

EMPIRICAL ARTICLE

Test–retest reliability of the SVO-Slider Measure and other psychometric measures of personality traits

Debora Frei¹, Stefan Wehrli², Alexander Ehler³, Fabian Winter³, and Heiko Rauhut³

¹Institute of Sustainable Development, ZHAW Zürich, Winterthur, Switzerland; ²Decision Sciences Laboratory, ETH Zürich, Zürich, Switzerland and ³Department of Sociology, UZH Zürich, Zürich, Switzerland

Corresponding author: Debora Frei; Email: frde@zhaw.ch

Received: 22 June 2024; **Revised:** 4 April 2025; **Accepted:** 4 April 2025

Keywords: social value orientation (SVO); Slider Measure (SLM); measurement methods; reliability; individual differences

Abstract

This study presents findings from a 4-year panel study examining three major questions regarding the measurement of social value orientation (SVO). First, we investigate the test–retest reliability of the Slider Measure (SLM, Murphy et al., 2014) over a period of up to 4 years in a large, demographically diverse sample. Second, we compare the stability of the SLM to related measurements of prosociality and distributional preferences along the behavior–behavioral tendency–trait continuum, including single behaviors (e.g., the Dictator Game and the Prisoner’s Dilemma), alternative behavioral tendencies (e.g., survey-based measures of altruism), and broader personality traits (e.g., Big-Five, HEXACO, Dark Factor *D*). Third, we explore differences in individual trajectories of SVO, focusing on how age and gender influence its stability and change over time. Our study thus complements earlier research on the stability of the SLM by extending the time period and depth of analysis, and putting the measure in the context of other related measures. The results show a considerable degree of stability, higher than all behavioral games, but often lower than fully fledged measures of personality traits. Furthermore, we find that age has a stabilizing effect on behavior in the SLM. With regard to gender, we find that women behave generally more prosocial than men but that they do not differ in their stability. We conclude that the SLM is a suitable method for assessing individual SVO over longer time periods and is best thought of as covering a sweet spot between stable personality traits and immediate behavioral expressions.

1. Introduction

Understanding how much we care about the impact of our behavior on others is fundamental to understanding human decision-making and addressing social and global challenges. From individual environmental preservation to income redistribution, differing perspectives on fairness and concern for others often shape individual behavior and societal outcomes. These distributional concerns, commonly referred to as social preferences or Social Value Orientation (SVO), provide valuable insights into the motivations behind individual decision-making and collective action. In particular, the SVO Slider Measure (SLM hereafter; Murphy et al., 2011) has become a widely used incentivizable tool for assessing these distributional preferences. However, while prior research has established its short-term reliability, key questions about its long-term reliability, comparative standing to other measures, and demographic influences remain unanswered. In this article, we present findings from a 4-year panel

study that explores these facets, focusing on the reliability of the SLM over time, the demographic influences on its fluctuations, and its position within the broader landscape of prosociality measures.

The SLM is an incentive-compatible measure of distributional preferences, involving several decisions on allocating points or real money between oneself and another person, typically a stranger. To date, only a handful of studies have explicitly examined the test–retest reliability of the SLM. Murphy et al. (2011) conducted three sessions 1 week apart, finding a high test–retest correlation (Pearson's $r = 0.915$), particularly when compared to other popular measures of prosociality like the Ring Measure (RM) ($r = 0.599$, Liebrand, 1984) and the Triple Dominance Measure (TDM) (Goodman and Kruskal's $\gamma = 0.391$, Van Lange et al., 1997). However, studies on the SLM over longer time frames are rare. One exception is Bakker and Dijkstra (2021), who employed a 4-wave design spanning 16 months to assess the temporal reliability of SVO and compare it with the RM and TDM. Due to the limited sample size of individuals participating in all 4 waves, their analysis focused mainly on the first 2 waves, 3 months apart, which included 89 respondents. Results showed high consistency, with 77.9% of participants remaining in the same distributional category, and a moderate correlation between waves ($r = 0.698$). Similarly, de Matos Fernandes et al. (2022) investigated the test–retest reliability of prosociality among 495 participants in a Dutch sample over six months. They assessed the consistency of the SLM in monthly intervals, finding that continuous social preferences were moderately correlated ($r \geq 0.66$) across waves. Höglinger and Wehrli (2016) use the first 2 waves of the present study to show that social preferences elicited with the SVO-Slider on MTurk have a similar distribution to those found in laboratory settings with a high test–retest reliability over 1 week (Pearson's $r = 0.79$). However, the SVO measure correlates only weakly with self-reported prosocial behavior items but, interestingly, considerably with the survey response time.

We extend the previous literature in three ways: First, we are able to report results on various aspects of the test–retest stability of the SLM over a much longer time period than in previous studies. Second, we contextualize our findings on SLM stability within the broader field of personality measurement by comparing our results with data from other measures of prosociality. Specifically, we collected behavioral measures, such as the Dictator Game (DG) and the Prisoner's Dilemma (PD), data on various self-reported behaviors, and a shortened version of the Big-Five Inventory (BFI-10). Furthermore, we compare our results with those of other studies using established personality measures, including the full Big-Five Inventory (Specht et al., 2011), the HEXACO model (Dunlop et al., 2021), and the Dark Factor of Personality (D , Zettler et al., 2021). Third, we contribute to the literature by not only reporting the test–retest reliabilities of the SLM across our entire sample, but by disaggregating the results to examine (i) differences in prosociality by age and gender, (ii) differences in stability by age and gender, and (iii) differences in the stability of individual items.

Our results indicate that on an individual level, behavior tends to become more self-serving over the six waves, yet older individuals are slightly more prosocial. This might suggest that different effects play a role, such as exposure vs. cohort effects, differing in size and direction. Moreover, women display consistently higher prosocial tendencies than men. When comparing measurement methods, personality scales like the Big-Five, HEXACO or the Dark Factor D , their relevant aspects exhibit higher stability than the SLM. However, comparable experimental games that focus on prosocial and cooperative behavior like the DG and PD capture more variable, context-dependent behavior. Thus, the SLM occupies a distinct position between these approaches, measuring incentivizable behaviors with greater reliability than single-game setups while reflecting specific behavioral tendencies rather than broad traits. This makes it a valuable tool for studying prosociality in a behavioral setting over time.

2. Methodology

2.1. Sample

We collected our data using the online labor market Amazon Mechanical Turk (MTurk) in six waves between November 2015 and December 2019 (see Höglinger and Wehrli, 2016 for more details on

Table 1. Sample description.

Wave	$\Delta_t(W_1, W_t)$	<i>N</i>	Age (Mean, SD)	Female (in %)	College (in %)
1	0	1,009	35.81 (11.46)	49.75	61.05
2	1 Week	820	36.10 (11.44)	48.90	63.05
3	26 Weeks (1/2 Year)	618	37.77 (11.72)	48.70	64.89
4	52 Weeks (1 Year)	662	37.20 (11.66)	50.45	62.99
5	104 Weeks (2 Years)	451	38.09 (12.22)	49.89	65.19
6	208 Weeks (4 Years)	397	38.77 (12.03)	49.87	65.24

Note: Descriptive data for each wave. College in % is percentage with college education.

specifics of the implementation and further assessed constructs, and Table 1 for demographics of our sample). For our study, we posted a HIT (Human Intelligence Task) open to US-residents asking for participation in a ‘Study on Decision Making’ that required completing 2 online surveys—one immediately, the other 1 week later. A base payment of \$2 (USD) was offered for completing the 2 surveys, with the possibility of earning up to an extra \$3 on various decision tasks (see Höglinger and Wehrli (2017) for an analysis of the first 2 waves). Participation in later waves was achieved by re-inviting participants at comparable incentives via the platform’s personalized invitation system.

We collected data over 4 years across six waves, with increasing intervals between each wave. Starting with $N=1,009$ participants in wave 1, we observed a gradual decline in participation over time, accompanied by slight shifts in the distributions of age, gender, and education. Specifically, being in college significantly reduced the likelihood of dropout in waves 2 and 3, while older individuals were more likely to drop out between waves 3 and 6 (see Table A10 in the Appendix). This is not unusual in longitudinal studies of this duration. Importantly, although the mean SVO angle decreases over time (see Table A4), we found no evidence of systematic dropout based on the SVO ($p > 0.05$ in all waves), ensuring the validity of our key variable of interest. This is corroborated when looking at the group means of all those individuals who participated in all six waves, where we also find a slight decrease in group means (see Table A9). Table 1 gives an overview of the sample. All data, methods and materials are publicly available at <https://osf.io/ygfwf>.

2.2. Measures

2.2.1. SVO SLM

The SLM measures a person’s social preferences based on a series of six primary allocation decisions that lead to an individual SVO angle (SVO°). In each decision, participants have to allocate points between themselves and another unknown individual (see Figure 1a). For example, in Item 5, the decision maker has to choose an amount between 50 and 100 points the other person receives, while their own payoff decreases by 1 point for every point the other’s payoff rises. They thus have to make a trade-off between their own welfare and that of the other person.

Figure 1b shows the six items and their specific payoff possibilities in a self/other allocation plane. The six connecting lines depict the payoff combinations that emerge when comparing the 4 idealized SVOs *competitive*, *individualistic*, *prosocial*, and *altruistic* with each other. The assessment of the six allocation decisions allows for the calculation of the individual SVO score (SVO°) and thus for a fine-grained differentiation of the SVOs. Geometrically, SVO° denotes the angle between a horizontal line and a line connecting the midpoint of the circle (50,50) and a point somewhere on the outer circle. SVO° thus indicates where someone’s social preferences fall on the altruism-competitiveness spectrum.¹

¹The formula for calculating the prosociality score is: $SVO^\circ = \arctan\left(\frac{(A_O - 50)}{(A_S - 50)}\right)$, where A_O and A_S denote the average allocation across all six sliders to the other and oneself, respectively. The idealized SVOs can be used to calculate cutoffs between different fairness types: competitive individuals with an $SVO^\circ < -12.04^\circ$, individualists with SVO° between

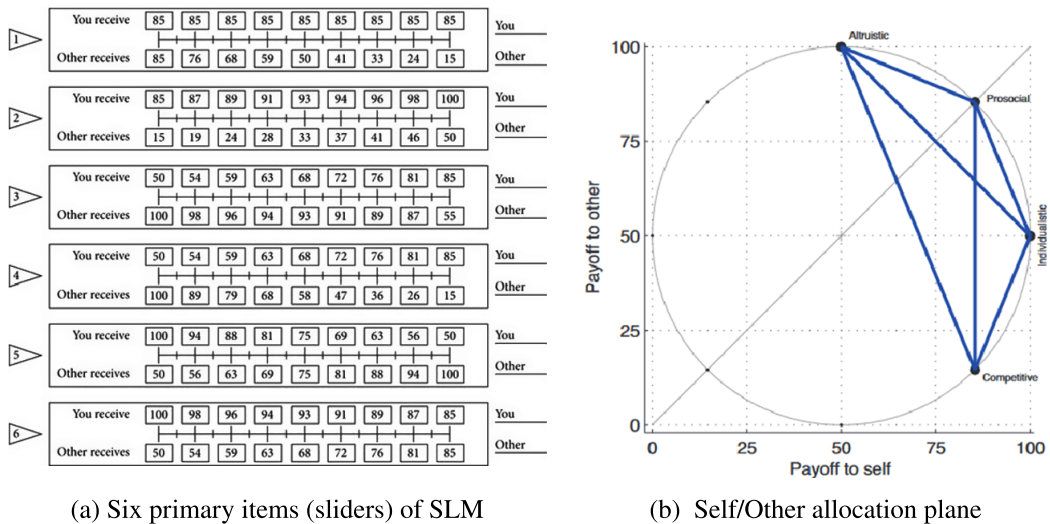


Figure 1. *SVO Slider Measure (quoted from Murphy & Ackermann (2014)).*

During our study, participants were randomly matched with anonymous partners and presented with the six allocation questions. At the end of the study, we randomly selected one decision that determined how much participants would be paid, with 100 points equal to \$0.50.

2.2.2. The PD

The PD is a measure of cooperativeness in mixed-motive situations (Rapoport and Chammah, 1965). Conceptually, it is closely related to the SVO, which is considered a good predictor of cooperative behavior (Balliet et al., 2009; Bogaert et al., 2008). In our implementation of the PD, two individuals independently decide whether to cooperate or not. If both cooperate, they both receive a moderate reward (60 each). If only one person cooperates, the cooperator receives nothing (0), while the defector receives a large reward (100). If both defect, they both receive a smaller reward than if they had cooperated (20 each). The PD was employed in waves 1, 2, 5, and 6.

2.2.3. DG

The DG is a measure of prosociality within a resource allocation scenario. In our implementation of the DG, one participant (the ‘dictator’) is given 100 points. The dictators can divide these points in 10-point increments between themselves and the receiver. The receiver has no say in how the points are allocated and must accept the offer made by the dictator. The DG was employed in waves 2 to 6.

2.2.4. Survey questions on altruistic behavior

As a fourth measure of prosociality, we employed survey questions on self-reported acts of altruism using a 14-item altruism scale from the 2002 General Social Survey (Smith et al., 2012). In the survey, participants were requested to report how frequently they have engaged in a particular altruistic behavior within the previous year, ranging from ‘never’ to ‘more than once a week’. For example, one of the survey items was ‘During the past 12 months, how often have you given food or money to a homeless person?’ (see complete survey questions in Table A1 in the Appendix). We evaluated the self-reported acts of altruism in all six waves.

(−12.04°, 22.45°), prosocial individuals with SVO° between (22.45°, 57.15°) and altruistic individuals with SVO° > 57.15°. We will focus on the continuous measure SVO° throughout this article.

2.2.5. A reduced version of the Big-Five Inventory (BFI-10)

Finally, we deployed a 10-item Big-Five Inventory (BFI-10, Rammstedt and John, 2007). Due to time restrictions, we chose this largely reduced version of the Big-Five instead of the more comprehensive Big-Five inventory by John et al. (1991). The BFI-10 is designed and validated to assess five important personality traits: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism on a Likert-scale from 1 to 5 (see complete survey questions in Table A2 in the Appendix). We used the BFI-10 as a benchmark for a well-established measurement method of personality with good psychometric properties (Rammstedt et al., 2013; Rammstedt et al., 2023). The BFI-10 questions were evaluated in waves 3 to 6.

3. Results

In order to verify the quality of the SLM as a valid measure of SVOs, we first examine general distributional findings and continue with the discussion of the test–retest reliability of the SLM as a whole and by specific items. We then compare the reliability of the SLM with other personality measures from our survey and the literature. Finally, we assess socio-demographic differences in the SLM’s test–retest reliability and substantive developments of SVO by age and gender in more detail.

3.1. Distribution of SVO°

When looking at the distribution of SVO° (see Figure 2), we find a wide range of SVO angles between -16.26° (competitive) and 61.39° (altruistic) with an average SVO angle across all waves and participants of $M(3957) = 23.86^\circ$ ($SD = 13.96$). Thus, on average, individuals have prosocial preferences. Across the six waves, most participants exhibit either individualistic or prosocial preferences, with only a few showing competitive tendencies and only one participant in 1 wave showing altruistic preferences. Notably, we observe two distinct spikes in the distribution of SVO angles near 10° and near 35° , indicating that there were consistently sized subsets of participants who exhibited either almost perfectly individualistic or prosocial preferences (i.e., maximized own or joint payoffs). Overall, fluctuations between the 4 ‘ideal types’ are minimal and range between 13.66% and 22.44% (see Table A3 for all percentage changes). All these findings replicate central findings of earlier studies on the SLM (de Matos Fernandes et al., 2022; Ehler et al., 2021; Liu et al., 2024; Van Doesum et al., 2021).

When focusing on the mean level consistency (Roberts et al., 2006), we find that the mean angle decreases slightly over the six waves, indicating a slight shift toward more selfish scores. When examining wave by wave, we observe an increase in the mean SVO angle between waves 1 and 2, but this trend is not sustained, as we find that the mean SVO angle between waves 2 and 3 decreases significantly ($t(556) = 3.30, p < 0.01$) and decreases afterwards even further. However, the effect sizes, i.e., Cohen’s d ’s, are rather small (all $d < 0.125$), so we can conclude that the group remains mostly at the same level and tends to become only slightly more selfish over time (see Table A4).

3.2. Test–retest reliability of SVO° and the individual sliders

3.2.1. General SVO°

To assess the test–retest reliability of the SLM over time, we leverage our extensive longitudinal data set by analyzing the rank-order stability of SVO°. The rank-order stability, or Spearman’s r , assesses the extent to which individuals maintain their relative standing in comparison to others with regard to their SVO° between 2 waves (Roberts et al., 2006). As shown in Table 2, there exists a strong positive correlation between SVO° across successive waves with all correlation coefficients above $r = 0.7$ (all correlations are statistically significant with $p < .01$).

To integrate our findings with other studies and emphasize long-term effects, we next focus on changes within the 2-year and 4-year intervals, enabling us to specifically explore patterns and

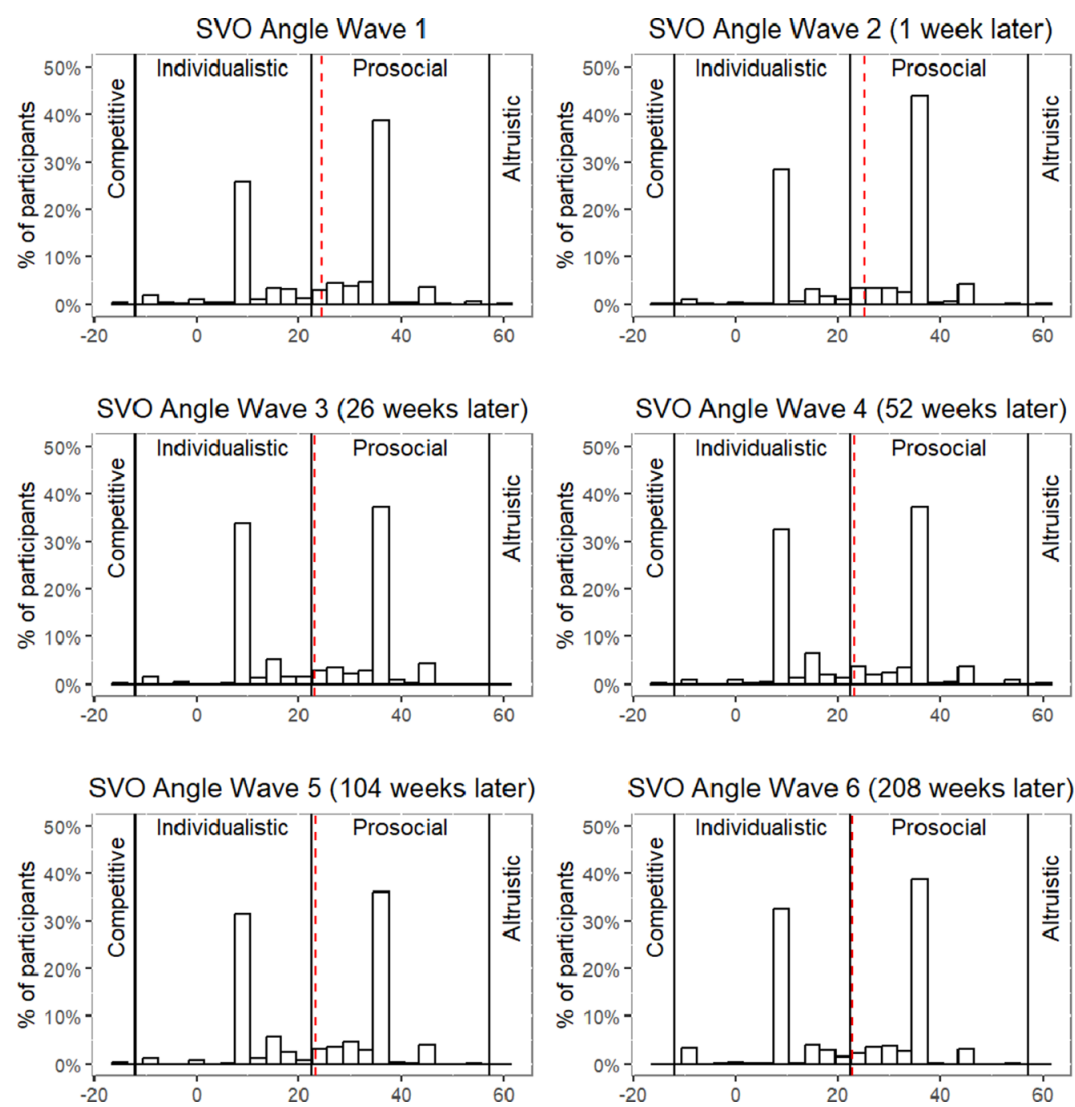


Figure 2. Histogram of SVO angles per wave (Mean is shown as red, dotted line).

Table 2. Rank-order stability between waves.

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6
Wave 1	1					
Wave 2	0.79***†	1				
Wave 3	0.67***†	0.71***†	1			
Wave 4	0.66***†	0.71***†	0.80***†	1		
Wave 5	0.64***†	0.68***†	0.73***†	0.81***†	1	
Wave 6	0.58***†	0.64***†	0.72***†	0.68***†	0.78***†	1

Note: Spearman's r between waves (* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$).

† Bonferroni-adjusted p -values $p_B < 0.05$.

transformations over the longer term. Thus, the rank-order stability between waves 1 and 5 (2 years) is $r_{2y} = 0.64$ and $r_{4y} = 0.58$ between waves 1 and 6 (4 years). These findings suggest that individuals tend to maintain their relative rank in the sample compared to other participants over time.

3.2.2. Individual sliders

Looking more closely at the six sliders that constitute the SVO (i.e., the trade-off between each combination of orientations), we find that the slider that considers the trade-off between altruism and competitiveness exhibits the highest rank-order stability for the 2- and 4-year comparison ($r_{2y} = 0.57$, $N = 451$, $p < 0.01$; $r_{4y} = 0.55$, $N = 397$, $p < 0.01$) (see Table A7 for all rank-order stabilities). In fact, all sliders in which one has the choice of whether or not to take the other's benefit into account (i.e., the slider, *Prosociality-Individualism*, *Altruism-Individualism*, *Prosociality-Competitiveness*) show a high test-retest reliability over time. This suggests that people are quite stable in their decisions about whether or not to consider the benefit of another person. However, when having to choose the actual degree of consideration (behaving prosocially vs. altruistically, or individualistically vs. competitively), people fluctuate more ($r_{2y} = 0.38$, $N = 451$, $p < 0.01$; $r_{4y} = 0.24$, $N = 397$, $p < 0.01$ and $r_{2y} = 0.26$, $N = 451$, $p < 0.01$; $r_{4y} = 0.31$, $N = 397$, $p < 0.01$). Hence, the decision whether to consider the benefit of others at all seems to be an underlying, stable tendency, while the exact form it takes is more variable and thus more situational.

3.3. Comparison with other prosociality measures and psychometric constructs

To contextualize the test-retest reliability of SVO, we first compare our findings about the SLM to several incentivized and unincentivized measures of altruism and cooperation in our data. When taking into account all rank-order stabilities, we find that on average, the BFI-10 items show a substantially higher rank-order stability than the SLM. This is consistent with the literature, in which the short version of the Big-Five is considered as very consistent over time (Rammstedt et al., 2023; Rieger et al., 2017; Robinson, 2009). Compared to the other cooperation measurement methods, we find that the SLM's rank-order stability is substantially higher than the DG and PD and comparable to the Survey-measured Altruism method (see Figure 3).²

When only taking the 2- and 4-year time frames into account, we find a similar picture regarding the rank-order stability of the BFI-10, the DG and the PD (see Table 3). As we assessed the BFI-10 starting from wave 3, the time spans are reduced by half a year. All five components of the Big-Five exhibit high reliability across the 6 waves, with nearly all rank order stabilities exceeding the 0.75 threshold and thus showing very good reliability, despite the reduced version of the Big-Five we used in our study (see also Table 3). This is consistent with previous results and trait assessments of the Big-Five (Cobb-Clark and Schurer, 2011). Notably, the 2 dimensions of the BFI-10 most closely related to the SLM, Agreeableness and Conscientiousness, show reliabilities between the SLM and the BFI-10's most stable dimensions Extraversion and Neuroticism. This suggests that preferences related to 'social' traits might be more malleable than more personality-centered traits. Thus, compared to the BFI-10, the SLM is more variable over the course of 4 years. The altruism-survey questions exhibit a slightly higher rank-order stability than the SVO score for comparisons of 2 and 4 years.

While the informative value of the BFI-10 is limited due to its brevity, a comparison of our data with other long-term studies shows a similar picture. Table 4 contrasts our results on the SLM ($r_{2y} = 0.64$, and $r_{4y} = 0.58$) with three well-established personality trait measures closely related to the concepts measured in the SLM. A comparison to the full inventory of the Big-Five is based on data by Specht et al. (2011) collected over 4 years. They find test-retest reliability scores between $r_{4y} = 0.64$ ($p < 0.001$) for Conscientiousness and $r_{4y} = 0.74$ ($p < 0.001$) for Extraversion. Agreeableness, the arguably

²As the questions relate to self-reported behavior in the past year, those items that are easily verifiable (such as 'How often did you donate to a charitable organization in the past year?') are the least susceptible to change. In contrast, items that relate to activities that are not so easily traceable (such as 'How often did you lend a tool to someone you don't know well?') exhibit greater variability.

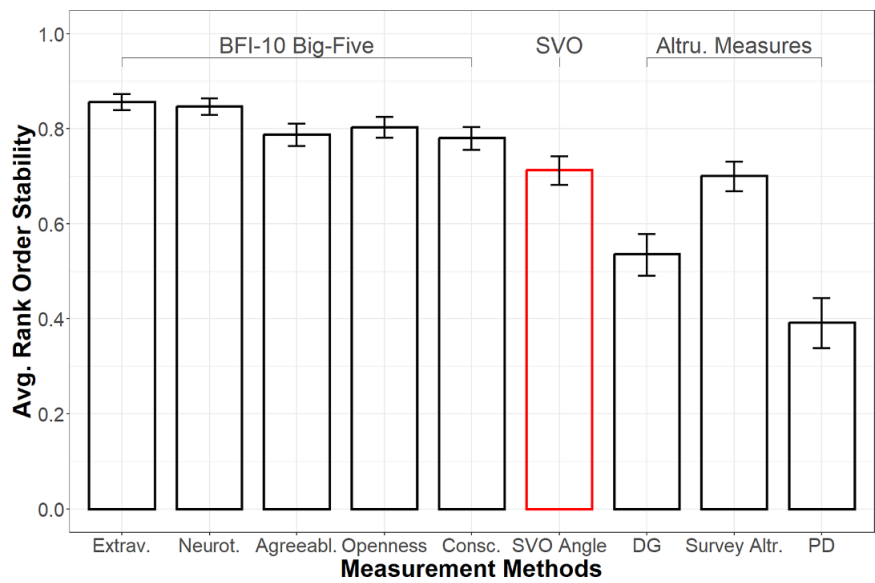


Figure 3. Fisher-z corrected ROS of different measures including BFI–10 (4 waves, short version of Big-Five), SLM (6 waves, SVO Angle), DG (5 waves), Survey-measured altruism (6 waves, Survey Altru.) and Prisoner’s Dilemma (4 waves, PD).

Table 3. Rank-order stability: SVO vs. BFI–10 and other cooperation measurements.

Measure	N	Time in years	Rank-order stability	p-value
SVO Slider Measure (SLM)	451	2	0.64	<0.01
	397	4	0.58	<0.01
BFI–10 (short Big-Five)				
Neuroticism	371	1.5	0.84	<0.01
	328	3.5	0.82	<0.01
Extraversion	373	1.5	0.87	<0.01
	330	3.5	0.84	<0.01
Openness	371	1.5	0.81	<0.01
	328	3.5	0.78	<0.01
Agreeableness	371	1.5	0.79	<0.01
	329	3.5	0.74	<0.01
Conscientiousness	372	1.5	0.76	<0.01
	327	3.5	0.74	<0.01
Other Behavioral Measures				
Dictator Game (DG)	397	2	0.45	<0.01
	346	4	0.45	<0.01
Prisoner’s Dilemma (PD)	447	2	0.33	<0.01
	395	4	0.29	<0.01
Survey-based Altruism	216	2	0.70	<0.01
	194	4	0.71	<0.01

Table 4. Rank-order stability: Comparison of SVO with measurements of other studies.

Measure	Time in years	Rank-order stability	<i>p</i> -Value
SVO Slider Measure	2	0.64	<0.001
	4	0.58	<0.001
Big-Five (Specht et al., 2011)			
Emotional stability	4	0.73	<0.001 [†]
Extraversion	4	0.74	<0.001 [†]
Openness	4	0.72	<0.001 [†]
Agreeableness	4	0.68	<0.001 [†]
Conscientiousness	4	0.64	<0.001 [†]
HEXACO (Dunlop et al., 2021)			
Honesty-humility	2	0.801	<0.001 [†]
Agreeableness	2	0.749	<0.001 [†]
Conscientiousness	2	0.743	<0.001 [†]
Dark Factor <i>D</i> (Zettler et al., 2021)	4	0.81	<0.01

Note: [†] Bonferroni-adjusted *p*-values $p_B < 0.05$.

most closely related construct to the SLM, shows $r_{4y} = 0.68$ ($p < 0.001$). Dunlop et al. (2021) tested the test–retest reliability of the Honesty-humility, Agreeableness, and Conscientiousness scales from the HEXACO inventory over a period of 2 years among young adults and found $r_{2y} > 0.74$ for all 3 dimensions (see Table 4). Finally, Zettler et al. (2021) report the test–retest reliability of the Dark factor *D* over a period of 4 years with $r_{4y} = 0.81$ ($p < 0.01$), an excellent degree of reliability. Taken together, these instruments designed to measure personality traits have considerably higher test–retest reliability than the already fairly reliable SLM.

3.4. Age and gender differences in stability of SLM

To assess how SVO angles vary with time, gender and age at wave 1, we employ a linear mixed-effects model with random intercepts and random slopes for time (see Table A5). The inclusion of random intercepts allows each individual to have a unique baseline SVO angle, capturing the inherent variability in starting levels. Additionally, by incorporating random slopes for time, we account for the heterogeneity in individual trajectories, and, thus, the rate at which SVO angles change over time. The fixed effects include baseline age (i.e., age at wave 1), the time interval since the first wave (in years), and gender. A preliminary analysis on whether nonlinear effects of age might improve model fit was conducted by testing quadratic and cubic specifications. However, both AIC and BIC were lowest for the linear model, and likelihood ratio tests indicated that adding quadratic or cubic terms did not significantly improve model fit (all $p > 0.05$). Therefore, we retained the linear specification for age in our final model.

When assessing the random effects, we find substantial variability between individuals in both their starting SVO levels (*Intercept* variance = 142.95, SD = 11.96) and their trajectories over time (*TimeInterval* variance = 4.66, SD = 2.16, corr = −0.25). The negative correlation between the random intercept and slope suggests that, throughout the sample, those individuals with initially higher SVO angles tend to experience a steeper decline in SVO angles over time.

When assessing the fixed effects, we observe two counteracting effects of time: a negative longitudinal effect (waves) and a positive cross-sectional effect (age). On the one hand, participants become more egoistic over the 4 years of the study, highlighted by the significant decline in the SVO

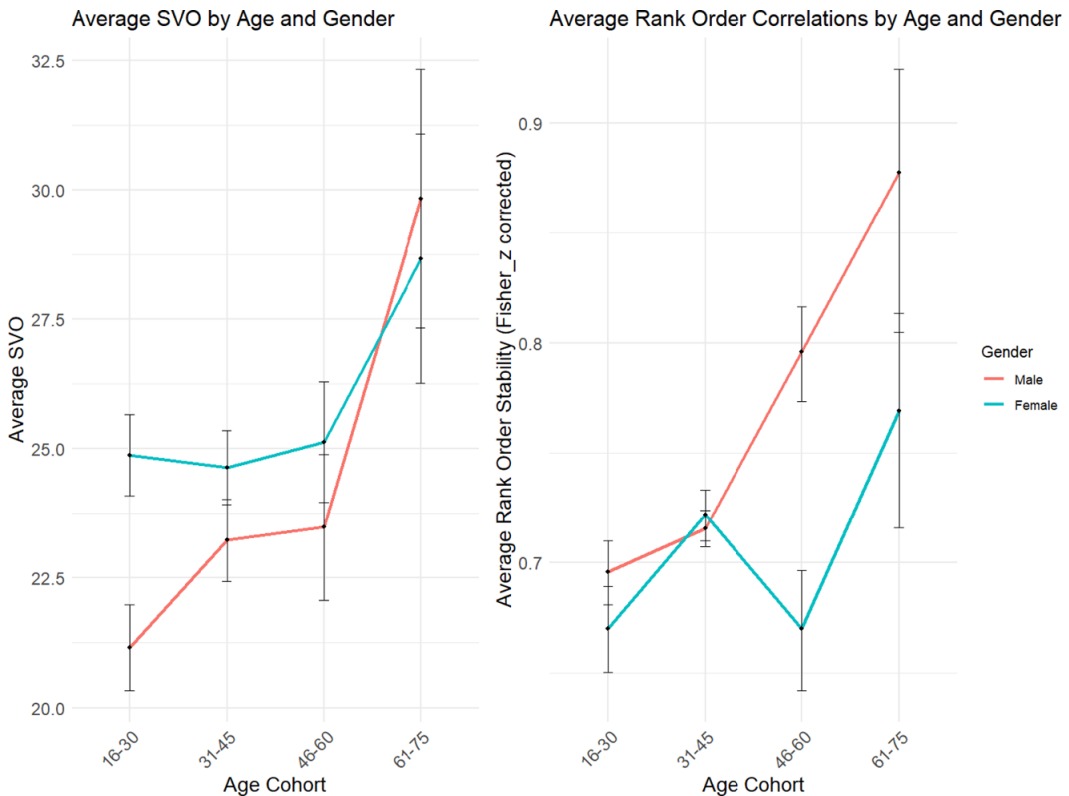


Figure 4. (a) Average SVO angles by age and gender across all waves and (b) average rank order stability by age cohort and gender (Fisher z corrected).

angles over time ($TimeIntervalYears = -0.577$, $SE = 0.140$, $p < 0.001$). This could be considered a longitudinal effect, potentially driven by exposure to the task or short term changes beyond the measurement procedure.

On the other hand, older individuals have slightly higher SVO scores ($Age\ in\ 10\text{-}year\ steps = 1.05$, $SE = 0.48$, $p = 0.029$). This cross-sectional effect could indicate a cohort effect. It is worth noting that the longitudinal effect is 5 times bigger than the cross-sectional effect, but both effects are substantively small (0.5 resp. 0.1 degrees per year). These two effects, however, are also diminishing with age: we find that age moderates the overall decline in SVO scores over time ($Time\ Interval * Age = 0.218$, $SE = 0.115$, $p = 0.0582$). Therefore, the negative effect of time on the SVO angle is less pronounced in older individuals, which also contributes to their higher SVO angle. Regarding gender differences, women have higher SVO scores than men ($Female = 2.55$, $SE = 0.78$, $p < 0.001$), suggesting greater prosocial behavior among women. However, we find no gender differences in the age-related increase in SVO ($Age * Female = -0.77$, $SE = 0.66$, $p = 0.242$). Similarly, Figure 4a shows no gender differences in the older age categories except for the youngest cohort. Thus, both genders show similar trajectories over time.

To further evaluate SVO stability of participants over time, we calculate the *SVO Change*. This variable represents the change in each participant's SVO angle for each wave combination. Thus, we use a multilevel model to examine whether these changes are influenced by the time interval between assessments, as well as by the participants' age and gender (see Table A6 Model 2). As expected, longer intervals between assessments are associated with greater variability in SVO ($Time\ Interval: 0.532$, $p < 0.001$). Furthermore, older participants change their SVO angles less ($Age: -0.453$, $p = 0.0336$). On the other hand, gender does not play a role regarding the stability of SVO scores ($p > 0.05$; see

also Figure 4b for a visualization). Thus we can conclude that age is not only relevant for the level of prosociality but also for its stability.

4. Discussion

The importance of investigating prosocial behavior and attitudes has grown over the past decades, particularly in the context of recent discussions on environmental protection, climate change and migration (Van Lange and Rand, 2022). The need to study prosociality is driven by the fact that many goods and resources are becoming increasingly scarce and distributional decisions and compromises have to be made, which are often affected by the different prosocial preferences of individuals. Despite a plethora of tools for measuring distributional preferences as an expression of prosociality, the reliability of these instruments remains uncertain, and determining the most effective among them is a pressing challenge. Notably, few of these tools have been tested over a long period but rather in very short time frames. We, therefore, conducted a 4-year-long comprehensive analysis, gathering data in 6 waves on various metrics of prosociality.

Overall, we observe a slight overall decrease in the sample mean of prosociality. Furthermore, we observed characteristic distributional spikes formed by large groups of individualists and prosocials across the waves. This corresponds to the results of most other studies using the SLM, which also find two large groups within the individualists and prosocials and hardly categorize any individual as competitive or altruistic (Murphy and Ackermann, 2014; Murphy et al., 2011; Rahal et al., 2020; Van Doesum et al., 2021).

On the individual level, the SLM shows high reliability, both in the short and in the long term, indicated by the high rank-order stability across all waves. The findings underscore the reliability of the SLM as a measure of distributional preferences and are consistent with previous studies (e.g., de Matos Fernandes et al., 2022; Murphy et al., 2011).

When positioning the SLM in the landscape of different measures (behavioral expressions) of prosociality, we find that the reliabilities of measures such as the BFI-10, the Big-Five or HEXACO are clearly higher than the reliability of the SLM. On the other hand, other incentivized games that measure cooperation or prosociality, such as the frequently employed DG or the Prisoner's Dilemma (PD), exhibit a markedly higher fluctuation over time. This suggests that the SLM can assess prosociality (regarding distributional preferences) more reliably than single-choice games such as the DG or the PD. Regarding the survey-based measurement of altruism, some variation in answers was expected, as the questions refer to very specific behaviors that may change depending on specific life situations. For example, certain behaviors such as 'watering someone's plants' or 'helping someone to find a job' depend very much on whether the current life situation in terms of time, circumstances, and opportunity allows it and if one can keep track of such events. The fact that the individuals nevertheless answered quite consistently indicates the reliability of the survey measure and its possibility of capturing altruism not only on a situational basis.

Interestingly, we find two counteracting effects associated with time. On average, our participants become (slightly) more egoistic with each wave of participation over up to 2 years. This effect could be due to some form of task familiarity or external effects (e.g., changes in the economy or shifts in social discourses over fairness). On the other hand, older people are somewhat more prosocial than younger participants. While we cannot experimentally test for these effects, they could for instance be driven by cohort effects.

Finally, we find that age plays a significant role in SVO stability, with older individuals displaying more consistent orientations compared to younger cohorts, who show greater variability. This variability in younger cohorts could hint at their ongoing identity development during early adulthood, where social contexts and experiences play a major role (Roberts and DelVecchio, 2000). Gender differences are minor, with women being more prosocial in general but not significantly more variable overall. Item-specific analyses indicate that individuals are particularly stable in deciding whether or not to consider another's benefit, compared to decisions about the actual magnitude of the other's benefit.

Overall, our results suggest that SVO stability is shaped by demographic factors and specific thematic characteristics. Future studies focusing on younger populations or life transitions could provide further insight into when and how SVOs stabilize (e.g., Ehler et al., 2020).

The study is subject to some limitations. The sample is composed exclusively of US-American individuals. Although several population groups are represented, such as South Americans and individuals from the Asia-Pacific region, the US still mainly reflects a highly individualistic culture (Hofstede et al., 2010). A cross-cultural study could be beneficial at this point and better capture the idiosyncrasies of different cultures. Furthermore, long-term assessments often suffer from the fact that many participants leave the study between waves, and dropping-out might correlate with the dependent variable. In our study, however, we find only very limited evidence for selective non-participation. Accordingly, our study clarifies and contributes to the question of the reliable measurement of social preferences and provides the possibility to use this construct more effectively in research and hopefully to help address pressing social and economic issues.

Our results indicate that it might be most fruitful to think of the SLM as a reliable instrument for the measurement of a specific behavioral tendency that bears significant implications, particularly in incentivized contexts. In contradistinction to comprehensive traits, which encompass consistent patterns of thought, emotion, and behavior (Funder and Fast, 2010), the SLM focuses exclusively on behavior. It has been shown to offer higher reliability than single games such as the DG or the PD. It is imperative to recognize the SLM as a behavioral manifestation of a trait (e.g., Agreeableness, the D factor, etc.), characterized by its high reliability. However, it is crucial to note that the SLM captures only a single facet of that trait. Other behavioral expressions, such as dishonesty, cheating, or vengeance, are not intentionally measured by the SLM. Given its narrow focus on a specific behavioral expression, the SLM is not expected to match the stability of the underlying trait itself. Nevertheless, it provides a significantly more stable measure of behavior compared to many alternative methods. For instance, typical games have shown test–retest correlations of $r_{2y} = .45$ for the DG and $r_{2y} = .33$ for the PD, whereas the SLM demonstrates considerably greater stability.

Data availability statement. Please retrieve the data from the following link: <https://osf.io/ygfwf/>.

Acknowledgements. We are very grateful for the extensive and helpful feedback from Antonia Velicu, Julia Jerke, Justus Rathmann, Nico Mutzner and Vincent Oberhauser and the fruitful discussion with participants from the course *Advanced Topics in Social Theory and Quantitative Methods*.

Funding statement. This work was supported by Grants BSSGIO_155981 and 10001A_176333 from the Swiss National Science Foundation.

Competing interests. We declare that this research was conducted in the absence of any commercial and financial relationship that could be construed as a potential conflict of interest.

Ethical standards. This research meets all ethical guidelines, including adherence to the legal requirements of the studied countries.

References

- Bakker, D. M., & Dijkstra, J. (2021). Comparing the slider measure of social value orientation with its main alternatives. *Social Psychology Quarterly*, 84(3), 235–245. <https://doi.org/10.1177/01902725211008938>
- Balliet, D., Parks, C., & Joireman, J. (2009). Social value orientation and cooperation in social dilemmas: A meta-analysis. *Group Processes & Intergroup Relations*, 12(4), 533–547. <https://doi.org/10.1177/1368430209105040>
- Bogaert, S., Boone, C., & Declerck, C. (2008). Social value orientation and cooperation in social dilemmas: A review and conceptual model. *British Journal of Social Psychology*, 47(3), 453–480.
- Cobb-Clark, D. A. & Schurer, S. (2011). The stability of big-five personality traits. *Economics Letters*, 115(1), 11–15.
- de Matos Fernandes, C. A., Bakker, D. M., & Dijkstra, J. (2022). Assessing the test-retest reliability of the social value orientation slider measure. *Judgment and Decision Making*, 17(1), 31–49. <https://doi.org/10.1017/S1930297500009013>
- Dunlop, P. D., Bharadwaj, A. A., & Parker, S. K. (2021). Two-year stability and change among the honesty-humility, agreeableness, and conscientiousness scales of the hexaco100 in an Australian cohort, aged 24–29 years. *Personality and Individual Differences*, 172, 110601.

- Ehlert, A., Böhm, R., Fleiß, J., Rauhut, H., Rybníček, R., & Winter, F. (2021). The development of prosociality: Evidence for a negative association between age and prosocial value orientation from a representative sample in Austria. *Games*, 12(3), 67. <https://doi.org/10.3390/g12030067>
- Ehlert, A., Kindschi, M., Algesheimer, R., & Rauhut, H. (2020). Human social preferences cluster and spread in the field. *Proceedings of the National Academy of Sciences*, 117(37), 22787–22792.
- Funder, D. C. & Fast, L. A. (2010). Personality in social psychology. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., pp. 668–697). John Wiley & Sons, Inc.. <https://doi.org/10.1002/9780470561119.socpsy001018>
- Hofstede, G. H., Hofstede, G. J. & Minkov, M. (2010). *Cultures and organizations: Software of the mind: Intercultural cooperation and its importance for survival* (3rd ed). McGraw-Hill.
- Höglinger, M., & Wehrli, S. (2016). A study on decision making. Documentation, ETH Zürich.
- Höglinger, M., & Wehrli, S. (2017). Measuring social preferences on Amazon mechanical turk. In B. Jann and W. Przepiorka (Eds.), *Social dilemmas, institutions, and the evolution of cooperation*, Berlin, Boston: De Gruyter Oldenbourg, 527–546. <https://doi.org/10.1515/9783110472974-025>
- John, O. P., Donahue, E. M., & Kentle, R. L. (1991). Big five inventory. *Journal of Personality and Social Psychology*. <https://doi.org/10.1037/07550-000>
- Liebrand, W. B. G. (1984). The effect of social motives, communication and group size on behaviour in an N-person multi-stage mixed-motive game. *European Journal of Social Psychology*, 14(3), 239–264. <https://doi.org/10.1002/ejsp.2420140302>
- Liu, Y., Stivers, A. W., Murphy, R. O., Van Doesum, N. J., Joireman, J., Gallucci, M., Aharonov-Majar, E., Athenstaedt, U., Bai, L., Böhm, R. et al. (2024). Wherefore art thou competitors? How situational affordances help differentiate among prosociality, individualism, and competition. *European Journal of Personality*, 0(0). <https://doi.org/10.1177/08902070241298850>
- Murphy, R. O., & Ackermann, K. A. (2014). Social value orientation: Theoretical and measurement issues in the study of social preferences. *Personality and Social Psychology Review*, 18(1), 13–41. <https://doi.org/10.1177/1088868313501745>
- Murphy, R. O., Ackermann, K. A., & Handgraaf, M. J. J. (2011). Measuring social value orientation. *Judgment and Decision Making*, 6(8), 771–781. <https://doi.org/10.1017/S1930297500004204>
- Rahal, R.-M., Fiedler, S., & De Dreu, C. K. W. (2020). Prosocial preferences condition decision effort and ingroup biased generosity in intergroup decision-making. *Scientific Reports*, 10(1). <https://doi.org/10.1038/s41598-020-64592-2>
- Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and German. *Journal of Research in Personality*, 41(1), 203–212. <https://doi.org/10.1016/j.jrp.2006.02.001>
- Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C., & Kovaleva, A. (2013). A short scale for assessing the big five dimensions of personality: 10 item big five inventory (bfi-10). *Methods, Data Analyses*, 7(2), 17.
- Rammstedt, B., Roemer, L., & Lechner, C. M. (2023). Consistency of the structural properties of the BFI-10 across 16 samples from eight large-scale surveys in Germany. *European Journal of Psychological Assessment* 40(3), pp. 204–215. <https://doi.org/10.1027/1015-5759/a000765>
- Rapoport, A., & Chammah, A. M. (1965). *Prisoner's dilemma: A study in conflict and cooperation* (Vol. 165). University of Michigan Press.
- Rieger, S., Göllner, R., Spengler, M., Trautwein, U., Nagengast, B., & Roberts, B. W. (2017). Social cognitive constructs are just as stable as the big five between grades 5 and 8. *AERA Open*, 3(3), 2332858417717691.
- Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*, 126(1), 3.
- Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin*, 132(1), 1–25. <https://doi.org/10.1037/0033-2909.132.1.1>
- Robinson, O. C. (2009). On the social malleability of traits: Variability and consistency in big 5 trait expression across three interpersonal contexts. *Journal of Individual Differences*, 30(4), 201–208.
- Smith, T. W., Marsden, P., Hout, M., & Kim, J. (2012). *General social surveys*. National Opinion Research Center.
- Specht, J., Egloff, B., & Schmukle, S. C. (2011). Stability and change of personality across the life course: The impact of age and major life events on mean-level and rank-order stability of the big five. *Journal of Personality and Social Psychology*, 101(4), 862.
- Van Doesum, N. J., Murphy, R. O., Gallucci, M., Aharonov-Majar, E., Athenstaedt, U., Au, W. T., Bai, L., Böhm, R., Bovina, I., Buchan, N. R., Chen, X.-P., Dumont, K. B., Engelmann, J. B., Eriksson, K., Euh, H., Fiedler, S., Friesen, J., Gächter, S., Garcia, C., & Van Lange, P. A. M. (2021). Social mindfulness and prosociality vary across the globe. *Proceedings of the National Academy of Sciences*, 118(35), e2023846118. <https://doi.org/10.1073/pnas.2023846118>
- Van Lange, P. A. M., De Bruin, E. M. N., Otten, W., & Joireman, J. A. (1997). Development of prosocial, individualistic, and competitive orientations: Theory and preliminary evidence. *Journal of Personality and Social Psychology*, 73(4), 733–746. <https://doi.org/10.1037/0022-3514.73.4.733>
- Van Lange, P. A., & Rand, D. G. (2022). Human cooperation and the crises of climate change, covid-19, and misinformation. *Annual Review of Psychology*, 73, 379–402. <https://doi.org/10.1146/annurev-psych-020821-110044>
- Zettler, I., Moshagen, M., & Hilbig, B. E. (2021). Stability and change: The dark factor of personality shapes dark traits. *Social Psychological and Personality Science*, 12(6), 974–983.

A. Appendix

A.1 Survey items: Altruism, BIG5

Table A1. *Past altruistic behavior.*

During the past 12 months, how often have you done each of the following things:
Given food or money to a homeless person
Returned money to a cashier after getting too much change
Allowed a stranger to go ahead of you in line
Done volunteer work for charity
Given money to a charity
Offered your seat in a bus or in a public place to a stranger who was standing
Looked after a person's plants, mail or pets while they were away
Carried a stranger's belongings, like groceries, a suitcase or shopping bag
Given direction to a stranger
Let someone you didn't know well borrow an item of some value like dishes or tools
Help someone outside of your household with housework or shopping
Lent quite a bit of money to another person
Spent time talking to a person who was a bit down or depressed
Helped someone to find a job
Answer categories: 'Never', 'Once (year)', 'At least 2-3x (year)', 'Once (month)', 'Once (week)', 'More than once (week)'.

Table A2. *Big-Five Inventory questions.*

Dimension	Question
Extraversion	I see myself as: is reserved (R)
	I see myself as: is outgoing, sociable
Agreeableness	I see myself as: is generally trusting
	I see myself as: tends to find fault with others (R)
Conscientiousness	I see myself as: tends to be lazy (R)
	I see myself as: does a thorough job
Neuroticism	I see myself as: is relaxed, handles stress well (R)
	I see myself as: gets nervous easily (R)
Openness to Experience	I see myself as: has few artistic interests
	I see myself as: has an active imagination

Note: Items containing an (R) are reversed coded.

A.2 Additional analyses: Group change and multilevel analyses

Table A3. *Percentage change of group membership between waves.*

	1	2	3	4	5	6
1	0.00					
2	13.66	0.00				
3	19.90	19.21	0.00			
4	20.54	19.73	14.29	0.00		
5	20.62	20.54	16.98	14.25	0.00	
6	22.17	22.44	15.27	18.54	14.43	0.00

Note: % of participants that changed their SVO group between waves.

Table A4. *SVO angle means.*

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6
Mean	24.47	25.18	23.02	23.27	23.16	22.72
(SD)	(14.19)	(13.77)	(13.84)	(13.83)	(13.80)	(14.23)
T-Stat W_1, W_t		-1.67	2.61**	3.30***†	1.76	2.44*
Cohen's d W_1, W_t		-0.051	0.10	0.086	0.093	0.125

Note: Pairwise t-tests * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

† Bonferroni-adjusted p-values $p_B < 0.05$.

The following Tables A5 and A6 depict multilevel-models assessing the influence of time interval, gender and age on the SVO angle and on the difference of the SVO angle between waves.

Table A5. *SVO Scores: Influence of age, gender, and time.*

Model	Dependent variable:		
	(1)	SVO Score (2)	(3)
Constant	23.967*** (0.393)	23.092*** (0.564)	23.123*** (0.564)
Time interval		−0.525*** (0.137)	−0.577*** (0.140)
Age (at wave 1)		0.840* (0.330)	1.050* (0.480)
Female		2.527*** (0.785)	2.548*** (0.784)
Age (at wave 1)* Female			−0.772 (0.660)
Time interval * Age (at wave 1)			0.218. (0.115)
Random effects			
Residual Std.DV	7.203	7.171	7.175
ID Std.DV	11.979	11.959	11.955
Groups	1,040	1,002	1,002
Observations	3,957	3,943	3,943
Log likelihood	−14,772.740	−14,701.970	−14,700.250
Akaike Inf. Crit.	29,555.480	29,419.940	29,420.490
Bayesian Inf. Crit.	29,586.900	29,470.180	29,483.290

Note: $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; Time interval depicts the time (in years) passed since first wave; Age is centered and depicts 10 year steps.

Table A6. Stability of SVO scores: Influence of age, gender, and time.

Model	Dependent variable:		
	Absolute change in SVO		
	(1)	(2)	(3)
Constant	5.842*** (0.249)	5.308*** (0.372)	5.262*** (0.373)
Time interval		0.532*** (0.116)	0.578** (0.120)
Age (at wave 1)		−0.453* (0.213)	−0.316 (0.235)
Female		0.079 (0.499)	0.074 (0.499)
Time interval * Age (at wave 1)			−0.129 (0.095)
Random effects			
Residual Std.DV	6.526	6.521	6.521
ID Std.DV	5.289	5.231	5.232
Groups	1022	922	922
Observations	4,207	4,202	4,202
Log likelihood	−14,409.610	−14,380.920	−14,381.440
Akaike Inf. Crit.	28,829.220	28,777.850	28,780.880
Bayesian Inf. Crit.	28,860.940	28,828.590	28,837.970

Note: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; Time interval depicts the time (in years) passed since first wave; Age is centered and depicts 10 year steps.

A.3 Item specific analysis of stability

Table A7. Rank-order stability of 6 sliders.

#	Slider	r_{2y}	r_{4y}
1	Prosocial - Competitive	0.43***†	0.51***†
2	Individualistic - Competitive	0.26***†	0.31***†
3	Altruistic - Prosocial	0.38***†	0.24***†
4	Altruistic - Competitive	0.57***†	0.55***†
5	Altruistic - Individualistic	0.55***†	0.48***†
6	Prosocial - Individualistic	0.56***†	0.45***†

Note: Spearman's r between waves (* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$);

† Bonferroni-adjusted p -values $p_B < 0.05$.

To assess influences of other factors such as age or gender, we build individual linear mixed models for each item with absolute differences between assessment intervals as the dependent variable and age, gender, interval length, and their interaction as fixed effects. Each model includes a random intercept

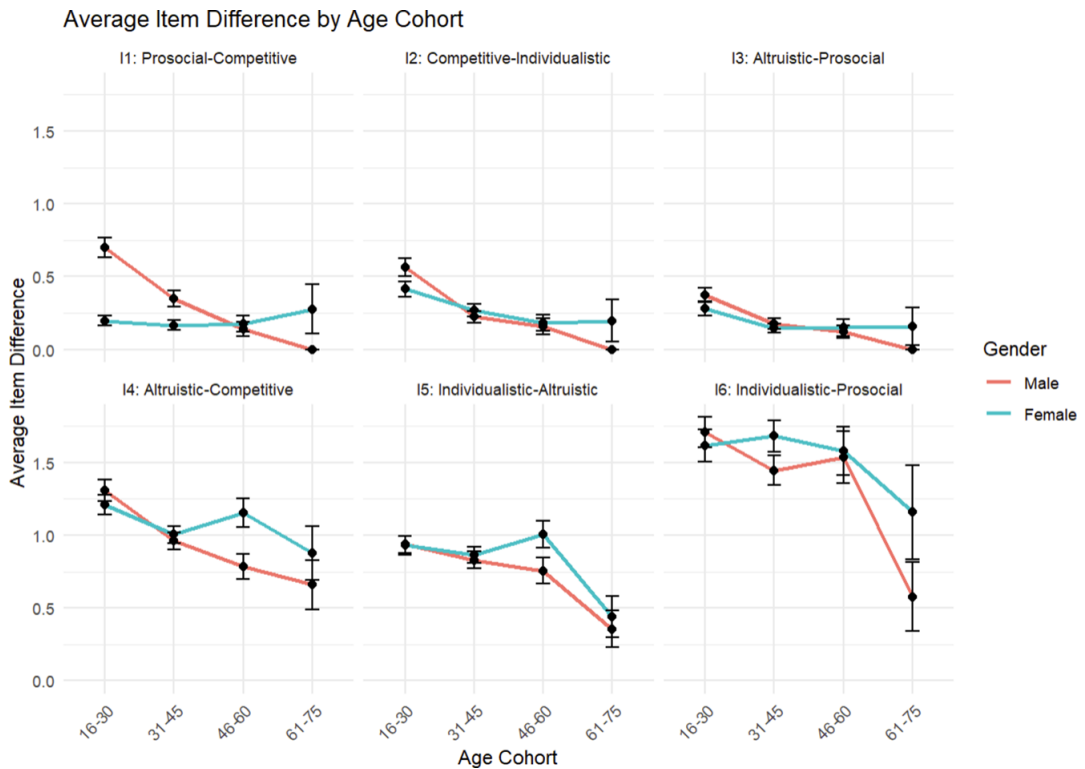


Figure A1. Itemwise assessment of stability according to age cohort and gender.

for participants to account for individual differences in baseline stability. As expected, interval length has a strong influence on nearly all items, with longer intervals between assessments leading to reduced stability. Only the comparison *Altruism-Prosociality* is unaffected by interval length ($\text{coef.} = -0.0001$, $p > 0.1$), suggesting that this comparison may capture a behavior that remains stable over time.

Age emerges as a significant stabilizing factor for the comparison between *Competitiveness* and any other value orientation. Older participants showed less difference between assessments ($\text{coef.} (\text{Comp-Prosoc}) = -0.019$; $\text{coef.} (\text{Comp-Indiv}) = -0.015$; $\text{coef.} (\text{Comp-Altru}) = -0.015$, $p < 0.001$). This suggests that younger people vary more in their decisions about whether or not to engage in competitive behavior. Thus, age seems to have a stabilizing effect on competitiveness.

Gender differences are most prominent for the comparison *Prosociality-Competitiveness*, where females demonstrate significantly greater stability than males ($\text{coef.} = -0.989$, $p < 0.001$). For this comparison, also a significant age-gender interaction is found ($\text{coef.} = 0.021$, $p < 0.001$), indicating that age has a stabilizing effect for men regarding the choice between prosociality and competitiveness (see Figure A1).

A.4 Analysis for full participants

Table A8. *Group change: Percentage change of group membership between waves.*

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6
Wave 1	0.00					
Wave 2	11.48	0.00				
Wave 3	22.95	20.49	0.00			
Wave 4	18.85	18.85	13.93	0.00		
Wave 5	21.31	21.31	15.16	13.11	0.00	
Wave 6	22.95	22.95	16.39	17.21	14.75	0.00

Note: % of participants (participating in all 6 waves) that changed their SVO group between waves.

Table A9. *SVO angle: Means per wave for full participants..*

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6
Mean (SD)	23.52 (14.02)	23.10 (13.46)	22.59 (13.67)	22.59 (13.44)	23.45 (14.08)	22.03 (14.80)
Coh.d W_1, W_t	-	0.030	0.067	0.067	0.005	0.103
T-Stat W_1, W_t	-	0.85	1.23	1.34	0.09	1.80

Note: Pairwise t-tests $*p < 0.05$; $**p < 0.01$; $***p < 0.001$.

[†] Bonferroni-adjusted p-values $p_B < 0.05$.

A.5 Additional data on sample

Table A10. *Drop out: Influence of age, gender, college, and SVO angle.*

	<i>Dependent variable:</i>			
	Drop-Out in W_t			
	Estimate	Std. Error	z value	Pr(> z)
Wave 2				
Constant	−0.7530	0.3161	−2.38	0.0172
SVO Angle	−0.0078	0.0058	−1.34	0.1811
Age	−0.0144	0.0076	−1.89	0.0588
Female	0.2857	0.1670	1.71	0.0871
College _{t-1}	−0.3457	0.1667	−2.07	0.0381
Wave 3				
Constant	1.0568	0.2680	3.94	0.0001
SVO angle	0.0040	0.0048	0.84	0.3995
Age	−0.0442	0.0066	−6.73	0.0000
Female	0.2070	0.1353	1.53	0.1258
College _{t-1}	−0.3112	0.1371	−2.27	0.0232
Wave 4				
Constant	0.8032	0.2705	2.97	0.0030
SVO Angle	−0.0018	0.0048	−0.37	0.7084
Age	−0.0376	0.0066	−5.68	0.0000
Female	−0.0111	0.1370	−0.08	0.9352
College _{t-1}	−0.1544	0.1393	−1.11	0.2678
Wave 5				
Constant	1.5880	0.2567	6.19	0.0000
SVO Angle	0.0052	0.0047	1.12	0.2643
Age	−0.0377	0.0059	−6.41	0.0000
Female	0.0248	0.1314	0.19	0.8500
College _{t-1}	−0.2612	0.1350	−1.94	0.0530
Wave 6				
Constant	2.0315	0.2642	7.69	0.0000
SVO Angle	0.0040	0.0048	0.84	0.4017
Age	−0.0442	0.0060	−7.39	0.0000
Female	0.0556	0.1345	0.41	0.6796
College _{t-1}	−0.2139	0.1386	−1.54	0.1228

Note: Influence of sociodemographic variables and SVO angle on probability to drop out.

Table A11. Sample description (including dropouts).

Wave	$\Delta(W_1, W_t)$	N	Participants			Non-participants		
			Age (Mean, SD)	Female %	College	Age	Female	College
1	0	1009	35.81 (11.46)	49.75%	61.05%	.	.	.
2	1 Week	820	36.10 (11.44)	48.90%	63.05%	33.88 (11.03)	53.44%	52.38%
3	1/2 Year	618	37.77 (11.72)	48.71%	64.87%	32.37 (10.02)	51.41%	54.99%
4	1 Year	662	37.20 (11.66)	50.45%	62.99%	32.61 (10.07)	48.41%	57.35%
5	2 Year	451	38.09 (12.22)	49.89%	65.18%	33.59 (10.19)	49.64%	57.71%
6	4 Years	397	38.77 (12.03)	49.87%	65.24%	33.44 (10.38)	49.67%	58.33%

Note: Descriptive variables for each wave. College in % depict the percentage with college education.