Personality and Individual Differences 259 (2026) 113854

Contents lists available at ScienceDirect

Personality and Individual Differences

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

Beyond good and evil: A person-centred analysis of benevolent vs. malevolent interpersonal dispositional styles

David Skvarc [a,*], Brittany Patafio [a], Richelle Mayshak [a], Shannon Hyder [a], Scott Barry Kaufman [b], Craig Neumann [c]

[a] School of Psychology, Deakin University, Australia
[b] Columbia University, United States of America
[c] Department of Psychology, University of North Texas, United States of America

ARTICLE INFO

Keywords:
Dark Triad
Light Triad
Personality traits
Latent profile analysis
Prosocial behaviour
Aggression

ABSTRACT

Recent research has identified three unique profiles that reflect variation in benevolent versus malevolent dispositions. These dispositions (aka "light" vs. "dark" propensities) go beyond general personality trait domains and offer the opportunity to study individuals in terms of affiliative versus aversive interpersonal styles. To further this line of research, we performed latent profile analysis with a general population sample ($N = 2332$; $M_{age} = 39.45$, $SD = 17.12$) and assessed interpersonal dispositions reflecting benevolent (Kantianism, Humanism, Faith in Humanity) and malevolent (Machiavellianism, Narcissism, Psychopathy) features. Results replicated the previously reported three-class solution corresponding to three subtypes: 1) those with higher benevolent traits and lower malevolent traits (20.7%) those with the opposite profile (8.7%), and those with approximately equal and moderate levels across all features (70.6%). The obtained three-class structure broadly matched other normative samples, indicating good validity. External validation was provided by significant and substantial differences across these three subtypes corresponding to prosocial behaviour, relational aggression perpetration and victimisation, social exclusion, vulnerable narcissism (but not grandiose), and sadism. The findings provide further insight into affiliative vs. aversive dispositional interpersonal styles, with implications for demographic differences, as well as a utilitarian interpretation of exclusionary behaviours.

Interest in understanding benevolent vs. malevolent personality traits has increased dramatically in recent years (Neumann et al., 2025; Neumann & Ngo, 2025). Benevolent traits, such as Kantianism, Humanism, and Faith in Humanity, capture affiliative prosocial orientations rooted in compassion and moral regard for others, while malevolent dispositions, including Machiavellianism, Psychopathy, and Narcissism, reflect traits associated with aversive interpersonal exploitation of others and self-interest (Kaufman et al., 2019; Neumann et al., 2025). Although considerable research has established the predictive validity of malevolent traits for antisocial and aggressive outcomes (Muris et al., 2017; Skvarc et al., 2025), empirical work on the nature of benevolent traits remains comparatively limited, focused primarily on subjective or intrapersonal outcomes (Ng et al., 2024). A deeper understanding of individuals and their moral (or amoral) behaviours may be gained by studying trait or dispositional profiles among individuals (i.e., sub-groups of persons with specific interpersonal styles). The present study extends this emerging literature by examining whether previously identified profiles of benevolent and malevolent traits generalise to socially embedded outcomes that reflect individuals lived interpersonal experiences. Specifically, rather than focusing on internal psychological processes, this study adopts a person-centred approach to identify dispositional profiles and evaluates their associations with peer-related social behaviours and experiences, including prosocial engagement, relational aggression, victimisation, and social exclusion. In doing so, the study tests the ecological and interpersonal validity of benevolent-malevolent profiles and clarifies how distinct dispositional configurations manifest in everyday social functioning.

The evidence and significance of malevolent personality traits within non-clinical samples is documented by Paulhus and Williams (2002) given links to antisocial behaviour, social and physical aggression, and interpersonal exploitation (Muris et al., 2017). However, only recently were benevolent personality traits proposed as a theoretically grounded counterpoint, reflecting a benevolent social orientation toward others (Kaufman et al., 2019). Specifically, the Light Triad Scale (Kaufman

* Corresponding author.
E-mail address: david.skvarc@deakin.edu.au (D. Skvarc).

https://doi.org/10.1016/j.paid.2026.113854
Received 13 November 2025; Received in revised form 19 April 2026; Accepted 21 April 2026

et al., 2019) has demonstrated promising psychometric properties across a range of populations, with studies demonstrating its discriminant validity from the Dark Triad and broader personality frameworks (Jones & Paulhus, 2014). However, whether Light-Dark trait configurations predict meaningful differences in social experiences and interpersonal outcomes remains an open empirical question, particularly when examined at the level of person-centred profiles rather than isolated traits.

Research in forensic psychology has shown the value of using a person-centred approach for studying subtypes of psychopathic personality (Neumann et al., 2024; Roy et al., 2023); however, little research has sought to uncover aversive (or affiliative) subtype profiles among individuals with malevolent vs. benevolent dispositions within the general population. Benevolent and malevolent trait configurations may shape interpersonal behaviour through distinct mechanisms: benevolent traits promote genuine prosociality via empathy and moral regulation, whereas malevolent traits can motivate strategic prosociality when it serves self-interest or reputational goals (Jones & Paulhus, 2014). Such motivational conflicts are central to understanding how ostensibly similar behaviours arise from divergent dispositional bases. For example, a person who highly values others would be more predisposed to engaging in prosocial behaviours, whereas we would expect the opposite from a person with low empathy. However, it is less clear what the association between prosocial behaviour and a manipulative disposition may be, since the motivation for such behaviours may be grounded in self-aggrandisement rather than altruism. Recently, Patafio et al. (2025) demonstrated differential associations of benevolent and malevolent traits with a range of interpersonal outcomes in a large community sample, with malevolent features emerging as particularly strong predictors of relational aggression, social exclusion, and the strategic use of prosocial behaviour. While this variable-centred regression approach was well suited to identifying the unique contributions of individual traits, it cannot determine how benevolent and malevolent characteristics co-occur within individuals. While factorial designs interaction effects can explore such trait combinations to some degree, they are practically limited to small numbers of indicator variables and assume uniform trait levels across individuals. In contrast, person-centred approaches directly model configurations of traits, allowing examination of whether variation in the light-dark personality space is organised into qualitatively distinct profiles.

Recently, Neumann et al. (2020) conducted a multi-sample person-centred study, which included two large general population samples as well as a supplementary sample of U.S. Senators. These investigators uncovered three latent classes (or subtypes) with each reflecting relatively distinct profiles of malevolent vs. benevolent dispositions. One of the profiles consisting of higher benevolent and lower malevolent features was associated with a range of positive correlates (e.g., self-esteem, empathy), suggestive of greater psychological maturation (Neumann et al., 2020), while a profile with elevated malevolent and lower benevolent features reported lower life satisfaction and a readiness to do harm to others. However, while most distal outcomes examined by Neumann and colleagues were based around psychological processes and beliefs (with the notable exception of general aggression), the comparative experiences of social relationships within profiles were unknown. Moral dispositions, and their composition, might theoretically be expected to be related to more socially oriented behavioural experiences, such as experiences of relational aggression where the target is someone's relationships, prosocial behaviours where there are inherent moral impacts, or sensitivity to peer rejection (perception of peer exclusion). Further, missing from much recent work are trait sadism and vulnerable narcissism. Sadism and vulnerable narcissism index theoretically distinct forms of malevolence and narcissistic functioning that are not captured by the Dark Triad alone. Sadism is moderately correlated with dark traits but reflects enjoyment of others' suffering rather than instrumental antagonism (Bonfá-Araujo et al., 2022), whereas vulnerable narcissism is empirically separable from grandiose narcissism (e.g. Bryce et al., 2023) and associated with heightened sensitivity to rejection and interpersonal withdrawal (Abdelrahman et al., 2024). Including these constructs as distal outcomes therefore provides a stringent test of whether light–dark personality profiles reflect qualitatively meaningful interpersonal styles rather than simple differences in trait magnitude.

Building on this work, the present study applies latent profile analysis (LPA) to the joint light–dark trait space to test whether distinct benevolent, balanced, and malevolent profiles can be identified and whether these profiles show meaningful differences in relational aggression perpetration and victimisation. We intended to replicate earlier work by Neumann et al. (2020), and extend our understanding of how these profiles might associate with other traits and behaviours that have yet to be examined. In addition to the core Light and Dark Triad personality traits, we incorporated closely related constructs such as sadism and vulnerable narcissism to provide a more comprehensive view of personality. This study also examined how these profiles were associated with relational aggression perpetration, social exclusion, prosocial behaviour, victimisation, and relevant demographic variables (age, gender, student status, employment type) where available, aiming to advance our understanding of how specific configurations of affiliative and aversive dispositions shape social behaviours.

1. Method

1.1. Sample

A convenience sample of 2332 Australian adults aged 18–82 years ($M = 39.45$, $SD = 17.12$) were recruited in three independent cohorts in 2021. Two cohorts were recruited from an Australian University in 2021 (Trimester 1, $n = 613$; Trimester 2 = 707), and a third through social media ($n = 1012$) during the second half of the same year. All participants provided positive confirmation of informed consent before commencing participation. Most participants identified as female ($n = 1572$, 67.4%) or male ($n = 721$, 30.9%). Between 83 and 85% completed all responses for the benevolent and malevolent personality scales ($n = 1939$ to 1997) with no difference in completeness between samples. In accordance with institutional ethics guidelines at the university, RA victimisation was only measured in the Trimester 2 and social media samples ($n = 1461$, 85% of the two samples combined, no differences in response rate), and exclusion and sadism were only measured in the social media sample. We obtained complete data for exclusion and sadism for $n = 829$ (82%) and $n = 653$ (65%) of the social media sample, respectively.

1.2. Aggression constructs (relational aggression perpetration, victimisation, and exclusion)

Subscales within the Self-Report of Aggression and Social Behaviour Measure (SRASBM; Morales & Crick, 1998) were used to measure all aggression constructs. Nine items examined RA perpetration (e.g., "I have spread rumours about a person just to be mean"), and four items each examined RA victimisation (e.g., "A friend of mine has gone 'behind my back' and shared private information about me with other people"), and the perceptions of exclusion (e.g., "I get mad or upset if a friend wants to be close friends with someone else"). Responses were scored on a 7-point Likert scale ranging from 1 (not at all true) to 7 (very true), with higher scores indicating higher levels of RA perpetration, RA victimisation, or the experience of exclusion (referred to as "exclusion" from here on), respectively. Scale reliability was good for RA perpetration ($\alpha = 0.79$, $\omega = 0.81$) and RA victimisation ($\alpha = 0.82$, $\omega = 0.83$), and considered acceptable for Exclusion ($\alpha = 0.71$, $\omega = 0.76$).

1.3. Benevolent personality constructs

Benevolent personality was measured via the 12-item Light Triad

Scale (LTS; Kaufman et al., 2019). The LTS contains three subscales, each containing four items, measuring Faith in Humanity (e.g., "I tend to see the best in people"), Humanism (e.g., "I tend to treat others as valuable"), and Kantianism (e.g., "I prefer honesty over charm"). The current study utilised this scale in two ways: as an overall measure of benevolent personality (whereby all items were averaged into a total score), and for each of the three subscales individually (whereby the items for each trait were averaged into a total score for each trait separately). Responses were scored on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree), with higher scores indicating greater benevolence. Reliability was acceptable for the scale overall ($\alpha = 0.74$, $\omega = 0.76$), but 'questionable' for faith in humanity ($\alpha = 0.66$, $\omega = 0.69$) and humanism ($\alpha = 0.66$, $\omega = 0.67$) subscales, and 'poor' for the Kantianism subscale ($\alpha = 0.50$, $\omega = 0.51$). However, an additional examination of the mean inter-item correlation for Kantianism ($r = 0.20$) suggests that the scale falls well within the expected range for an acceptable unidimensional measure (Clark & Watson, 2019).

Prosocial behaviour was measured by averaging the 11-item prosocial behaviour subscale within the SRASBM (Morales & Crick, 1998; e.g., "I make other people feel welcome"). Responses are scored on a 7-point Likert scale ranging from 1 (not at all true) to 7 (very true), with higher scores indicating greater prosociality. Scale reliability was good in the current study ($\alpha = 0.81$, $\omega = 0.84$).

1.4. Malevolent personality constructs

The 27-item Short Dark Triad (Jones & Paulhus, 2014) was used to measure Machiavellianism (e.g., "It's not wise to tell your secrets"), psychopathy (e.g., "People who mess with me always regret it"), and grandiose narcissism (e.g., "I insist on getting the respect I deserve"). Each subscale contained nine items, which were averaged into a total score for each trait separately. Responses are scored on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree), with higher scores indicating higher levels of each malevolent trait. Reliability was acceptable for all subscales: Machiavellianism ($\alpha = 0.74$, $\omega = 0.75$), psychopathy ($\alpha = 0.73$, $\omega = 0.76$), grandiose narcissism ($\alpha = 0.71$, $\omega = 0.72$).

The 10-item Hypersensitivity Narcissism Scale (Hendin & Cheek, 1997) was used to measure vulnerable narcissism, whereby items were averaged to create a total vulnerable narcissism score (e.g., "I often interpret the remarks of others in a personal way"). Responses are scored on a 5-point Likert scale ranging from 1 (very uncharacteristic or untrue, strongly disagree) to 5 (very characteristic or true, strongly agree), with higher scores indicating greater vulnerable narcissism characteristics. Scale reliability was acceptable in the current study; $\alpha = 0.75$, $\omega = 0.75$.

The 10-item Short Sadistic Impulse Scale (O'Meara et al., 2011) was used to measure Sadism, with all items averaged to create a total Sadism score (e.g., "People would enjoy hurting others if they gave it a go"). Items are scored dichotomously, with participants responding either "like me" (1) or "unlike me" (0) to each statement. Higher scores indicate greater sadistic tendencies. Scale reliability was good in the current study; $\alpha = 0.79$, $\omega = 0.84$.

1.5. Analysis strategy

1.5.1. Gaussian mixture modelling

We performed a latent profile analysis (LPA) in MPLUS to provide initial specification and performed confirmatory Gaussian mixture modelling including three-step BCH auxiliary checks using the mclust (Scrucca et al., 2023) and MASS (Venables & Ripley, 2002) packages in R. We used six continuous indicator variables: the benevolent traits (Faith, Humanism, and Kantianism), and the malevolent traits (Machiavellianism, Psychopathy, and Grandiose Narcissism), and explored two to four class solutions. Given the intercorrelation among the Light and Dark Triad indicators (Neumann et al., 2020), solutions assuming independent (zero) covariances were not pursued. Inspection revealed significant deviation from normality on all indicator variables except for grandiose narcissism (negative skew for Light Triad, positive for dark), and so these variables were transformed using the 'bestNormalize' package (Peterson, 2021) before modelling. We also performed iterative analysis comparing number of classes with variance-covariance parameters, and selected the model based on BIC and ICL, entropy, highest classification accuracy, sufficient group size (i.e. $\geq$5%), and best theoretical sense based on Neumann et al. (2020). Our sample of 2003, with 1800 completing all indicator variables, is suitably powered for our expected high degree of separation between theoretically expected classes (see Nylund-Gibson & Choi, 2018).

To compare classes on distal outcomes (vulnerable narcissism, prosocial behaviour, RA perpetration and victimisation, exclusion perceptions, and sadism) while adjusting for age, gender, employment, and sample, we used the BCH three-step correction. We first residualised each distal outcome on the covariates and then applied BCH weights in an intercept-only model to obtain covariate-adjusted, misclassification-corrected class means (Asparouhov & Muthén, 2014).This approach corrects for classification error and yields covariate-adjusted class means. We calculated standardised mean differences (SMD) for between-group pairwise differences. Ferguson (2009) conservatively suggests that SMD = 0.41, 1.15, and 2.7 represent small, moderate, and large effects respectively. For variables measured across all samples, missingness was low and sporadic within samples. 'Sample' was included as a covariate in adjusted analyses to account for unmeasured confounds but removed in models where variables were measured in only one sample. Analyses used all available cases per outcome (listwise deletion), which is appropriate given the low non-design missingness and inclusion of sample as a covariate.

2. Results

The VEE model with three classes (70.6%, 20.6%, and 8.7% of the sample) was selected as the final solution. This model had the most optimal BIC and ICL among all models, good entropy (80%) and good average posterior probability (91.6%). The 3-class VEE model was therefore retained as the most parsimonious, stable, and theoretically coherent solution (Supplemental Table 1).

We label the classes as "Benevolent" (higher benevolent/lower malevolent traits; 20.7%), "Malevolent" (inverse pattern; 8.6%), and "Balanced" (moderate levels on both; 70.6%). Demographic and outcome characteristics are shown in Tables 1–4 and Fig. 1. While most BCH and raw means did not substantially differ, RA victimisation was significantly impacted by the BCH adjustment (Supplemental Table 2. Output for between-class pairwise comparisons is provided in Table 5. The Benevolent class reported significantly higher benevolent characteristics (Light Triad, prosocial behaviour) and lower malevolent

Table 1
Descriptives of total sample.

	N	Missing	Mean	SD	Min	Max
Age (years)	2332	0	39.45	17.12	18	82
Vulnerable narcissism	1897	435	2.89	0.62	1	5
Light Triad (total)	1989	343	3.99	0.46	1.25	5
Faith	1997	335	3.61	0.70	1	5
Humanism	1997	335	4.17	0.54	1.25	5
Kantianism	2000	332	4.21	0.57	1	5
Prosocial behaviour	2004	328	5.80	0.70	1.18	7
Proactive RA	2026	306	1.49	0.67	1	7
RA total perpetration	2016	316	1.61	0.66	1	6.78
RA victimisation	1461	871	2.74	1.60	1	7
Exclusion	829	1503	2.53	1.01	1	6.75
Machiavellianism	1941	391	2.68	0.58	1	5
Grandiose Narcissism	1939	393	2.89	0.39	1.67	5
Psychopathy	1939	393	2.29	0.46	1	5
Sadism	653	1679	1.44	0.49	1	3.70

Table 2
Frequencies of total sample.

	Label	N	%
Sample	2021 T1 (student)	613	26.3%
	2021 T2 (student)	707	30.3%
	2022 (social media)	1012	43.4%
Gender	Male	721	30.9%
	Female	1572	67.4%
	Non-binary	11	0.5%
	Prefer to self-describe	18	0.8%
	Prefer not to say	10	0.4%
Student	Higher education	1256	53.9%
	Not in higher education	1075	46.1%
Employment	Full time employed	556	32.4%
	Part time/casual	731	42.5%
	Unemployed	431	25.1%
Class	Benevolent	393	20.70%
	Balanced	1343	70.60%
	Malevolent	165	8.7%

characteristics (Dark Triad, sadism, vulnerable narcissism, relational aggression, and exclusion) than the Balanced and Malevolent classes (Table 4). The Balanced class followed a similar pattern. RA victimisation were highest for the Benevolent and lowest for the Malevolent classes, and grandiose narcissism did not differ between classes. Effect sizes ranged from small to large but were typically robust ($p < .001$).

3. Discussion

Our aim was to investigate combinations of personality traits through the lens of malevolent and benevolent dispositions and extend from other work by comparing self-reports of relational aggression victimisation and perpetration, exclusion, and prosocial behaviours, as well as including the additional theoretically important malevolent traits of vulnerable narcissism and sadism. Our findings are broadly aligned with our expectations that three distinct clusters would emerge: (1) those characterised predominantly by higher levels of benevolent traits and lower levels of malevolent traits, (2) those characterised by the inverse combination of traits, and (3) those with moderate levels across all traits. Around one-fifth of our sample were classified as Benevolent, just under three-quarters as Balanced, and around one-twelfth as Malevolent. The Benevolent class skewed strongly female, whereas the ratio of male to female in the Malevolent and Balanced classes mirrored the overall sample. Likewise, our three classes broadly follow expectations regarding arguably moral behaviours. However, some interesting patterns with regards more socially complicated experiences (e.g., victimisation) suggests contextually important factors such as individual motivations might explain deviations from a basic

Table 3
Class characteristics (categorical).

		Balanced		Benevolent		Malevolent		*Proportional comparisons*		
		N	%	N	%	N	%	B vs Mid	B vs Mal	Mid vs Mal
Sample	2021 T1	353	26.3%	121	30.8%	51	30.9%	0.089	1.000	0.241
	2021 T2	386	28.7%	166	42.2%	50	30.3%	<0.001	0.011	0.744
	2022	604	45.0%	106	27.0%	64	38.8%	<0.001	0.008	0.154
Gender		459	34.2%	66	16.8%	52	31.5%	<0.001	<0.001	0.552
	Male	867	64.6%	318	80.9%	108	65.5%	<0.001	<0.001	0.888
	Female	2	0.1%	4	1.0%	1	0.6%			
	Non-binary	8	0.6%	4	1.0%	3	1.8%			
	Prefer to self-describe	7	0.5%	1	0.3%	1	0.6%			
	Prefer not to say	682	50.8%	279	71.2%	103	62.4%	<0.001	0.053	0.006
Student		661	49.2%	113	28.8%	62	37.6%			
	Yes	332	26.1%	63	16.6%	44	27.5%	<0.001	0.006	0.772
	No	402	31.6%	154	40.6%	49	30.6%	0.001	0.036	0.877
Employment		256	20.1%	54	14.2%	21	13.1%	0.013	0.835	0.045
	Full time	353	26.3%	121	30.8%	51	30.9%	0.089	1.000	0.241
	Part time	386	28.7%	166	42.2%	50	30.3%	<0.001	0.011	0.744
	Unemployed	604	45.0%	106	27.0%	64	38.8%	<0.001	0.008	0.154
	Student	283	22.2%	108	28.5%	46	28.7%	0.014	1	0.08

Note: B = Benevolent, Bal = Balanced, Mal = Malevolent classes. Columns B vs Bal, B vs Mal, and Bal vs Mal represent the unadjusted *p*-values for pairwise comparisons of proportions between the classes Benevolent and Balanced, Benevolent and Malevolent, and Balanced and Malevolent. Gender comparisons outside the gender binary were not performed due to small cell size. 2021 samples were students recruited over two trimesters (T1 and T2).

Table 4
Class descriptives and continuous characteristics.

	Balanced				*Benevolent*				*Malevolent*			
	M	SD	Min	Max	M	SD	Min	Max	M	SD	Min	Max
Age (years)	40.97	17.15	18	82	33.55	15.12	18	82	36.47	14.58	18	69
Vulnerable narcissism	2.86	0.58	1	4.7	2.83	0.67	1	4.9	3.27	0.62	1.7	4.7
Light Triad (total)	3.94	0.35	2.5	4.92	4.45	0.32	3.42	5	3.31	0.44	1.25	4.08
Faith	3.65	0.52	2.5	5	4.03	0.61	2.25	5	2.19	0.39	1	2.75
Humanism	4.01	0.42	1.5	4.75	4.87	0.14	4.5	5	3.78	0.64	1.25	4.75
Kantianism	4.17	0.54	1.75	5	4.45	0.54	2	5	3.97	0.7	1	5
Prosocial behaviour	5.72	0.64	1.18	7	6.25	0.51	4.45	7	5.39	0.85	2.45	7
Proactive RA	1.47	0.59	1	5	1.4	0.58	1	4.5	1.76	0.93	1	6.75
RA total perpetration	1.58	0.58	1	5	1.5	0.59	1	4.22	1.96	0.94	1	6.78
RA victimisation	2.63	1.52	1	7	3.01	1.83	1	7	3.19	1.63	1	7
Exclusion	2.48	0.9	1	6.75	2.48	1.18	1	6.75	3.05	1.28	1.25	6
Machiavellianism	2.68	0.54	1.11	4.44	2.53	0.62	1	4.89	3.01	0.61	1.67	4.78
Grandiose narcissism	2.89	0.37	1.67	3.89	2.92	0.46	1.67	4.33	2.84	0.41	2	4
Psychopathy	2.27	0.43	1	4.33	2.22	0.47	1.11	3.89	2.55	0.52	1.22	4.11
Sadism	1.68	0.39	1	3.5	1.56	0.32	1	3	1.92	0.52	1.2	3.2

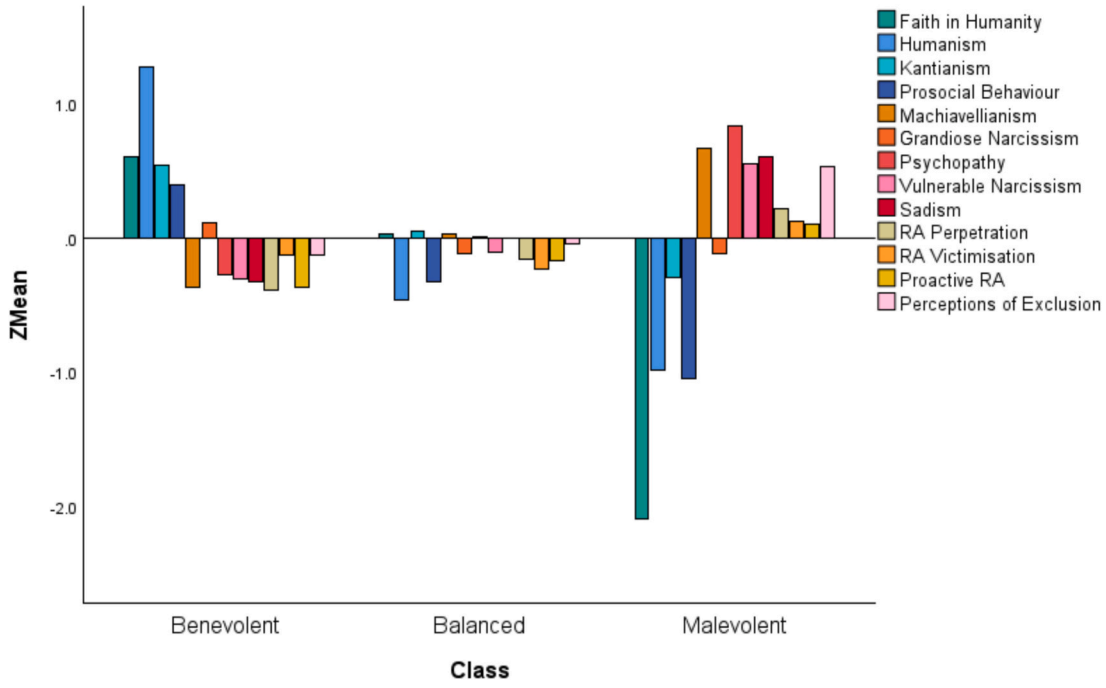

Fig. 1. Indicator and distal outcomes across classes. Scores are standardised to account for different scales.

Table 5
Pairwise comparisons for distal and indicator variables.

	Group	N	Adj. mean	Raw mean	SD	Benevolent − Balanced	Benevolent − Malevolent	Balanced − Malevolent
Prosocial behaviour	Benevolent	382	6.19	6.25	0.51	MD = −0.45, SMD = −0.65 [−0.71, −0.6], $t(1778) = -23.53$, $p < .001$	MD = −0.82, SMD = −1.2 [−1.26, −1.13], $t(1778) = -37.23$, $p < .001$	MD = −0.37, SMD = −0.54 [−0.61, −0.47], $t(1778) = -15.5$, $p < .001$
	Balanced	1325	5.74	5.72	0.64			
	Malevolent	164	5.37	5.39	0.85			
RA perpetration	Benevolent	387	1.48	1.5	0.59	MD = 0.11, SMD = 0.17 [0.11, 0.23], $t(1793) = 5.74$, $p < .001$	MD = 0.41, SMD = 0.64 [0.56, 0.72], $t(1793) = 15.58$, $p < .001$	MD = 0.3, SMD = 0.47 [0.39, 0.55], $t(1793) = 11.88$, $p < .001$
	Balanced	1334	1.59	1.58	0.58			
	Malevolent	165	1.89	1.96	0.94			
Victimisation	Benevolent	269	2.89	3.01	1.83	MD = −0.22, SMD = −0.13 [−0.21, −0.06], $t(1363) = -3.48$, $p < .001$	MD = 0.21, SMD = 0.13 [0.05, 0.21], $t(1363) = 3.18$, p 0.002	MD = 0.42, SMD = 0.26 [0.19, 0.34], $t(1363) = 7.17$, $p < .001$
	Balanced	985	2.67	2.63	1.52			
	Malevolent	114	3.09	3.19	1.63			
Vulnerable narcissism	Benevolent	379	2.79	2.83	0.67	MD = 0.08, SMD = 0.12 [0.06, 0.19], $t(1753) = 3.62$, $p < .001$	MD = 0.45, SMD = 0.74 [0.67, 0.81], $t(1753) = 20.64$, $p < .001$	MD = 0.38, SMD = 0.62 [0.55, 0.68], $t(1753) = 18.9$, $p < .001$
	Balanced	1305	2.87	2.86	0.58			
	Malevolent	160	3.25	3.27	0.62			
Exclusion	Benevolent	104	2.42	2.48	1.18	MD = 0.06, SMD = 0.06 [−0.04, 0.17], $t(764) = 1.17$, p 0.242	MD = 0.51, SMD = 0.52 [0.4, 0.64], $t(764) = 8.57$, $p < .001$	MD = 0.45, SMD = 0.46 [0.35, 0.56], $t(764) = 8.37$, $p < .001$
	Balanced	601	2.48	2.48	0.9			
	Malevolent	64	2.94	3.05	1.28			
Sadism	Benevolent	79	1.6	1.56	0.32	MD = 0.09, SMD = 0.22 [0.12, 0.31], $t(632) = 4.4$, $p < .001$	MD = 0.25, SMD = 0.63 [0.51, 0.75], $t(632) = 10.16$, $p < .001$	MD = 0.17, SMD = 0.41 [0.29, 0.53], $t(632) = 6.64$, $p < .001$
	Balanced	505	1.69	1.68	0.39			
	Malevolent	53	1.85	1.92	0.52			
Faith	Benevolent	379	4.03		0.53	MD = 0.38, SMD = 0.71 [0.58, 0.84], $t(1803) = 10.79$, $p < .001$	MD = 1.83, SMD = 3.43 [3.27, 3.59], $t(1803) = 41.3$, $p < .001$	MD = 1.45, SMD = 2.72 [2.59, 2.84], t (1803) = 41.98, $p < .001$
	Balanced	1273	3.65		0.53			
	Malevolent	160	2.2		0.53			
Humanism	Benevolent	379	4.84		0.4	MD = 0.83, SMD = 2.09 [2.02, 2.16], $t(1803) = 56.21$, $p < .001$	MD = 1.07, SMD = 2.69 [2.45, 2.94], $t(1803) = 21.52$, $p < .001$	MD = 0.24, SMD = 0.6 [0.35, 0.85], $t(1803) = 4.71$, $p < .001$
	Balanced	1273	4.01		0.4			
	Malevolent	160	3.78		0.4			
Kantianism	Benevolent	379	4.46		0.54	MD = 0.28, SMD = 0.52 [0.4, 0.63], $t(1803) = 8.73$, $p < .001$	MD = 0.47, SMD = 0.86 [0.64, 1.09], $t(1803) = 7.56$, $p < .001$	MD = 0.19, SMD = 0.35 [0.14, 0.55], $t(1803) = 3.27$, p 0.001
	Balanced	1273	4.18		0.54			
	Malevolent	160	3.99		0.54			
Machiavellianism	Benevolent	379	2.54		0.56	MD = −0.13, SMD = −0.23 [−0.36, −0.1], $t(1803) = -3.57$, $p < .001$	MD = −0.46, SMD = −0.82 [−1.02, −0.62], $t(1803) = -7.95$, $p < .001$	MD = −0.33, SMD = −0.59 [−0.76, −0.41], $t(1803) = -6.61$, $p < .001$
	Balanced	1273	2.67		0.56			
	Malevolent	160	3		0.56			
Grandiose narcissism	Benevolent	379	2.91		0.38	MD = 0.01, SMD = 0.03 [−0.1, 0.16], $t(1803) = 0.46$, p 0.643	MD = 0.08, SMD = 0.2 [−0.01, 0.4], $t(1803) = 1.89$, p 0.059	MD = 0.06, SMD = 0.17 [−0.01, 0.34], $t(1803) = 1.85$, p 0.064
	Balanced	1273	2.9		0.38			
	Malevolent	160	2.83		0.38			
Psychopathy	Benevolent	379	2.22		0.44	MD = −0.05, SMD = −0.12 [−0.24, 0], $t(1803) = -2$, p 0.046	MD = −0.33, SMD = −0.74 [−0.95, −0.54], $t(1803) = -7.04$, $p < .001$	MD = −0.27, SMD = −0.62 [−0.81, −0.43], $t(1803) = -6.54$, $p < .001$
	Balanced	1273	2.27		0.44			
	Malevolent	160	2.55		0.44			

Note: Adj. Means are BCH-adjusted for distal outcomes, and MANCOVA estimated marginal means for light and dark traits. Covariate adjustments are for age, gender, employment, and sample. All *p*-values for pairwise comparisons are unadjusted for multiple tests.

benevolent, balanced and malevolent behavioural profile. Implications for understanding who is perpetrating socially disruptive behaviours points toward an approach which combines both benevolent/malevolent profiling as well as contextual motivations.

Behaviourally, our findings also largely align with expectations. Prosocial behaviour was highest for the Benevolent class and lowest for the Malevolent class, with the reverse pattern observed for relational aggression perpetration. Vulnerable narcissism and sadism also differed across classes in the expected directions. Our finding that grandiose narcissism was diminished in the Malevolent class was unexpected, though the magnitude of the difference was small and would not have persisted with error rate adjustment. Effectively, grandiose narcissism did not differ across classes in the adjusted models, which distinguishes it from the other Dark Triad traits. This suggests that while Machiavellianism and psychopathy are core to the malevolent profile, grandiose narcissism may function differently, perhaps reflecting the class's shared reliance on self-enhancement strategies (Benevolent and Balanced classes also score near normative levels). Grandiose narcissism may be less aversive in interpersonal contexts than Machiavellianism or psychopathy, and might even confer social advantages (e.g., confidence, assertiveness) that are not uniquely tied to a malevolent profile. Thus, all three classes – including the Benevolent – may endorse grandiose narcissism at similar levels because it serves a self-presentational function regardless of underlying prosocial or antagonistic orientation. Alternatively, the distinctive behavioural patterns associated with vulnerable narcissism (e.g. Bryce et al., 2023) may indicate a more suitable indicator of aversive profiles compared to grandiose. Demonstrably, sadism – arguably less socially desirable and less susceptible to self-presentational distortion – showed clear differentiation across all three classes. Within this finding, it should be noted that the dichotomous nature of the sadism scale may have influenced findings as self-report measures can exacerbate social desirability effects (Stöber et al., 2002). This highlights the value of incorporating multi-informant or behavioural measures of social exclusion in future research.

Relational aggression victimisation was highest in the Malevolent class and lowest in the Balanced class, with the Benevolent class falling in between. This pattern suggests different social dynamics across profiles. Malevolent individuals, who also reported the highest perpetration, may experience victimisation as a consequence of their own provocative or retaliatory behaviour consistent with a 'provocation-victimisation' cycle (Widom, 1989). Their elevated vulnerable narcissism may further heighten sensitivity to perceived slights, increasing both conflict and reported victimisation. In contrast, the Balanced class reported the lowest victimisation alongside moderate perpetration, suggesting better social skills, conflict management, or protective social networks that buffer against retaliation. The Benevolent class, despite being the least aggressive, showed intermediate victimisation. This may reflect their high trust, forgiveness, and prosocial orientation, which could make them more vulnerable to exploitation by others — a 'cooperative victim' pattern. The weaker perpetration-victimisation correlation within the Balanced class (if observed) could indicate greater heterogeneity, with some members resembling Benevolent (victimised due to trust) and others resembling Malevolent (perpetrating due to antagonism), thereby attenuating the overall association. Collectively, these findings suggest that victimisation is not simply a function of one's own aggression but is shaped by complex interpersonal dynamics that differ across latent profiles.

Aversive traits showed only moderate differences between classes (largest SMDs between 0.6 and 0.8), whereas affiliative and prosocial traits differed substantially. The magnitude of these differences suggests that the latent profiles are primarily organised along a dimension of prosocial versus antagonistic worldviews, with the Malevolent class not only acting more aggressively but also holding fundamentally different beliefs about human nature and moral obligation. Such effect sizes are rare in personality research and demonstrate the clinical or practical significance of the classification. The observed pattern supports a unified model in which malevolence is best conceptualised as a deficit in an evolved prosocial baseline, rather than the presence of unique pathological characteristics (Jensen et al., 2014). Dark traits such as narcissism and sadism may represent deficits or attenuations of this baseline, not additive pathological structures (Hepp & Niedtfeld, 2022). Because deficit-based constructs exhibit constrained variance (there are only so many ways to express profoundly low faith in humanity), aversive traits show only moderate discriminative capacity between classes. Conversely, prosociality allows substantial heterogeneity, as benevolent individuals may express humanism and trust through diverse frameworks, generating the large effect sizes observed. Consequently, the Malevolent class is defined not only by moderately elevated dark traits, but more fundamentally by the collapse of light traits.

Our modelling replicates the three-profile solution of Neumann et al. (2020), supporting the Benevolent, Balanced, and Malevolent configurations. Class sizes differ from previous studies, likely due to demographic and sampling differences: our sample is older and has more females (Jonason & Davis, 2018), and we recruited university students and community members via social media, whereas Neumann used mass online and popular psychology websites, which may attract individuals motivated to explore light and dark traits. The Malevolent class was the smallest (8.7%), consistent with the rarity of high Dark Triad and low Light Triad traits. Its demographic profile diverges from Neumann: in our sample it is age-intermediate (36 years) and gender-balanced (33% male), whereas Neumann's dark trait class was youngest and predominantly male. The Benevolent class is youngest and most female (81%). These differences likely stem from our broader community recruitment, which yielded an overall sample with ~33% male — a baseline mirrored by the Malevolent and Balanced classes. Neumann's strong male skew in the dark trait class may therefore be specific to their online self-selected sample rather than a universal feature of the malevolent profile.

3.1. Limitations

The cross-sectional design precludes causal inferences. Self-report measures may be subject to social desirability, particularly for grandiose narcissism, though the consistency of our results with prior work mitigates this concern. We did not adjust for multiple testing; however, most comparisons were highly significant ($p < .001$) and would survive correction. The key exception is grandiose narcissism, which has been discussed. As previously discussed, sampling context likely contributed to differences in class proportions compared with past studies. One indicator, Kantianism, showed modest internal consistency ($\alpha = 0.50$), though its mean inter-item correlation was acceptable and between-class differences were significant (smallest SMD ≈ 0.3). In a person-centred framework, measurement error generally attenuates rather than inflates class differences, so results involving this indicator are likely conservative. Nevertheless, measurement imprecision could still affect profile estimation; these findings should therefore be interpreted with appropriate caution.

The robust standardised mean differences observed for the indicator variables reflect the nature of latent profile analysis, which maximises separation on the indicators used to define the classes. These differences are descriptive of the class structure rather than independent effect sizes and should not be interpreted as evidence of strong external associations. Finally, although allowing within-class correlations (VEE) is theoretically justified given known associations among Light and Dark Triad traits and resulted in good model fit, the classes should be understood as idealised patterns along largely continuous dimensions, and class membership as probabilistic. Except for victimisation, BCH-corrected means were very similar to naive estimates, suggesting that classification uncertainty had minimal impact on group-level comparisons. Future research may benefit from integrating person-centred and variable-centred approaches to capture both profile structure and underlying dimensional variation.

4. Conclusions

This study extends research on benevolent and malevolent traits by identifying three distinct profiles with meaningful trait and behavioural correlates. While regression analyses (Patafio et al., 2025) establish the relative strength of light vs. dark predictors, our person-centred approach reveals how these traits co-occur within individuals, identifying qualitatively distinct configurations. As expected, benevolent individuals reported higher prosociality and lower relational aggression; malevolent individuals showed the opposite. Grandiosity does not appear to reliably differentiate the classes unlike the other aversive traits and may even reflect under-reporting. Relational aggression perpetration increased across classes, whereas victimisation was elevated in the Benevolent and Malevolent classes, suggesting victimisation may be exacerbated by both affiliative and averse traits in combination. These findings highlight the value of examining light and dark traits together, alongside demographic and contextual influences. Recognising trait configurations rather than isolated dimensions may help identify individuals at risk of socially disruptive behaviour. Future work using behavioural and multi-informant methods could clarify how these profiles manifest across settings and inform interventions to mitigate the relational costs of malevolent dispositions.

CRediT authorship contribution statement

David Skvarc: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Brittany Patafio:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Richelle Mayshak:** Writing – review & editing, Writing – original draft, Resources, Investigation, Data curation, Conceptualization. **Shannon Hyder:** Writing – review & editing, Writing – original draft, Validation, Software, Resources, Investigation, Formal analysis, Data curation, Conceptualization. **Scott Barry Kaufman:** Writing – review & editing, Writing – original draft, Validation, Investigation, Conceptualization. **Craig Neumann:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Formal analysis, Conceptualization.

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.

Acknowledgements

This research was funded by internal support at the School of Psychology, Deakin University.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.paid.2026.113854.

Data availability

An analytical dataset is uploaded

References

Abdelrahman, R. M., Ahmed, M., Tayim, N., & Kordbagheri, M. (2024). Identification of the core characteristics of vulnerable/hypersensitive narcissism and its association with the dark triad in a large international sample: A network analysis study. *Psychiatric Quarterly, 95*(3), 415–431.

Asparouhov, T., & Muthén, B. (2014). Auxiliary variables in mixture modeling: Using the BCH method in Mplus to estimate a distal outcome model and an arbitrary secondary model. *Mplus Web Notes, 21*(2), 1–22.

Bonfá-Araujo, B., Lima-Costa, A. R., Hauck-Filho, N., & Jonason, P. K. (2022). Considering sadism in the shadow of the Dark Triad traits: A meta-analytic review of the Dark Tetrad. *Personality and Individual Differences, 197*, Article 111767.

Bryce, C. J., Skvarc, D. R., King, R. M., & Hyder, S. (2023). Don't set me off—Grandiose and vulnerable dimensions of narcissism are associated with different forms of aggression: A multivariate regression analysis. *Current Psychology, 42*(12), 10177–10185.

Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment, 31*(12), 1412.

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice, 40*(5), 532.

Hendin, H. M., & Cheek, J. M. (1997). Assessing hypersensitive narcissism: A reexamination of Murray's Narcism Scale. *Journal of Research in Personality, 31*(4), 588–599.

Hepp, J., & Niedtfeld, I. (2022). Prosociality in personality disorders: Status quo and research agenda. *Current Opinion in Psychology, 44*, 208–214.

Jensen, K., Vaish, A., & Schmidt, M. F. (2014). The emergence of human prosociality: Aligning with others through feelings, concerns, and norms. *Frontiers in Psychology, 5*, 822.

Jonason, P. K., & Davis, M. D. (2018). A gender role view of the Dark Triad traits. *Personality and Individual Differences, 125*, 102–105.

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.

Kaufman, S. B., Yaden, D. B., Hyde, E., & Tsukayama, E. (2019). The Light vs. Dark Triad of personality: Contrasting two very different profiles of human nature. *Frontiers in Psychology, 10*, 467.

Morales, J. R., & Crick, N. R. (1998). *Self-report of aggression and social behavior measure. Unpublished measure*. University of Minnesota, Twin Cities Campus.

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

Neumann, C., & Ngo, D. (2025). Malevolent vs. benevolent dispositions and conservative political ideology in the Trump era. *Journal of Research in Personality*. ISSN: 0092-6566, *118*, 104638. https://doi.org/10.1016/j.jrp.2025.104638

Neumann, C. S., Kaufman, S. B., & Ten Brinke, L. (2025). Citizens in democratic countries have more benevolent traits, fewer malevolent traits, and greater well-being. *Scientific Reports, 15*(1), 13346. https://doi.org/10.1038/s41598-025-97001-7

Neumann, C. S., Kaufman, S. B., ten Brinke, L., Yaden, D. B., Hyde, E., & Tsykayama, E. (2020). Light and dark trait subtypes of human personality—A multi-study person-centered approach. *Personality and Individual Differences, 164*, Article 110121.

Neumann, C. S., Salekin, R. T., Commerce, E., Charles, N. E., Barry, C. T., Mendez, B., & Hare, R. D. (2024). Proposed Specifiers for Conduct Disorder (PSCD) scale: A latent profile analysis with at-risk adolescents. *Research on Child and Adolescent Psychopathology, 52*(3), 369–383. https://doi.org/10.1007/s10802-023-01126-0

Ng, N. L., Neumann, C. S., Luke, D. M., & Gawronski, B. (2024). Associations of aversive ('dark') traits and affiliative ('light') traits with moral-dilemma judgments: A preregistered exploratory analysis using the CNI model. *Journal of Research in Personality, 109*, Article 104450. https://doi.org/10.1016/j.jrp.2023.104450

Nylund-Gibson, K., & Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science, 4*(4), 440.

O'Meara, A., Davies, J., & Hammond, S. (2011). The psychometric properties and utility of the Short Sadistic Impulse Scale (SSIS). *Psychological Assessment, 23*(2), 523.

Patafio, B., Skvarc, D., Mayshak, R., Harries, T., Curtis, A., Benstead, M., … Hyder, S. (2025). Dark and light personalities: A utilitarian perspective on their impact on relational aggression. *Personality and Individual Differences, 242*, Article 113209.

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

Peterson, R. A. (2021). Finding optimal normalizing transformations via bestNormalize. *The R Journal, 13*(1), 310–329. https://doi.org/10.32614/RJ-2021-041

Roy, S., Neumann, C. S., & Hare, R. D. (2023). Validating latent profiles of the Psychopathy Checklist-Revised with a large sample of incarcerated men. *Personality Disorders, Theory, Research, and Treatment, 14*(6), 649–659. https://doi.org/10.1037/per0000633

Scrucca, L., Fraley, C., Murphy, T. B., & Raftery, A. E. (2023). *Model-based clustering, classification, and density estimation using mclust in R*. Chapman and Hall/CRC. https://doi.org/10.1201/9781003277965. ISBN 978-1032234953 https://mclust-org.github.io/book/.

Skvarc, D., Patafio, B., Hyder, S., Harries, T., Curtis, A., Benstead, M., & Mayshak, R. (2025). Relational aggression and its association with other forms of aggression: An applied latent profile analysis. *Behavioral Science, 15*(12), 1736.

Stöber, J., Dette, D. E., & Musch, J. (2002). Comparing continuous and dichotomous scoring of the Balanced Inventory of Desirable Responding. *Journal of Personality Assessment, 78*(2), 370–389.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). New York: Springer. ISBN 0-387-95457-0.

Widom, C. S. (1989). The Cycle of Violence. *Science, 244*(4901), 160–166. http://www.jstor.org/stable/170278.