Introduction and Validation of the Short Antinatalism Scale (S-ANS)

Philipp Schoenegger[1], Maximilian Maier[2], and Rafael Valdece Sousa Bastos[3]

[1]Department of Philosophy/School of Economics and Finance, University of St Andrews, Fife, UK
[2]Division of Psychology and Language Sciences, University College London, UK
[3]Department of Psychology, São Francisco University, São Paulo, Brazil

Abstract: Antinatalism is the view that procreation is morally wrong. This paper introduces and validates the Short Antinatalism Scale (S-ANS) that allows researchers to measure antinatalist views. We conducted four preregistered studies with a total of 1,088 participants. First, we ran a study on Prolific (*N* = 296) and conducted an exploratory factor analysis of an initial scale including 22 items drawn from the philosophical literature on antinatalism. In Study 2, we conducted a confirmatory factor analysis of a reduced 12-item scale, also on Prolific (*N* = 396). Based on a Mokken scale analysis, we further reduced the scale to a 5-item version which we tested in a second confirmatory factor analysis, Study 3, on Prolific (*N* = 297), where we also aimed to provide evidence of validity. The results indicated excellent model fit (RMSEA = 0.012) and evidence for validity (with life satisfaction, affective empathy, and conservatism correlating negatively with antinatalism). Lastly, we conducted Study 4 with a sample of self-identified antinatalists on Reddit (*N* = 99) to provide additional evidence of validity. We find that the instrument is measurement invariant between self-described antinatalists and the general population and that antinatalists score significantly higher on the scale (*d* = 2.80). This provides evidence in favor of reliability and validity with respect to the final 5-item Short Antinatalism Scale (S-ANS). We hope that the S-ANS, which is freely available to all researchers, advances rigorous research into antinatalism and its determinants across a variety of fields that relate to the value of life and procreation.

Keywords: antinatalism, scale development, personality, empathy, life satisfaction

Antinatalism is the philosophical position that procreation is morally wrong. This claim was already discussed in the 19th century by John Stuart Mill (1892) who stated that "to bring a child into existence without a fair prospect of being able to [care for it] is a moral crime" (Mill, 1892, p. 62). More recent philosophical work has provided various justifications for the general claim that bringing a child into existence is always morally wrong. For example, Shiffrin (1999) argues that there is a moral challenge inherent in being brought into a state (i.e., existence) that one has not consented to, and children do not have the ability to consent to being born. Benatar (1997, 2008) advances a different type of argument, according to which procreation is morally wrong for the following reasons: The expected amount of good in one's life is never sufficiently high to outweigh the expected amount of bad, which implies that bringing someone into existence is always morally wrong. This argument by Benatar (1997) is called the asymmetry argument: Their main claim there is that there is a central asymmetry between good and bad (pleasures and pains) regarding a person that might come into existence. Benatar's claim is that while the absence of pain is good, the absence of pleasure is not bad. As such, while it would be good to come into existence and experience pleasure, nothing bad happens when one does not come into existence. However, coming into existence and experiencing pain would be bad, but not experiencing this pain would be good. As such, it is better not to come into existence and any act that does bring a being into existence (i.e., procreation) is morally wrong. Additional justifications of antinatalism are too plenty to discuss in detail in this introduction (see MacIver, 2015; Rulli, 2016; Singh, 2018).

In the contemporary philosophical discourse, general (or a priori) antinatalism is a central question of population ethics, a branch of philosophical ethics that deals with questions concerning the welfare, identities, and/or numbers of people that exist or will exist. While there has been significant philosophical debate relating to these questions (Arrhenius, 2000; Greaves, 2017; Parfit, 1984; Zuber et al., 2021), there has also recently been an upswing of empirical social science work that investigates related questions. For example, Schoenegger and Grodeck (forthcoming) outlined how lay people respond when their ethical intuitions about populations conflict with their endorsed moral principles. Furthermore, Spears (2017) investigated population ethical views regarding policy choices, while Caviola et al. (2022)

offered a more general picture of intuitions about population ethics, ranging from the value of average welfare to the focus on currently existing lives.

While there has not been much further explicit empirical work on population ethics yet, the present paper also fits in with parts of the already existing psychological literature on a broader set of topics. First, a lot of reasoning about the value of life is often captured by overgeneralization of one's own life. This type of reasoning is particularly easily biased in cases where one's own life might be especially difficult, perhaps due to depression. Moore and Fresco (2012) reviewed the literature on depressive realism, i.e., the view that those who are higher in depressive attitudes perceive some aspects of reality more clearly and find that while both depressed and nondepressed individuals exhibit positivity bias, depressed individuals exhibited it to a significantly lower extent; the relationship between antinatalist views and depression has also been established before (Schönegger, 2021). Second, as Caviola et al. (2022) pointed out, reasoning regarding population ethics (and by extension antinatalism) may also be influenced by negativity bias (Rozin & Royzman, 2001), the bias that negative events are judged as outweighing positive events of same magnitude. Negativity bias has also been researched in the context of moral behavior (Riskey & Birnbaum, 1974) and on moral judgments (Tappin & Capraro, 2018). Our present work on antinatalism may thus also inform our understanding of negativity bias and overgeneralisation and thus contribute to the wider psychological literature.

Importantly, while there has not been much psychological work directly on antinatalism, the small empirical literature specifically on antinatalism has increased in scope over the past few years. For example, Brown (2020) recently investigated the relationship between optimism and support for antinatalism, finding that optimism about future children reduces support for antinatalism. Furthermore, Schönegger (2021) showed that antinatalist views stand in a strong relationship to dark triad personality traits and depressive attitudes. Importantly though, all these studies that investigate antinatalism have used ad hoc measures of antinatalism which had no accompanied tests regarding their reliability and validity. This makes comparisons of findings across studies more difficult while also raising concerns about bad measurement practices that put this young literature on potentially shaky methodological foundations (cf. Flake & Fried, 2020; Lilienfeld & Strother, 2020).

The purpose of this paper was to address this lack of a rigorous scale measuring antinatalism. As discussed above, the central construct being measured is antinatalism. In this paper, we develop and present the Short Antinatalism Scale (S-ANS) to measure antinatalist views. The intended use of this measure is to aid further research and thus the uptake by academic researchers. As such, the scale is openly and freely available to all researchers. Our target population is a general lay population, as we are primarily interested in antinatalist attitudes of a general population. The overarching goal of this paper is to provide a short scale while also investigating reliability and validity, which allows researchers studying antinatalism, population ethics, or any related research on bringing new lives into existence to accurately measure antinatalist views. We also take this scale to allow those researching lay opinions on broader topics relating to procreation to accurately control for heterogeneity in antinatalist views, thus allowing for a better delineation of any given.

In the present paper, we report three studies conducted on Prolific ($N = 296$, $N = 396$, and $N = 297$) as well as one study conducted on Reddit with a community of self-described antinatalists ($N = 99$). All four studies have received ethics approval and were preregistered at the Open Science Framework (OSF). The resulting 5-item Short Antinatalism Scale (S-ANS) showed exceptional model fit. Furthermore, we also found strong evidence for construct validity as we have observed negative relationships of antinatalist views to variables where one would expect such a relationship. Specifically, we study the relationships to life satisfaction (Margolis et al., 2019) – because those lower in life satisfaction may have a more negative outlook on life and on procreation, conservatism (Everett, 2013) – because standard conservative values of the family are pronatalist, and empathy (Carré et al., 2013) – as there has been previous research showing a negative relationship between dark triad traits and as such low empathy on the one hand and antinatalism on the other (Schönegger, 2021). Studying a community of self-described antinatalists in our fourth study ($N = 99$), we found that self-described antinatalists also scored substantially higher on the S-ANS than the general population. We take our paper to provide evidence in favor of reliability and validity of the S-ANS that measures antinatalist views, which is now freely available for all researchers.

Study 1

Generating the Initial Item Pool

The authors generated an initial pool of 22 items, the final version of which can be found in Table 1. To arrive at these items, PS and MM developed two sets of items independently based on the existing antinatalism literature. In the following discussion, we merged these two lists, removed duplicates, and added additional variations and items based on the discussion. We also added reverse-scored items based on the initial set and the discussion. Then, we sent these items to an expert in the field, the philosopher Theron Pummer. Based on his comments, we further

Table 1. Factor loadings for 22-item scale and 20-item scale

Items	Model 1		Model 2	
	FL	h²	FL	h²
(1) Because humans cause so much harm to the environment, it is morally wrong to have children.	.86	.75	.86	.75
(2) Bringing a person into existence is usually a considerable net benefit to that person. (R)	−.50	.25	−.49	.24
(3) In becoming a parent, one creates not only a child and a relationship, but oneself, one creates a new ideally better self-identity. (R)	−.37	.14		
(4) It is better not to exist than to exist because only in existence can there be pain. If one does not exist, one cannot be in pain. As such, it is morally wrong to have children.	.84	.71	.84	.71
(5) It is usually morally permissible to bring a child into this world (R)	−.57	.33	−.58	.33
(6) Bringing someone into existence means that they will eventually suffer in one way or another, which is why it is morally wrong to have children.	.88	.77	.88	.78
(7) It is morally wrong to bring a child into this world because a child cannot consent to being born.	.86	.74	.86	.75
(8) It is morally wrong for those unable to care for their children to have children.	.34	.11		
(9) Bringing a new child into this world is a morally good thing. (R)	−.60	.37	−.60	.36
(10) Irrespective of how one's life goes, it is always better to exist than to not have existed. (R)	−.52	.27	−.52	.27
(11) Having a child in the developed world is morally wrong as it impacts the environment at least to the same extent as excessive consumption behavior.	.85	.72	.85	.72
(12) The absence of pleasure is not bad unless there is somebody for whom this absence is a deprivation. This is why not being born is not bad.	.43	.19	.43	.19
(13) People have a moral right to have children. (R)	−.40	.16	−.40	.16
(14) It is always morally wrong to bring a child into this world.	.77	.60	.77	.60
(15) People should adopt rather than have children themselves.	.54	.29	.54	.29
(16) Bringing a child into existence is always a considerable net harm to that child.	.77	.59	.77	.59
(17) Conceiving and raising children is immoral because it requires money, time, and energy that could be spent on giving to charitable organizations that save lives and reduce suffering.	.81	.65	.81	.65
(18) The absence of pain is good, even if that good is not enjoyed by anyone. This is why not being born is good.	.65	.42	.65	.42
(19) Some people's choices to have children result in other people's suffering by depriving them of basic rights like the right to food.	.60	.35	.60	.36
(20) No matter your economic circumstances, having a child is never bad thing. (R)	−.41	.17	−.40	.16
(21) It is better to exist than not to exist because only in existence can there be pleasure. If one does not exist, one cannot have pleasure. As such, it is morally good to have a child. (R)	−.43	.19	−.43	.18
(22) It would be good if we all stopped having children.	.78	.60	.78	.61

Note. Factor loadings (FL) and communality scores (h2) for the full 22-item scale and the 20-item reduced scale.

adapted the item set. The final 22 items presented here capture antinatalism sufficiently in depth and breadth such that a large variety of different philosophical motivations for antinatalism were included (such as consent-based claims, suffering-based claims, global, as well as local versions). All items are formulated in clear language that is accessible to nonexperts.

Our initial item pool included a total of 13 items relating to a priori antinatalism (claims about the universal and general wrongness of procreation), five of which were

reverse scored, as well as nine items relating to a posteriori antinatalism (claims about the circumstantial wrongness of procreation for reasons such as human rights and the environment), three of which were reverse scored. Furthermore, the individual items were sometimes phrased as more complex arguments, other times as moral pronouncements or statements of fact. The topics covered by the items were adoption, the relationship between procreation and global poverty, the impact of procreation on the environment, one's relationship to one's child, and the general moral status of procreation.

Methods

We recruited 303 participants (50.8% female) via Prolific who were between 18 and 82 years old ($M = 44.76$, $SD = 16.08$). Seven participants failed the attention check that asked people to indicate *strongly agree* on a 7-point Likert scale leaving the final sample at 296. This attention check was used for all three studies on Prolific. All participants had a minimum approval rate of 95% on Prolific and had participated in at least 10 studies prior. All participants were paid £0.50 for the completion of the study. They were presented with the 22 initial items. In this and all further studies, participants were asked to agree or disagree with the specific items on a 5-point Likert scale ranging from *strongly disagree* (1) to *strongly agree* (5). All items were presented in a random order to participants. At the end of the survey, participants were asked to complete a short demographic questionnaire.

We sought evidence of structure-based validity via an exploratory factor analysis based on polychoric correlations using the R (R Core Team, 2021) package psych (Revelle, 2022). We investigated the adequacy of the data and sample through the Kaiser–Meyer–Olkin criterion as well as Bartlett's sphericity tests. We used maximum likelihood (ML) factor extraction method with Oblimin rotation. Furthermore, we drew on parallel analysis based on minimum rank factor analysis with 500 simulations for factor retention (Timmerman & Lorenzo-Seva, 2011) via the EFA.MRFA R package (Navarro-Gonzalez & Lorenzo-Seva, 2021) and FACTOR software (version factor.12.03.02; Ferrando Piera & Lorenzo-Seva, 2017) to calculate the eigenvalues and explained variance based on eigenvalues. Lastly, we used the Tucker–Lewis index (TLI) as well as root-mean-square error of approximation (RMSEA) to evaluate model fit.

Results

Our results showed robust sample and data adequacy, with Kaiser–Meyer–Olkin = .92 as well as Barlett's sphericity $\chi^2(231, N = 296) = 2,957.2$, $p < .001$. Parallel analysis suggested extracting a single factor, with the explained variance for this first observed factor being 45.1% (eigenvalue = 9.93). The second factor explained 8.84% variance (eigenvalue = 1.94), while the third factor explained 5.95% of the variance (eigenvalue = 1.31). The upper bound of the 95% CI of the corresponding explained variance of the first simulated factor was 9.86%, while the second and third were 9.86% and 9.23%, respectively. This led to the following model fit indices, overall indicating poor model fit: TLI = .762, RMSEA = .118 (CI 90% .112–.126). When extracting the factors with all items, the factor loadings ranged from .40 to .88. We excluded two items that presented factor loadings of < .40. The fit indices remained below the cut-off points for good model fit. The results were, TLI = .79, RMSEA = .119 (CI 90% .111–.127), though reliability indices were rather robust: α = .91, ω = .95, and greatest lower bound of reliability, GLB = .85. Below see Table 1 with the factor loadings for all items (Model 1) and with those items that had a factor loading < .40 excluded (Model 2). To reduce the scale further and to generate hypotheses for Study 2, we conducted a non-preregistered confirmatory factor analysis (CFA). In this analysis, we found better fit statistics for the model with items whose factor loadings ≥ .60 (Model 3): $\chi^2(54, N = 296) = 163.838, p < .001$, CFI = 0.943, TLI = 0.930, RMSEA = 0.083 (CI 90% 0.069–0.098), SRMR = 0.039.

Discussion

The data suggested that the extraction of a single factor was most appropriate. While the results indicated good sample and data adequacy, the overall model fit was poor. To reduce the scale and increase model fit, we removed all items with a factor loading of .40 or smaller. However, this also did not improve model fit meaningfully. To arrive at a shortened version that researchers can more easily implement and to improve model fit going forward, we shortened the scale and included only items with a factor loading of .60 or greater in an attempt to eliminate items that loaded worst on the single factor. In non-preregistered CFA, we found improved fit which motivated us to continue working on this scale in this manner. This new 12-item scale was then used for CFA in Study 2 to test our item choice in a new data set and to provide model fit estimations in a new data set.

Study 2

The goal of this study was to conduct a CFA based on the results from Study 1. We used the reduced 12-item scale extracted in Study 1.

Methods

We recruited 399 participants (48.8% female) via Prolific who were between 18 and 84 years old (M = 34.34, SD = 11.98). Three participants failed the attention check, leaving the final sample at 396 participants. All participants had a minimum approval rate of 95% on Prolific and had participated in at least 10 studies before. They were paid £0.50 for the completion of the study. We also excluded those who had participated in Study 1. All 12 items were presented in a random order. At the end of the survey, participants were presented with a short demographic questionnaire.

In this study, we ran a CFA for the single-factor model identified in Study 1 (with factor loadings ≥ .60). This analysis was conducted in R's lavaan package (Rosseel, 2012). We assessed model fit using ML estimator via a combination of the following statistics: chi-squared (χ^2), Tucker–Lewis index (TLI), comparative fit index (CFI), root-mean-square error of approximation (RMSEA), and standardized root-mean-square residuals (SRMR). We identified the cut-off points for most of our indices via McNeish and Wolf's (2021) dynamic fit index cut-off. This applies to SRMR, RMSEA, and CFI (using ML estimation). This dynamic cut-off has three distinct levels: Level 1, Level 2, and Level 3 (with the first level being defined as having 1/3 of items with a residual correlation of .3 that one has failed to include in one's model, respectively, with second and third level being defined as 2/3 and 3/3 of items, respectively). We further set the cut-off for TLI at .90 in line with Brown's (2015) recommendation. We also ran a CFA using WLSMV estimator and considering items to be ordered. When using WLSMV, we draw on Brown's (2015) recommendation: TLI and CFI > 0.90, RMSEA < 0.08 with CI 90% not surpassing 0.10, and SRMR < 0.08.

Additionally, we conducted a Mokken scale analysis (MSA) to select the best items from this instrument (Mokken, 2011, see Franco et al., 2022 for a review). MSA is a nonparametric item response theory (IRT) model, which employs less restrictive assumptions about data that are often done by parametric statistics (in the IRT case, the relation between the probability of obtaining a particular score on a particular item of a latent variable depends on an item response function that often follows a logistic function or a normal ogive model, while in MSA only imposes order restrictions of item response functions, so they can have any shape). The montone homogeneity model (tested here) tests the three central item response theory assumptions: unidimensionality (each item is only measuring one latent trait), local independence (the item responses only depend on the latent trait being measured), and latent monotonicity (the item step response functions are nondecreasing functions of the latent trait). We implemented a genetic

algorithm to conduct the Mokken scale analysis because it has been shown to perform well in recovering the correct dimensionality of scales (Straat et al., 2013). Following Straat and colleagues' recommendations, we use a scalability coefficient of .30 as our cut-off point. We further conduct a manifest monotonicity test (Junker & Sijtsma, 2000), which involves a regression between the scores of individual items and the residual scores (of omitting selected items from the total test score). To adjust for the fact that the number of respondents at specific score levels can be small, we group respondents with adjacent residual scores until a minimum proportion of individuals per score is greater than $n/5$ (Sijtsma & Molenaar, 2002). We used this approach to test the assumption of nonintersection of item response functions using rest scores. Furthermore, we also assessed the reliability of the scale through the Molenaar–Sijtsma reliability statistic (MS), which is an unbiased estimator of test score reliability (Molenaar & Sijtsma, 1984; Molenaar & Sijtsma, 1988) as well as Cronbach's α (Cronbach, 1951) and Guttman's λ^2 (Guttman, 1945). We conducted the MSA and the reliability analyses in R (R Core Team, 2021) with the Mokken package (Van der Ark, 2012).

Results

Our CFA of the shortened 12-item scale again showed poor model fit using ML estimator across a variety of tests: χ^2(54, N = 396) = 301.624, p < .001, CFI = 0.921, TLI = 0.904, RMSEA = 0.108 (CI 90% 0.096–0.120), SRMR = 0.043. The dynamic fit index cut-offs for Level 1 were as follows: CFI = 0.986, RMSEA = 0.044, SRMR = 0.028; Level 2: CFI = 0.972, RMSEA = 0.064, SRMR = 0.031; Level 3: CFI = 0.951, RMSEA = 0.086, SRMR = 0.035. As for the CFA using WLSMV estimator, we found poor fit for the RMSEA indicator but good fit for the remaining statistics: χ^2(60, N = 396) = 341.241, p < .001, CFI = 0.973, TLI = 0.966, RMSEA = 0.116 (CI 90% 0.104–0.128), SRMR = 0.042. For factor loadings of this 12-item scale, see Table 2.

Because the first model showed poor model fit, we ran the same analysis on a second model where some items regarding morality had their residuals correlated (Item 6 correlated with Items 1 and 9). This was not preregistered and was as such fully exploratory. The model fit was markedly better but still remained inadequate: χ^2(52, N = 396) = 189.432, p < .001, CFI = 0.956, TLI = 0.945, RMSEA = 0.082 (CI 90% 0.069–0.094), SRMR = 0.039. The dynamic fit index cut-offs for Level 1 were as follows: CFI = NONE, RMSEA = NONE, SRMR = 0.024; Level 2: CFI = 0.987, RMSEA = 0.044, SRMR = 0.028; Level 3: CFI = 0.972, RMSEA = 0.066, SRMR = 0.031. Dynamic cut-off point results of NONE

Table 2. Factor loadings for 12-item scale

Items	FL ML	FL WLSMV
(1) Because humans cause so much harm to the environment, it is morally wrong to have children.	.807	.876
(2) It is better not to exist than to exist because only in existence can there be pain. If one does not exist, one cannot be in pain. As such, it is morally wrong to have children.	.816	.880
(3) Bringing someone into existence means that they will eventually suffer in one way or another, which is why it is morally wrong to have children.	.829	.878
(4) It is morally wrong to bring a child into this world because a child cannot consent to being born.	.813	.865
(5) Bringing a new child into this world is a morally good thing. (R)	−.475	−.537
(6) Having a child in the developed world is morally wrong as it impacts the environment at least to the same extent as excessive consumption behavior.	.782	.869
(7) It is always morally wrong to bring a child into this world.	.804	.868
(8) Bringing a child into existence is always a considerable net harm to that child.	.767	.817
(9) Conceiving and raising children is immoral because it requires money, time, and energy that could be spent on giving to charitable organizations that save lives and reduce suffering.	.764	.849
(10) The absence of pain is good, even if that good is not enjoyed by anyone. This is why not being born is good.	.682	.748
(11) Some people's choices to have children result in other people's suffering by depriving them of basic rights like the right to food.	.601	.677
(12) It would be good if we all stopped having children.	.738	.804

Note. Factor loadings (FL) for the 12-item scale using maximum likelihood (ML) and weighted least squares mean and variance adjusted (WLSMV).

indicate that this fit index is unable to differentiate between well-fitting and ill-fitting models for our specific model at this concrete level of misspecification. See OSF Appendix A, https://osf.io/rs23g/ (Maier et al., 2022), for a figure of the comparison of the fit index distributions for the *true* empirical model and the *misspecified* empirical model for the first model and the second model. As for the CFA of this model using WLSMV estimator, we also found poor fit for the RMSEA indicator but good fit for the remaining statistics: $\chi^2(62, N = 396) = 227.378$, $p < .001$, CFI = 0.983, TLI = 0.979, RMSEA = 0.092 (CI 90% 0.080–0.105), SRMR = 0.036.

Lastly, our Mokken scale analysis results indicated that Items 1, 5, 9, 10, 11, and 12 should be excluded from further analysis since they did not form a scale (or formed a second scale). Importantly, the excluded items included both items referring to the general claim that procreation is always wrong – or the reversed item that it is good – (Items 5 and 12) as well as more local formulations relating to environmental and human rights-based accounts (Items 1 and 11). As such, this makes it quite unlikely that these items themselves might form a distinct factor. Going forward, we decided to maintain Items 2, 3, 4, 7, and 8. In those items, we did not observe any monotonicity violations and had overall good internal consistency measures (MS = 0.916, Cronbach's α = 0.911, Guttman's λ^2 = 0.912). In further exploratory analysis, we ran a third model based on the 5-item scale drawn from the Mokken scale analysis. We observed better model fit with the ML estimator: $\chi^2(5, N = 396) = 11.457$, $p = .043$, CFI = 0.995, TLI = 0.990, RMSEA = 0.057 (CI 90% 0.009–0.101), SRMR = 0.014.

The dynamic fit index cut-offs for Level 1 were as follows: CFI = 0.993, RMSEA = 0.068, SRMR = 0.015, Level 2: CFI = 0.98, RMSEA = 0.121, SRMR = 0.023. We also found good fit with WLSMV estimator: $\chi^2(25, N = 396) = 9.187$, $p = .102$, CFI = 0.999, TLI = 0.998, RMSEA = 0.046 (CI 90% 0.000–0.092), SRMR = 0.011. To test whether this better model fit simply arose because we selected for it post data collection, we ran a further preregistered study (Study 3) to test the 5-item scale specifically.

Discussion

The results of our CFA of the 12-item scale showed unsatisfactory model fit. While we could improve model fit with non-preregistered analyses by correlating some residuals, we drew on Mokken scale analysis to further reduce the scale. Based on this procedure, we removed a further seven items and constructed a short 5-item version of the antinatalism scale. To test the 5-item scale on a new sample, we preregistered and ran a second CFA (Study 3).

Study 3

The goals of this study were twofold. First, we aimed to evaluate the model fit of the 5-item version of the antinatalism scale on a new sample. Second, we aimed to test

for validity by investigating the relationship with other scales measuring related constructs: First, we measured the relationship between antinatalism and life satisfaction. We hypothesized that low life satisfaction would correspond to higher antinatalist views as those who suffer more in their own life might have a more negative outlook on the question of whether it is better to exist than not to exist (and thus procreation), as has generally been established in the overgeneralisation literature, e.g., in the case of depression (Moore & Fresco, 2012). Second, we investigated the relationship between antinatalism and conservatism because standardly conservative values, for example, relating to the upholding of family values, are in many ways antithetical to antinatalism. Third, we looked at the association between antinatalism and empathy. In this case, there were two plausible predictions. First, those high in empathy may be more sensitive to the risks of causing suffering by bringing life into this world, thus standing in a positive relationship to antinatalism. Second, this relationship may also be negative, as it has been previously found that that those low on empathy (i.e., high in dark triad traits) are more likely to hold antinatalist views (Schönegger, 2021). Therefore, we aimed to investigate which of these two potential relationships of antinatalism and empathy was empirically better supported and did not make a directional prediction. We did not include these measures in any earlier study as we wanted to establish a reliable scale before proceeding with this step and also not in Study 4 as correlations based on a restricted sample would be affected by collider bias (e.g., De Ron et al., 2021).

Methods

We recruited 300 participants (46.5% female) via Prolific who were between 18 and 74 years old ($M = 34.49$, $SD = 12.28$). Three participants failed the attention check, leaving the final sample at 297. All participants had a minimum approval rate of 95% on Prolific and had participated in at least 10 studies before. They were paid £0.50 for the completion of the study. We also excluded those who had participated in Studies 1 and 2. All five items were presented in a random order to participants. After the main scale, participants were represented with three additional scales in random order. First, they were represented with a scale measuring cognitive and affective empathy (Carré et al., 2013). This 20-item scale uses a 5-point Likert scale across all items (ranging from 1 = *strongly disagree* to 5 = *strongly agree*). The two-factor model used here consists of two components that assess empathy in young people and adults. These two components are affective empathy and cognitive empathy. (For reliability of these two factors, see Carré et al., 2013, p. 691, Appendix

B). Second, they were presented with a life satisfaction scale (Margolis et al., 2019). This scale consists of six items measuring life satisfaction on a 7-point Likert scale (ranging from 1 = *strongly disagree* to 7 = *strongly agree*) that also aim to account for acquiescence bias. It has high reliability at $\omega = .93$. Third, they were presented with a scale to measure conservatism (Everett, 2013). This scale aims to measure modern conceptualizations of conservatism on a 14-item scale and includes two subscales: social conservatism and economic conservatism. Participants rated the items on a scale from 0 to 100, where 0 was coded as feeling very negative toward an issue and 100 as feeling very positive. The scale has shown overall strong reliability at $\alpha = .88$ (with the social conservatism subscale showing $\alpha = .87$ and the economic conservatism subscale showing $\alpha = .70$). These were added to seek evidence of validity.

First, we conducted taxometric analyses using R's RTaxometrics package (Ruscio & Wang, 2021; for further reading, see Ruscio et al., 2013) to test whether our instrument is better described as having latent profiles or dimensions. Taxometric analysis has a core premise that not all individual differences are alike and tries to answer the question of whether the latent construct being measured falls along a continuous spectrum or if constructs form two (or more) separate groups of people. We find evidence for this via the comparison curve fit index (CCFI), where values below .40 indicate dimensions, values above .60 indicate latent profiles/classes, and values between .40 and .60 indicate ambiguity between the two (Ruscio et al., 2018). If we found evidence for latent profiles, we would conduct latent profile analysis, and if we found evidence for dimensions or if the evidence was ambiguous, we could conduct a CFA for the single-factor model. As before, the CFA was conducted in R (R Core Team, 2021), using the lavaan package (Rosseel, 2012) with the same goodness of fit statistics (χ^2, TLI, CFI, RMSEA, and SRMR), and using both ML and WLSMV estimators. As in previous studies, we calculate the cut-off points via McNeish and Wolf's (2021) dynamic cut-off indices for the ML estimator and uses Brown's (2015) recommendation for the WLSMV estimator. Second, we ran a structural equation model with a diagonally weighted least squares (DWLS) estimator using the lavaan package (Rosseel, 2012) in R. As before, we used the following statistics to determine the goodness of fit of the model: χ^2, TLI, CFI, RMSEA, and SRMR.

Results

Our taxometric analysis produced results that were ambiguous between latent profiles and dimensions, with

Table 3. Factor loadings for the Short Antinatalism Scale (S-ANS)

Items	FL ML	FL WLSMV
(1) It is better not to exist than to exist because only in existence can there be pain. If one does not exist, one cannot be in pain. As such, it is morally wrong to have children.	.753	.885
(2) Bringing someone into existence means that they will eventually suffer in one way or another, which is why it is morally wrong to have children.	.750	.893
(3) It is morally wrong to bring a child into this world because a child cannot consent to being born.	.760	.898
(4) It is always morally wrong to bring a child into this world.	.726	.884
(5) Bringing a child into existence is always a considerable net harm to that child.	.535	.714

Note. Factor loadings (FL) for the 5-item scale using maximum likelihood (ML) and weighted least squares mean and variance adjusted (WLSMV).

the comparison curve fit index (CCFI) results being, $M = 0.425$, MAMBAC = 0.372, MAXEIG = 0.463, LMode = 0.460. In line with our preregistration for values between .40 and .60, we treated the 5-item antinatalism scale as a dimensional model, thus conducting a CFA.

Our CFA results showed excellent model fit across all tests for the ML estimator: $\chi^2(5, N = 297) = 5.227$, $p = .389$, CFI = 1.00, TLI = 0.999, RMSEA = 0.012 (CI 90% 0.000–0.082), SRMR = 0.017. The dynamic fit index cut-offs for Level 1 were as follows: NONE; Level 2: CFI = 0.986, RMSEA = 0.073, SRMR = 0.027; Level 3: CFI = 0.97, RMSEA = 0.113, SRMR = 0.037. We also found good fit with the WLSMV estimator: $\chi^2(25, N = 297) = 7.710$, $p = .173$, CFI = 0.999, TLI = 0.998, RMSEA = 0.043 (CI 90% 0.000–0.099), SRMR = 0.015. See Table 3 for the standardized factor loadings of this 5-item scale, the Short Antinatalism Scale (S-ANS).

Furthermore, we ran a structural equation model using the DWLS estimator, correlating the antinatalism factor with empathy (both affective and cognitive), life satisfaction, and conservatism (both social and economic). We observe negative statistically significant correlations between antinatalism and affective empathy ($r = -.082$, $p < .001$), life satisfaction ($r = -.198$, $p < .001$), social conservatism ($r = -.425$, $p < .001$), and economic conservatism ($r = -.360$, $p < .001$). We did not observe a statistically significant relationship between antinatalist views and cognitive empathy ($r = .025$, $p = .159$).

Discussion

Our 5-item scale showed excellent model fit across all tests. This indicates that the items from the short scale were accurately captured by our single-factor model and was a suitable candidate to seek additional evidence of validity. We ran a structural equation model correlating antinatalism with a number of constructs that plausibly stand in a relationship to antinatalism: First, those lower on life satisfaction (Margolis et al., 2019) might be higher on

antinatalism due to (over)generalization from own experiences. Furthermore, previous research has found a relationship between depression and antinatalist views (Schönegger, 2021). Furthermore, those higher on conservatism (Everett, 2013) might be lower on antinatalism as standardly conservative values like the traditional family with several children are, at least in part, inconsistent with antinatalism. Again, we observed this relationship in the predicted direction. Lastly, we included a measure of both cognitive as well as affective empathy (Carré et al., 2013). We were agnostic as to which direction the effect would run with either type of empathy. On the one hand, previous research found a positive relationship of antinatalism to dark triad traits (and as such low empathy; Schönegger, 2021), though there is also a plausible explanation of expecting antinatalists to be exceptionally high in empathy which plays out in their concern for the unborn generation. Because we failed to find a significant relationship to cognitive empathy but did find a statistically significant negative relationship to affective empathy, we take these data to speak in favor of the former possibility. However, note that while this relationship was statistically significant, its effect size was rather small at −.08, which may make it not as practically relevant as our other relationships.

These results provide support for validity in showing correlations with theoretically related constructs. Because the relationships were all in the same direction as predicted by theory (where such directional predictions based on theory existed), we take this to provide evidence for the validity of this scale and as such, its potential applicability in other contexts. However, to provide further data on validity, we also ran a further study (Study 4) with the S-ANS on a sample of self-described antinatalists to provide additional evidence of the validity of this scale.

Furthermore, the items that make up this final 5-item scale capture the central antinatalism construct by focusing on the primary a priori formulations of antinatalism. Recall that all five items directly draw on the notion of moral wrongness of bringing a child into existence in most (or all) circumstances, while varying the reason (ranging from

pain-based arguments to consent-based claims). This captures the essence of the antinatalist view.

Study 4

The goals of this study were twofold. First, we aimed to test whether the measure was invariant between self-described antinatalists and a general nonantinatalists population (specifically, participants from Study 3). This investigation is crucial as it might be that the additional philosophical knowledge of antinatalists causes them to interpret the items in a different way, making a comparison to the general population difficult. After establishing that our measure is invariant, we aimed to provide additional evidence regarding validity by showing that self-described antinatalists would score higher on our scale than the general population.

Methods

We initially recruited 351 participants via the Antinatalism subreddit on https://www.reddit.com. However, data collection on Reddit faces issues of repeated submissions and potential bot submissions. Therefore, we used Google's invisible Re-Captcha bot detection software integrated in Qualtrics that estimates the probability of the participant being a human (with 1 being certainly a human and 0 certainly a bot). We excluded all participants with a bot score lower than .8 (48 participants). Furthermore, we removed participants that entered the study prior to activation of software that prevented repeated submissions by blocking participants from entering the study again (123 participants). The activation of this software was made in response to a sudden influx of low-quality data within the first hour of data collection that we had not anticipated. To ensure high data quality, we removed all entries submitted prior to the activation of this software, though all participants remained eligible for payment. This step was not preregistered, but we felt that deviating from the preregistered protocol was needed given this situation. In addition, we excluded participants that took less than or equal to 60 s to complete the survey (24 participants), that answered the attention check incorrectly (68 participants), or that did not self-identify as antinatalist (132 participants). Note that several participants may have been excluded based on more than one criterion, and as such, the sum of all these exclusions may not equal the total exclusions.

In total, we excluded 252 participants. This left us with 99 participants (40.4% female) between 18 and 59 years old ($M = 26.81$, $SD = 7.77$). Participants could enter a lottery for a £50 gift card if they wanted to. Participants were presented with the final five items before being shown a short demographic questionnaire and a question asking them if they self-identified as an antinatalist.

We tested the differential item functioning (DIF) to see if the instrument is invariant between antinatalists and nonantinatalists (Jaloto, 2021). To analyze DIF, we are following five steps: (1) item calibration and parameter estimation using the graded response model (2) scores estimation using expected a posteriori (EAP) method, (3) identification of items with DIF, (4) in case we find items with DIF, we will return to Step 1 considering that the item is not invariant and proceed with the following steps, and (5) model fit verification using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), expecting that the AIC and BIC for the last model is lower than the first model. All DIF analyses were made using R (R Core Team, 2021) packages mirt (Chalmers, 2012) and lordif (Choi et al., 2011). For a full specification of all three models, see OSF Appendix B, https://osf.io/rs23g/ (Maier et al., 2022; Jodoin & Gierl, 2001; Zumbo, 1999).

After detecting if items show DIF or not, we estimated subjects' EAP scores – the expected a posteriori scores, i.e., the posterior probability distribution of the latent scores of this trait – considering DIF's results and use them to test mean differences between antinatalists and nonantinatalists (from Study 3) with a Welch t test in R (R Core Team, 2021). The effect size was calculated by Cohen's d and Hedges's g using the effect size (Ben-Shachar et al., 2020) R package.

Results

First, we proceed with the DIF analysis with the five items of the S-ANS. We conclude that the Short Antinatalism Scale has five items that are invariant between self-described antinatalists and non-self-described antinatalists. For a full set of results of this measurement invariance analysis, see OSF Appendix C, https://osf.io/rs23g/ (Maier et al., 2022).

Second, we proceeded with the Welch t test with EAP scores to analyze mean differences between self-described antinatalists and non-antinatalists. In Figure 1, we depict a raincloud plot visualizing the distribution of EAP scores for antinatalists and the general population.

Overall, we find that self-described antinatalists ($M = 0.011$, $SD = 0.896$) scored significantly higher on the antinatalism scale than the lay population ($M = -2.708$, $SD = 0.994$), $t(184.76) = 25.435$, $p < .001$, Cohen's $d = 2.80$ [95% CI (2.50–3.10)], Hedges's $g = 2.80$ [95% CI (2.50–3.09)]. The plot was created with JASP (JASP Team, 2022).
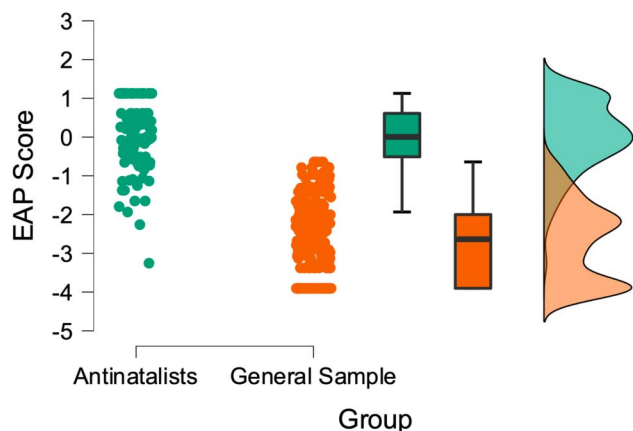
Figure 1. Distributions of EAP scores for antinatalists and the general population.

Discussion

These results provide additional evidence of the validity of our scale, as we find that the scale is measurement invariant between antinatalists and the general population. In addition, we find, as expected, that antinatalists score higher with a large effect size ($d = 2.80$). This further provides evidence of validity as we have shown that self-described antinatalists score considerably higher.

General Discussion

In this paper, we have introduced the Short Antinatalism Scale (S-ANS) and provided evidence in favor of its reliability and validity on Prolific samples and an antinatalist sample. While longer versions of this scale (22-item and 12-item) did not show good model fit, our final 5-item scale showed exceptional model fit, while also behaving as predicted in validation studies in both a Prolific and Reddit specialist antinatalist sample. As such, we take our scale to be a strong first contender for measuring antinatalist views rigorously, and we hope that researchers will pick up this freely available scale to employ it in their own research across a variety of fields where antinatalism is relevant as well as to fields whose area of research loosely relates to the value of life and procreation.

In favor of the adoption of this scale, we argue that controlling for antinatalist views has many applications across a wide variety of fields. For one, all research investigating population ethics specifically ought to control for antinatalist tendencies. This is primarily because aversion to certain increases in populations may be misattributed to factors like preference for increasing the average welfare when they might, at least in part, be

determined by antinatalist sentiments. Additionally, our research may also promise to provide a deeper understanding of the prochoice versus prolife debate as well as into the psychological mechanisms that drive these views (e.g., Rye & Underhill, 2020). Furthermore, we argue that this may also apply to a number of research areas where participant preferences about hypothetical scenarios may include population changes. These may include topics like increasing the fertility rates as a public policy (Kaplan & Lancaster, 2017) or interstellar expansion to avoid the negative effects of climate change. We argue that being able to control for antinatalist views in these areas is important even if there are very few self-described antinatalists in the general population. This is because our scale captures variation in agreement with antinatalism that captures the full spectrum of possible views, making it potentially useful to control for even when the study population does not hold conscious views regarding it. Furthermore, while we might have theoretical reasons to believe that antinatalist tendencies might underlie certain relationships, this has not been tested before. By controlling for antinatalism through the S-ANS, this is now possible. As such, we argue that in all these cases, properly controlling for antinatalist views may prove useful and doing so is now easily implementable via the Short Antinatalism Scale (S-ANS). For a clean version of the S-ANS to be straightforwardly adopted by other researchers, see OSF Appendix D, https://osf.io/rs23g/ (Maier et al., 2022).

This scale also has direct implications for the general psychological literature. First, it may help us directly address attitudes toward children that have been studied from a variety of angles like voluntary childlessness (Peterson & Engwall, 2019) or evolutionary psychology explanations of procreative behavior (Apostolou & Hadjimarkou, 2018; Brown & Keefer, 2020) from a new point of view. Second, our findings of the relationship of antinatalism to empathy, life satisfaction, and conservatism allow for a direct exchange between the results reported in this paper and these literatures in psychology more broadly, with potential intersections for further research, perhaps on the directional relationship between life satisfaction and antinatalist views. Third, our scale also enables proper research into the social perceptions of antinatalist tendencies and antinatalism generally. This has not been researched before, but with the S-ANS, psychological research into the social perceptions of those holding antinatalist views is now straightforwardly available and scientifically interesting.

Importantly, while the initial 22 items included a wide variety of different formulations of antinatalism, ranging from general formulations that procreation is always morally impermissible to more restricted and local claims of procreation's impermissibility in cases where doing so

damages the environment or where parents are unable to properly care for their children, the final S-ANS only includes items of the former kind. In other words, the final scale consists only of items that capture the more general (or a priori) formulations of antinatalism that state a categorical wrongness of procreation. One upside of this is that our scale narrows in on general or a priori antinatalism as opposed to including both general and local formulations that may be confounded by changing empirical realities and beliefs. This allows for the scale to be applied rather well in contexts where general antinatalism is of interest, like in population ethics.

There are some limitations to this scale. First, the Short Antinatalism Scale (S-ANS) is only measuring the construct of antinatalism and does not explicitly measure pronatalist views. After all, it is not evident that anyone who would score low on antinatalism would necessarily score high on pronatalism (though a certain negative relation should exist conceptually), as one may be averse to antinatalist reasoning as expressed in the philosophical literature and by self-ascribed antinatalists, but hold additional reasons to oppose procreation, perhaps above a certain threshold. As such, it is important to only apply this scale to measure antinatalist views and not use it to inversely measure pronatalist attitudes.

Second, this is a short-form scale; thus, it does not capture all arguments, behaviors, or attitudes related to antinatalist views, specifically local and conditional formulations of antinatalism that may heavily depend on cultural contexts. The focus of this scale is also apparent in the high reliability, which indicates that all items measure a similar construct. However, face validity shows that the items do incorporate different arguments in favor of antinatalism, and the scales show good evidence of validity in a variety of contexts. As such, we remain confident in our scale but encourage more research in creating other items that might capture antinatalist views in a different and perhaps broader manner.

Third, one may worry that our scale is too narrow as four of the five final items contain the concept of it being morally wrong to have children. On this worry, one may think that this would explain why these items showed the strongest results from a psychometric point of view, thus being primarily responsible for the one-factor structure observed here. While we agree that a multifactor structure might have been quite interesting, we argue that this worry can be rephrased as a genuine strength of the result. Rather than say that this reductionist result poses a challenge, we want to point out that the items that were selected for the final S-ANS are the clearest definitions and statements of a priori antinatalism. In other words, we argue that our scale captures exactly what it was intended to capture by narrowly measuring the universal attitude

that procreation is morally wrong. Furthermore, recall that one item of the final S-ANS does not use this formulation; instead, it refers to the fact that bringing a child into existence is always a net harm to that child, thus weakening this objection in the first place. Additionally, while items that specifically state that procreation is morally wrong are overrepresented in the S-ANS compared to the first item set, it is important to point out that three items that also specifically mention morally wrong did not fit our inclusion criteria for the final scale. These items (Items 1, 8, and 11 in the original list) are all a posteriori/conditional statements of antinatalism that nonetheless use the formulation of morally wrong. Specifically, they are concerned with harm to the environment, inability to properly raise children, and consumption behavior. As such, one can be quite confident that the final S-ANS did not only capture the moral wrongness element but rather properly identified universal antinatalist sentiments. As such, we conclude that the S-ANS narrowly and adequately captures the core antinatalist belief.

To conclude, this paper developed the Short Antinatalism Scale (S-ANS), which has shown excellent reliability and validity across a variety of studies. We hope that this freely available scale will advance measurement in research related to the value of life and procreation across several disciplines.

References

Apostolou, M., & Hadjimarkou, M. (2018). Domains of motivation in men and women for initiating and terminating procreation in an evolutionary perspective. *Marriage & Family Review*, 54(5), 486–506. https://doi.org/10.1080/01494929.2017.1414723

Arrhenius, G. (2000). An impossibility theorem for welfarist axiologies. *Economics and Philosophy*, 16(2), 247–266. https://doi.org/10.1017/s0266267100000249

Ben-Shachar, M. S., Lüdecke, D., & Makowski, D. (2020). Effectsize: Estimation of effect size indices and standardized parameters. *Joss*, 5(56), Article 2815. https://doi.org/10.21105/joss.02815

Benatar, D. (1997). Why it is better never to come into existence. *American Philosophical Quarterly*, 34(3), 345–355.

Benatar, D. (2008). *Better never to have been: The harm of coming into existence*. Oxford University Press.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford.

Brown, F. L. (2020). *The effects of optimism on antinatalism* [Master's thesis, University of Southern Mississippi]. https://aquila.usm.edu/masters_theses/746/

Brown, F. L., & Keefer, L. A. (2020). Anti-natalism from an evolutionary psychological perspective. *Evolutionary Psychological Science*, 6(3), 283–291. https://doi.org/10.1007/s40806-019-00226-9

Carré, A., Stefaniak, N., D'Ambrosio, F., Bensalah, L., & Besche-Richard, C. (2013). The Basic Empathy Scale in Adults (BES-A): Factor structure of a revised form. *Psychological Assessment*, 25(3), 679–691. https://doi.org/10.1037/a0032297

Caviola, L., Althaus, D., Mogensen, A. L., & Goodwin, G. P. (2022). Population ethical intuitions. *Cognition*, *218*, Article 104941. https://doi.org/10.1016/j.cognition.2021.104941

Chalmers, R. P. (2012). mirt: A Multidimensional Item Response Theory Package for the R environment. *Journal of Statistical Software*, *48*(1), 1–29. https://doi.org/10.18637/jss.v048.i06

Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software*, *39*(8), 1–30. https://doi.org/10.18637/jss.v039.i08

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. https://doi.org/10.1007/BF02310555

De Ron, J., Fried, E. I. & Epskamp, S. (2021). Psychological networks in clinical populations: Investigating the consequences of Berkson's bias. *Psychological Medicine*, *51*(1), 168–176. https://doi.org/10.1017/S0033291719003209

Everett, J. A. (2013). The 12 Item Social and Economic Conservatism Scale (SECS). *PLoS ONE*, *8*(12), Article e82131. https://doi.org/10.1371/journal.pone.0082131

Ferrando Piera, P. J., & Lorenzo Seva, U. (2017). Program FACTOR at 10: Origins, development and future directions. *Psicothema*, *29*(2), 236–240. https://doi.org/10.7334/psicothema2016.304

Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, *3*(4), 456–465. https://doi.org/10.1177/2515245920952393

Franco, V. R., Laros, J. A., & Bastos, R. V. S. (2022). Theoretical and practical foundations of Mokken scale analysis in psychology. *Paidéia (Ribeirão Preto)*, *32*, Article e3223. https://doi.org/10.1590/1982-4327e3223

Greaves, H. (2017). Population axiology. *Philosophy Compass*, *12*(11), Article e12442. https://doi.org/10.1111/phc3.12442

Guttman, L. (1945). A basis for analyzing test–retest reliability. *Psychometrika*, *10*(4), 255–282.

Jaloto, A. (2021). Funcionamento diferencial do item (DIF) e in-variância da medida [Differential item functioning (DIF) and measurement invariance]. In C. Faiad, M. N. Baptista, & R. Primi (Orgs.), *Tutoriais em Análise de Dados Aplicados à Psicometria* (pp. 268–291). Editora Vozes.

JASP Team. (2022). *JASP* (Version 0.17) [Computer software].

Jodoin, M. G., & Gierl, M. J. (2001). Evaluating type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection. *Applied Measurement in Education*, *14*(4), 329–349. https://doi.org/10.1207/S15324818AME1404_2

Junker, B. W., & Sijtsma, K. (2000). Latent and manifest monotonicity in item response models. *Applied Psychological Measurement*, *24*(1), 65–81. https://doi.org/10.1177/01466216000241004

Kaplan, H. S., & Lancaster, J. B. (2017). The evolutionary economics and psychology of the demographic transition to low fertility. In L. Cronk, N. Chagnon, & W. Irons (Eds.), *Adaptation and human behavior* (pp. 283–322). Routledge. https://doi.org/10.4324/9781351329200-18

Lilienfeld, S. O., & Strother, A. N. (2020). Psychological measure-ment and the replication crisis: Four sacred cows. *Canadian Psychology/Psychologie canadienne*, *61*(4), 281–288. https://doi.org/10.1037/cap0000236

MacIver, C. (2015). Procreation or appropriation? In S. Hannan, S. Brennan, & R. Vernon (Eds.), *Permissible progeny: The morality of procreation and parenting* (pp. 107–128). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199378111.003.0005

Maier, M., Schönegger, P., & Bastos, R. V. S. (2022). *Anti-natalism scale* [Data set]. https://osf.io/a2bpd

Margolis, S., Schwitzgebel, E., Ozer, D. J., & Lyubomirsky, S. (2019). A new measure of life satisfaction: The Riverside Life Satisfaction Scale. *Journal of Personality Assessment*, *101*(6), 621–630. https://doi.org/10.1080/00223891.2018.1464457

McNeish, D., & Wolf, M. G. (2021). Dynamic fit index cutoffs for confirmatory factor analysis models. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000425

Mill, J. S. (1892). *On liberty*. Longmans, Green and Co.

Mokken, R. J. (2011). *A theory and procedure of scale analysis*. De Gruyter Mouton.

Molenaar, I. W., & Sijtsma, K. (1984). Internal consistency and re-liability in Mokken's nonparametric item response model. *Tijdschrift Voor Onderwijsresearch*, *9*(5), 257–268.

Molenaar, I. W., & Sijtsma, K. (1988). Mokken's approach to reli-ability estimation extended to multicategory items. *Kwantita-tieve Methoden*, *9*(28), 115–126.

Moore, M. T., & Fresco, D. M. (2012). Depressive realism: A meta-analytic review. *Clinical Psychology Review*, *32*(6), 496–509. https://doi.org/10.1016/j.cpr.2012.05.004

Navarro-Gonzalez, D., & Lorenzo-Seva, U. (2021). *EFA.MRFA: Dimen-sionality assessment using minimum rank factor analysis*. R package version 1.1.2. https://CRAN.R-project.org/package=EFA.MRFA

Parfit, D. (1984). *Reasons and persons*. OUP.

Peterson, H., & Engwall, K. (2019). "Why would You want a Baby when You could have a Dog?" Voluntarily childless Women's "Peternal" feelings, longing and Ambivalence. *Social Sciences*, *8*(4), Article 126. https://doi.org/10.3390/socsci8040126

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Revelle, W. (2022). *psych: Procedures for personality and psycho-logical research* (Version 2.2.5). Northwestern University. https://CRAN.R-project.org/package=psych

Riskey, D. R., & Birnbaum, M. H. (1974). Compensatory effects in moral judgment: Two rights don't make up for a wrong. *Journal of Experimental Psychology*, *103*(1), 171–173. https://doi.org/10.1037/h0036892

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity domi-nance, and contagion. *Personality and Social Psychology Review*, *5*(4), 296–320. https://doi.org/10.1207/S15327957PSPR0504_2

Rulli, T. (2016). The ethics of procreation and adoption. *Philosophy Compass*, *11*(6), 305–315. https://doi.org/10.1111/phc3.12330

Ruscio, J., Carney, L. M., Dever, L., Pliskin, M., & Wang, S. B. (2018). Using the comparison curve fix index (CCFI) in taxometric an-alyses: Averaging curves, standard errors, and CCFI profiles. *Psychological Assessment*, *30*(6), 744–754. https://doi.org/10.1037/pas0000522

Ruscio, J., Haslam, N., & Ruscio, A. M. (2013). *Introduction to the taxometric method: A practical guide*. Routledge.

Ruscio, J., & Wang, S. (2021). *RTaxometrics: Taxometric analysis*. https://cran.r-project.org/web/packages/RTaxometrics/index.html

Rye, B. J., & Underhill, A. (2020). Pro-Choice and Pro-Life are not enough: An Investigation of abortion attitudes as a function of abortion prototypes. *Sexuality & Culture*, *24*, 1829–1851. https://doi.org/10.1007/s12119-020-09723-7

Schoenegger, P., & Grodeck, B. (forthcoming). Concrete over ab-stract: Experimental evidence of reflective equilibrium in pop-ulation ethics. In H. Viciana, F. Aguiar, & A. Gaitán (Eds.), *Issues in moral experimental philosophy*. Routledge.

Schönegger, P. (2021). What's up with antinatalists? An observa-tional study on the relationship between dark triad personality traits and antinatalist views. *Philosophical Psychology*, *35*(1), 66–94. https://doi.org/10.1080/09515089.2021.1946026

Shiffrin, S. V. (1999). Wrongful life, procreative responsibility, and the significance of harm. *Legal Theory*, *5*(2), 117–148. https://doi.org/10.1017/s1352325299052015

Sijtsma, K., & Molenaar, I. W. (2002). *Introduction to nonparametric item response theory*. Sage.

Singh, A. (2018). The Hypothetical consent objection to anti-natalism. *Ethical Theory and Moral Practice*, *21*(5), 1135–1150. https://doi.org/10.1007/s10677-018-9952-0

Spears, D. (2017). Making people happy or making happy people? Questionnaire-experimental studies of population ethics and policy. *Social Choice and Welfare*, *49*(1), 145–169. https://doi.org/10.1007/s00355-017-1055-7

Straat, J. H., van der Ark, L. A., & Sijtsma, K. (2013). Comparing optimization algorithms for item selection in Mokken scale analysis. *Journal of Classification*, *30*, 75–99. https://doi.org/10.1007/s00357-013-9122-y

Tappin, B. M., & Capraro, V. (2018). Doing good vs. avoiding bad in prosocial choice: A refined test and extension of the morality preference hypothesis. *Journal of Experimental Social Psychology*, *79*, 64–70. https://doi.org/10.1016/j.jesp.2018.06.005

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, *16*(2), 209–220. https://doi.org/10.1037/a0023353

Van der Ark, L. A. (2012). New developments in Mokken Scale analysis in R. *Journal of Statistical Software*, *48*(5), 1–27. https://doi.org/10.18637/jss.v048.i05

Zuber, S., Venkatesh, N., Tännsjö, T., Tarsney, C., Stefánsson, H. O., Steele, K., Spears, D., Sebo, J., Pivato, M., Ord, T., Ng, Y-K., Masny, M., MacAskill, W., Lawson, N., Kuruc, K., Hutchinson, M., Gustafsson, J. E., Greaves, H., Forsberg, L., … Asheim, G. B. (2021). What should we agree on about the repugnant conclusion? *Utilitas*, *33*(4), 379–383. https://doi.org/10.1017/S095382082100011X

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF) Logistic regression modeling as a unitary framework for binary and likert-type (ordinal) item scores*. Directorate of Human Resources Research and Evaluation National Defense HeadquartersDepartment of National Defense. https://faculty.educ.ubc.ca/zumbo/DIF/handbook.pdf

Publication Ethics

Informed consent was obtained from all participants included in the study.

All procedures in studies involving human participants were performed in accordance with the ethical standards of the institution's Human Research Ethics Committee. Specifically, we received ethics approval from the University Teaching and Research Ethics Committee (UTREC) at the University of St Andrews. Approval code: SA15754.

Authorship

Philipp Schoenegger: Conceptualization, Methodology, Analysis, Writing – Original Draft, Writing – Review & Editing, Funding Acquisition. Maxmiliam Maier: Conceptualization, Methodology, Analysis, Writing – Original Draft Writing – Review & Editing. Rafael Bastos: Methodology, Analysis, Writing – Original Draft, Writing – Review & Editing. Philipp Schoenegger and Maximilian Maier share first authorship.

Open Science

Open Data: The authors confirm that there is sufficient information for an independent researcher to reproduce all of the reported results, including codebook if relevant (Maier et al., 2022).

Open Materials: The authors confirm that there is sufficient information for an independent researcher to reproduce all of the reported methodology (Maier et al., 2022).

Preregistration of Study 1: https://doi.org/10.17605/OSF.IO/W54S3
Preregistration of Study 2: https://doi.org/10.17605/OSF.IO/JFU9R
Preregistration of Study 3: https://doi.org/10.17605/OSF.IO/MJ7KW
Preregistration of Study 4: https://doi.org/10.17605/OSF.IO/VAUMD

Philipp Schoenegger

Department of Philosophy/School of Economics and Finance
University of St. Andrews
Edgecliffe
KY16 9AL/ Castlecliffe
The Scores
Fife KY16 9AR
Scotland
UK
ps234@st-andrews.ac.uk