Dark Factor	-.82	-.86	-.83	-.75	-.61	**-.87**	-.53	-.76	-.79	-.62	-.64	-.92	-.85	-.49	-.71	*.85*	.53	.37	.74	.60	.44	.68
17. SDT Machiavellianism	-.55	-.57	-.54	-.41	-.52	**-.58**	-.46	-.84	-.39	-.53	-.50	-.51	-.71	-.41	-.70	.67	*.73*	.40	.46	.55	.38	.57
18. SDT Narcissism	-.22	-.32	-.33	-.24	-.03	**-.56**	.02	-.60	-.13	-.38	-.86	-.38	-.47	-.17	-.62	.46	.53	*.76*	.47	.50	.62	.33
19. SDT Psychopathy	-.79	-.84	-.77	-.85	-.49	**-.89**	-.44	-.84	-.69	-.74	-.68	-.68	-.72	-.49	-.78	.96	.64	.64	*.72*	.63	.53	.60
20. D12 Machiavellianism	-.70	-.78	-.76	-.66	-.45	**-.91**	-.31	-.98	-.58	-.47	-.71	-.55	-.71	-.37	-.89	.79	.78	.69	.90	*.68*	.56	.67
21. D12 Narcissism	-.44	-.55	-.51	-.52	-.16	**-.78**	-.09	-.66	-.30	-.44	-.88	-.38	-.57	-.28	-.88	.62	.58	.93	.82	.88	*.59*	.44
22. D12 Psychopathy	-.94	-.99	-.95	-.89	-.68	**-.99**	-.57	-.99	-.88	-.60	-.66	-.74	-.80	-.57	-.77	.92	.82	.47	.88	1.00	.71	*.65*

Note. Reliabilities are displayed at the diagonal. Below the diagonal, disattenuated correlations are displayed. Raw correlations are displayed above the diagonal. Squares depict correlations between scales of the same instrument domain. The total BFI-2 agreeableness domain scale score was computed *without* the final H facet. Correlations that were pre-registered are displayed in bold, correlations that did not conform to the pre-registered pattern are underlined. SDT = Short Dark Triad, D12 = Dirty Dozen.

importantly, the supplemental facet scale converged strongly with established questionnaire measures of dark personality traits/honesty-humility and provided incremental prediction of prosocial behavior beyond the three standard BFI-2 Agreeableness facet scales.

We therefore conclude that the supplemental facet scale increased the representation of dark personality/honesty-humility content captured by the BFI-2.

7. General discussion

7.1. Development of distinct BFI-2 Honesty-Humility scale

The present research had two overarching goals: to develop a supplemental Honesty-Humility scale for the BFI-2, and to gain conceptual insights into how honesty-humility and dark personality traits can be best incorporated within a Big Five framework. Regarding the first goal, we failed to develop an Honesty-Humility domain scale that was distinct from the BFI-2′s standard Big Five domains (Studies 1 and 2). However, we did succeed in developing and validating an Honesty-Humility facet scale to supplement the BFI-2′s Agreeableness domain (Study 3). This supplemental facet scale allows researchers interested in both the Big Five and personality variance represented by the Dark Triad and honesty-humility to efficiently capture this broad range of content with the BFI-2. This increases the usefulness of the BFI-2 as a comprehensive measure of personality trait assessment that is of high psychometric quality and also freely available. It also increases researchers' options regarding what scale to use when aiming to predict pro-social behaviors. Our results demonstrated that the inclusion of the new Honesty-Humility subscale increased the coverage of facet traits assessed by the NEO-PI-R and the HEXACO-PI, as well as the prediction of prosocial behavior in behavioral economic paradigms. In the following, we discuss implications, recommendations, and limitations of our research.

In Study 3, we provided substantial evidence for the reliability and structural validity of the new Honesty-Humility facet scale. The internal consistency of the new facet was 0.68, which is not high enough for individual assessment in high-stake situations but is comparable to the internal consistency of other BFI-2 facets and merits the use of the scale in research contexts. Regarding structural validity, we found that the items of the new Honesty-Humility facet scale correlated more strongly with each other than with items from other Agreeableness facets, but that these four facets jointly defined an Agreeableness factor in structural analyses. We therefore conclude that the supplemental Honesty-Humility facet scale effectively increases the scope of the BFI-2 Agreeableness domain. For researchers interested in incorporating this variance when administering the BFI-2′s short or extra-short forms (BFI-2-S or BFI-2-XS; Soto & John, 2017b), based on the results of item-level analyses we recommend supplementing the abbreviated forms with the supplemental items "Is arrogant, has a big ego" (for both the BFI-2-S and BFI-2-XS) and "Is sincere, genuine" (for the BFI-2-S only).

The present research provided evidence for the content and external validity of the supplemental BFI-2-H scale. That is, the new BFI-2-H facet correlated strongly with measures of dark personality and honesty-humility. Regarding the former, our new facet scale correlated strongly with various indicators of dark personality. Regarding the latter, convergent correlations with lexical H were stronger than correlations with the corresponding HEXACO H scale. Accordingly, the BFI-2 Agreeableness domain, with the H facet added, appears to represent both HEXACO H and Big Five Agreeableness content. The increased saturation of the BFI-2 with H content was also apparent in the pattern of convergent correlations with the NEO-PI-R facets. Specifically, the new H facet was more strongly associated with the Straightforwardness and Modesty facets of the NEO-PI-R, which have traditionally been linked to H content (Ashton & Lee, 2005). The BFI-2 Agreeableness domain expanded with the new H subscale thus also converged better with the NEO-PI-R coverage of the Agreeableness domain. Although we currently have no plans of doing so, it of course remains an open question whether the BFI-2 Agreeableness domain could be expanded even further, for example by adding content related to sentimentality.

7.2. Conceptual conclusions

Beyond the construction and validation of a supplemental Honesty-Humility scale for the BFI-2, our second overarching goal was to gain insights regarding the conceptual relations of honesty-humility and dark personality traits with the Big Five. What conceptual conclusionsfv, then, can we draw from the results of our scale development efforts? In Studies 1 and 2, we attempted to construct a supplemental Honesty-Humility domain, encompassing the facets of Modesty, Humility, and Greed Avoidance. However, while this new domain converged with both lexical and questionnaire measures of honesty-humility and dark traits, as a domain it failed to demonstrate sufficient independence from BFI-2 Agreeableness in structural analyses.

This failure offers one key insight: honesty-humility cannot be simply added to the Big Five as an "extra" orthogonal trait domain. Rather, it would require modification of the original scale content. One possibility is that creating space for an independent honesty-humility dimension requires redefining both agreeableness and neuroticism away from their Big Five conceptualizations (Ashton & Lee, 2007). Another possibility is that honesty-humility represents a late split from a broader agreeableness factor (Ashton, Lee, & Goldberg, 2004) and, as such, some original content must be removed from agreeableness to create space for an honesty-humility domain. For example, compassion content—which is central to Big Five agreeableness but falls on the border between the HEXACO agreeableness and honesty-humility—must be largely excluded. Either way, these operations cannot be repeated post hoc with the existing BFI-2 measure, which is based on prototypical definitions of

Table 7
Correlations, Multiple Regression Weights, and Confidence Intervals of Agreeableness/Honesty-Humility Scales Predicting Individual Differences in Social Value Orientation (Value of Outcomes for Others).

		Model 1		*Model 2*		*Model 3*		*Model 4*	
Scale	*r*	*β*	*ci*	*β*	*ci*	*β*	*ci*	*β*	*ci*
HEXACO Agreeableness	.14*	.05	[−05, 15]	.01	[−09, 11]				
HEXACO Honesty-Humility	.38*	.37*	[27, 47]	.29*	[19, 39]				
BFI-2 A-Honesty-Humility	.32*			.16*	[04, 28]			.31*	[19, 43]
BFI-2 A-Compassion	.20*					.12	[00, 24]	.02	[−10, 14]
BFI-2 A-Respectfulness	.17*					.08	[−04, 20]	−.04	[−16, 08]
BFI-2 A-Trust	.17*					.09	[−03, 21]	.06	[−06, 18]
*R*2		.15		.17		.05		.10	

Note. * $p < .05$. ci = 95% confidence interval.

the Big Five domains. Doing so would require us to change much more BFI-2 content (i.e., also the agreeableness and negative emotionality domains), which we deemed undesirable because the BFI-2 has already been widely adopted and is purposefully embedded in the theoretical framework of the Big Five. If users are willing the tolerate highly intercorrelated domains, then we recommend them to use the separate H domain scale of Study 2.

In contrast with our failure to construct an independent Honesty-Humility domain scale, in Studies 2 and 3 we successfully developed an Honesty-Humility facet scale to supplement the BFI-2 Agreeableness domain. What conceptual insights can this result offer? Some researchers (e.g., Lynam & Miller, 2019) have proposed that agreeableness is the positive pole of a broadly construed agreeableness-antagonism continuum, which encompasses not only agreeableness but also honesty-humility and the Dark Triad (Paulhus & Williams, 2002). Within this conceptualization, honesty-humility and dark personality traits are more accurately conceptualized as facets of agreeableness-antagonism than as defining a separate trait domain (Crowe et al., 2018). Consistent with this claim, our supplemental BFI-2-H subscale converged with the BFI-2 Agreeableness domain in structural analyses, and also correlated strongly with lexical and questionnaire measures of both honesty-humility and dark personality traits. We would like to emphasize that we did not see this outcome as a foregone conclusion. We (especially the first two authors) were quite optimistic about the prospects of developing a full-fledged Honesty-Humility domain to extend the BFI-2, and we were genuinely surprised by how difficult this proved to be .

7.3. Future research

Future research using the supplemental BFI-2-H facet scale can contribute to the emerging conceptual integration between the literature on "normal" vs. "dark" personality – which appear more like two sides of the same coin than as altogether different phenomena. One important nuance pertains to the position of narcissism on this continuum. As has been argued before (Back et al., 2013; Krizan & Herlache, 2018), narcissism consists of admiration/grandiosity and rivalry/entitlement, which are substantially correlated but have different nomological networks. In Study 3, we demonstrated that the rivalry aspects of narcissism performed best as items of the new HH subscale, and in Study 2 we found some evidence that admiration aspects, as reflected in the NEO-PI-R Modesty facet (e.g., "I have a very high opinion of myself') might instead form an independent domain in a sixth-factor solution, or alternatively are defined by low agreeableness and high extraversion in a five-factor solution (Paulhus, 2001). The Humility content of the new BFI-2-H facet thus pertains mostly to a tendency of avoiding rivalry (e.g., the denigration of others), rather than avoiding a tendency to boost the self.

In further considering the relations between agreeableness, honesty-humility, and the Dark Triad, one intriguing finding was that lexical markers of honesty-humility identified by Ashton, Lee, and Goldberg (2004), Ashton, Lee, Perugini, et al. (2004) converged somewhat more strongly with a preliminary version of the BFI-2-H scale (disattenuated $r = 0.80$) than with the HEXACO-PI Honesty-Humility scale (disattenuated $r = 0.72$). While both of these correlations reflect strong convergence, the somewhat lower value for the HEXACO-PI domain scale is noteworthy because of this inventory's roots in lexical analysis. Content analysis of the HEXACO items indeed confirmed that most items do not describe people using personality-relevant adjectives, but instead gauge people's reactions to hypothetical situations, as in: "I would never accept a bribe, even if it were very large." It is possible that such items allow the HEXACO Honesty-Humility scale to more directly assess relevant personality processes (e.g., moral disengagement). This may increase the scale's predictive validity, but raises questions about whether the HEXACO-PI should be regarded as a purely lexical personality instrument, or whether it mixes in additional content areas based on theoretical arguments. More research is needed on this question, which should ideally include lexical markers of the honesty-humility domain, multiple questionnaire measures (including the new BFI-2-H scale), and relevant outcome variables.

7.4. Strengths and limitations

The current study had a number of strengths. For example, we conducted three separate investigations with sample sizes large enough to provide high statistical power and precise effect size estimates. We also pre-registered all hypotheses for Study 3 and included a behavioral-economic paradigm to validate the new BFI-2-H facet scale. That said, the present research was also limited in some ways. One limitation was that we only assessed personality using self-reports, and even our behavioral outcome in Study 3 can be regarded as a quasi-self-report measure because no points were allocated to an actual interaction partner. Thus, future research is needed to test whether the present findings extend to behavioral measures and/or informant-reports. This may be especially valuable due to the evaluative and interpersonal nature of dark or antisocial personality traits (Vazire, 2010). Our reliance on self-reports might also have been a limiting factor because not all participants in Studies 2 and 3 were English native speakers. Finally, we did not include actual behavioral criteria in Study 3, instead relying on the allocation of imagined points distributed to imagined co-players. Future research should include the supplemental BFI-2-H facet scale in more objective and externally valid settings, such as in behavioral economic experiments with actual co-players and incentives, clinical and forensic settings, and high-stakes situations where people can violate trust and cheat others. More generally, our findings should be independently replicated in other samples before strong conclusions can be drawn.

A second limitation is that we did not administer all HEXACO-PI domain scales. This might have limited the generalizability of our efforts to find H-content that is independent of agreeableness. For example, we did not have direct benchmarks of the HEXACO blends of high A and N, which would perhaps have allowed more independent H content to emerge. At a more practical level, not including all HEXACO scales made it impossible to evaluate "how much" the supplemental BFI-2-H scale improved overall convergence between the BFI-2 and the HEXACO-PI (cf. Ashton et al., 2019). Perhaps future research comparing these instruments would indicate that the supplemented BFI-2 (especially when analyzed at the facet level) captures more of the HEXACO variance, although we expect that the sixth HEXACO factor would still capture some unique variance due to its more thorough assessment of honesty-humility.

8. Conclusion

Many researchers are interested in measuring both the Big Five personality domains and traits related to moral vs. immoral or antisocial behavior. To help accommodate this desire, we developed a supplemental Honesty-Humility facet scale for the Big Five Inventory–2. This

BFI-2 Honesty-Humility facet scale retained the conceptual coherence and strong psychometrics of the standard BFI-2 scales, effectively captured the content of the lexical honesty-humility dimension, converged with established measures of honesty-humility and the Dark Triad in expected ways, and predicted prosocial behavior within a resource allocation situation. We therefore conclude that this optional scale allows interested researchers to more directly assess the moral aspect of personality using the BFI-2. The scale development and validation process also yielded conceptual insights regarding how honesty-humility and dark personality traits can be incorporated within the BFI-2 framework. Specifically, our results suggest that—from a BFI-2 perspective—honesty-humility and the Dark Triad are better conceptualized as facets of agreeableness than as an independent trait domain beyond the Big Five.

9. Author Note

All authors came up with the idea for the study. JD collected the data, conducted the analyses, and wrote a first version of the MS. All authors then provided comments and suggestions, and approved the final version of the MS.

Data collection and preliminary analysis were sponsored by a grant of the Abbas Fund awarded to Jaap Denissen.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Full list of candidate Items, Item-Total Correlations, and item selections in Study 1 and 3

Facet scale / item	Corrected item-total correlation with facet scale (Study 1)	Selected for H domain and facet scales	Corrected item-total correlation with facet scale (Study 3)	Selected for Study A facet scale
Sincerity				
Study 1				
- Follows moral and ethical principles	0.33		0.46	
- Always tells the truth	0.41	X	0.66	
- Is always honest	0.42	X	0.65	
- Listens to their conscience	0.26		0.50	
- Is cunning, tricks people (R)	0.51	X	0.70	X
- Is crafty, sly (R)	0.34		0.42	
- Manipulates people to get their own way (R)	0.56	X	0.69	
- Tells lies (R)	0.47		0.65	
Study 3 additional items				
- Is straightforward, direct			0.17	
- Is sincere, genuine			0.69	X
- Can be devious, sneaky (R)			0.60	
- Deceives people to get what they want (R)			0.68	
Modesty				
- Never shows off	0.40	X	0.57	
- Feels like* an average, ordinary person	0.15		0.46	
- Is modest, humble	0.39	X	0.55	
- Never brags about their achievements	0.35		0.49	
- Can be conceited, stuck-up (R)	0.45		0.56	
- Is cocky, likes to show off (R)	0.58	X	0.68	
- Is arrogant, has a big ego (R)	0.51		0.64	X
- Tries to impress others (R)	0.44	X	0.46	
Study 3 additional items				
- Humble, doesn't think they're better than others			0.61	X
- Modest about their accomplishments			0.48	
- Likes to boast and brag (R)			0.66	
- Is snobbish, feels superior to others (R)			0.65	
Greed Avoidance				
- Has inexpensive tastes	0.16	X		
- Pays very little attention to money and possessions	0.18	X		
- Would never steal	0.30			
- Enjoys giving things away (R)	0.14			
- Is greedy, materialistic (R)	0.52	X		
- Likes to have expensive things (R)	0.36	X		
- Likes to own things that impress people (R)	0.49			
- Would do anything for personal gain (R)	0.41			

Note. Selection of items was based on item-total correlations but also on discriminant validity vis-à-vis other facets. * The formulation was "average, ordinary person" in Study 1 but this was changed to make the item more suitable for other-reports.

Appendix B. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jrp.2021.104147.

References

Ashton, M. C., & Lee, K. (2001). A theoretical basis for the major dimensions of personality. *European Journal of Personality, 15*(5), 327–353.

Ashton, M. C., & Lee, K. (2005). Honesty-Humility, the Big Five, and the Five-Factor Model. *Journal of Personality, 73*(5), 1321–1354. https://doi.org/10.1111/j.1467-6494.2005.00351.x

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

Ashton, M. C., & Lee, K. (2019). How well do Big Five measures capture HEXACO scale variance? *Journal of Personality Assessment, 101*(6), 567–573. https://doi.org/10.1080/00223891.2018.1448986

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

Ashton, M. C., Lee, K., de Vries, R. E., Perugini, M., Gnisci, A., & Sergi, I. (2006). The HEXACO model of personality structure and indigenous lexical personality dimensions in Italian, Dutch, and English. *Journal of Research in Personality, 40*(6), 851–875. https://doi.org/10.1016/j.jrp.2005.06.003

Ashton, M. C., Lee, K., Perugini, M., Szarota, P., de Vries, R. E., Di Blas, L., … De Raad, B. (2004a). A six-factor structure of personality-descriptive adjectives: Solutions from psycholexical studies in seven languages. *Journal of Personality and Social Psychology, 86*(2), 356–366. https://doi.org/10.1037/0022-3514.86.2.356

Ashton, M. C., Lee, K., & Goldberg, L. R. (2004b). A hierarchical analysis of 1,710 English personality-descriptive adjectives. *Journal of Personality and Social Psychology, 87*(5), 707–721. https://doi.org/10.1037/0022-3514.87.5.707

Ashton, M. C., Lee, K., & Visser, B. A. (2019). Where's the H? Relations between BFI-2 and HEXACO-60 scales. *Personality and Individual Differences, 137*, 71–75. https://doi.org/10.1016/j.paid.2018.08.013

Back, M. D., Küfner, A. C. P., Dufner, M., Gerlach, T. M., Rauthmann, J. F., & Denissen, J. J. A. (2013). Narcissistic admiration and rivalry: Disentangling the bright and dark sides of narcissism. *Journal of Personality and Social Psychology, 105*(6), 1013–1037. https://doi.org/10.1037/a0034431

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

Costa, P. T., & McCrae, R. R. (2008). The Revised NEO Personality Inventory (NEO-PI-R). In *The SAGE handbook of personality theory and assessment: Personality measurement and testing* (Vol. 2, pp. 179–198). Sage Publications, Inc.. https://doi.org/10.4135/9781849200479.n9

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.v86.510.1111/jopy.12358

Dahling, J. J., Whitaker, B. G., & Levy, P. E. (2009). The development and validation of a new Machiavellianism scale. *Journal of Management, 35*(2), 219–257.

Denissen, J. J. A., Geenen, R., Soto, C. J., John, O. P., & van Aken, M. A. G. (2020). The Big Five Inventory–2: Replication of psychometric properties in a Dutch adaptation and first evidence for the discriminant predictive validity of the facet scales. *Journal of Personality Assessment, 102*(3), 309–324. https://doi.org/10.1080/00223891.2018.1539004

Denissen, J. J. A., & Penke, L. (2008). Motivational individual reaction norms underlying the Five-Factor model of personality: First steps towards a theory-based conceptual framework. *Journal of Research in Personality, 42*(5), 1285–1302. https://doi.org/10.1016/j.jrp.2008.04.002

Gardiner, G., Sauerberger, K., & Funder, D. (2019). Towards meaningful comparisons of personality in large-scale cross-cultural studies. In A. Realo (Ed.), *In praise of an inquisitive mind: A Festschrift in honor of Jüri Allik on the occasion of his 70th birthday* (pp. 123–139). University of Tartu Press.

Glover, N., Miller, J. D., Lynam, D. R., Crego, C., & Widiger, T. A. (2012). The five-factor narcissism inventory: A five-factor measure of narcissistic personality traits. *Journal of Personality Assessment, 94*(5), 500–512.

Hilbig, B. E., Thielmann, I., Hepp, J., Klein, S. A., & Zettler, I. (2015). From personality to altruistic behavior (and back): Evidence from a double-blind dictator game. *Journal of Research in Personality, 55*, 46–50. https://doi.org/10.1016/j.jrp.2014.12.004

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

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 *Handbook of personality: Theory and research* (3rd ed., pp. 114–158). The Guilford Press.

Jonason, P. K., & Webster, G. D. (2010). The Dirty Dozen: A concise measure of the dark triad. *Psychological Assessment, 22*(2), 420–432. https://doi.org/10.1037/a0019265

Jones, D. N., & Paulhus, D. L. (2014). Introducing the Short Dark Triad (SD3): A Brief Measure of Dark Personality Traits. *Assessment, 21*(1), 28–41. https://doi.org/10.1177/1073191113514105

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

Krizan, Z., & Herlache, A. D. (2018). The narcissism spectrum model: A synthetic view of narcissistic personality. *Personality and Social Psychology Review, 22*(1), 3–31. https://doi.org/10.1207/s15327906mbr3902_8

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research, 39*(2), 329–358. https://doi.org/10.1207/s15327906mbr3902_8

Lee, K., & Ashton, M. C. (2005). Psychopathy, Machiavellianism, and Narcissism in the Five-Factor Model and the HEXACO model of personality structure. *Personality and Individual Differences, 38*(7), 1571–1582. https://doi.org/10.1016/j.paid.2004.09.016

Lee, K., & Ashton, M. C. (2018). Psychometric Properties of the HEXACO-100. *Assessment, 25*(5), 543–556. https://doi.org/10.1177/1073191116659134

Levenson, M. R., Kiehl, K. A., & Fitzpatrick, C. M. (1995). Assessing psychopathic attributes in a noninstitutionalized population. *Journal of Personality and Social Psychology, 68*(1), 151–158. https://doi.org/10.1037//0022-3514.68.1.151

Lilienfeld, S. O., & Andrews, B. P. (1996). Development and preliminary validation of a self-report measure of psychopathic personality traits in noncriminal population. *Journal of Personality Assessment, 66*(3), 488–524.

Lynam, D. R., Caspi, A., Moffitt, T. E., Raine, A., Loeber, R., & Stouthamer-Loeber, M. (2005). Adolescent psychopathy and the Big Five: Results from two samples. *Journal of Abnormal Child Psychology, 33*(4), 431–443.

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

Morey, L. C., & Boggs, C. D. (2004). *The Personality Assessment Inventory (PAI)*. John Wiley & Sons Inc.

Morizot, J. (2014). Construct validity of adolescents' self-reported Big Five personality traits: Importance of conceptual breadth and initial validation of a short measure. *Assessment, 21*(5), 580–606. https://doi.org/10.1177/1073191114524015

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

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

Murphy, R. O., Ackermann, K. A., & Handgraaf, M. (2011). Measuring social value orientation. *Judgment and Decision Making, 6*(8), 771–781.

Paulhus, D. L. (2001). Normal narcissism: Two minimalist accounts. *Psychological Inquiry, 12*(4), 228–230.

Paulhus, D. L., Neumann, C. S., & Hare, R. D. (2009). Manual for the self-report psychopathy scale.

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

Pincus, A. L., Ansell, E. B., Pimentel, C. A., Cain, N. M., Wright, A. G., & Levy, K. N. (2009). Initial construction and validation of the pathological narcissism inventory. *Psychological Assessment, 21*(3), 365–379.

Raskin, R. N., & Hall, C. S. (1979). A narcissistic personality inventory. *Psychological Reports, 45*(2), 590.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48*(2), 1–36. http://www.jstatsoft.org/v48/i02/.

Rosenthal, S. A, Hooley, J. M., Montoya, R. M., van der Linden, S. L., & Steshenko, Y. (2020). The Narcissistic Grandiosity Scale: A measure to distinguish narcissistic grandiosity from high self-esteem. *Assessment, 27*(3), 487–507. https://doi.org/10.1177/1073191119858410

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality, 47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

Soto, C. J., & John, O. P. (2017a). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology, 113*(1), 117–143. https://doi.org/10.1037/pspp0000096

Soto, C. J., & John, O. P. (2017b). Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality, 68*, 69–81. https://doi.org/10.1016/j.jrp.2017.02.004

Soto, C. J., & John, O. P. (2019). Optimizing the length, width, and balance of a personality scale: How do internal characteristics affect external validity? *Psychological Assessment, 31*(4), 444–459.

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), 330–459.

Thalmayer, A. G., & Saucier, G. (2014). The questionnaire big six in 26 nations: Developing cross-culturally applicable big six, big five and big two inventories. *European Journal of Personality, 28*(5), 482–496.

Vazire, S. (2010). Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology, 98*(2), 281.

Vize, C. E., Collison, K. L., Miller, J. D., & Lynam, D. R. (2021). The "core" of the dark triad: A test of competing hypotheses. *Personality Disorders: Theory, Research, and Treatment, 11*(20), 91–99. https://doi.org/10.1037/per0000386 (in press).

Zhao, K., Ferguson, E., & Smillie, L. D. (2017). Individual differences in good manners rather than compassion predict fair allocations of wealth in the dictator game. *Journal of Personality, 85*(2), 244–256. https://doi.org/10.1111/jopy.12