




Social Psychological Bulletin

Consistent Kindness: Money Allocation and Kind Act Decisions Are Regulated by a ‘Welfare Trade-Off Ratio’

Oliver Scott Curry^{1,2} , Chloe San Miguel² , James Wilkinson³ ,

Mehmet Necip Tunç⁴ 

[1] School of Anthropology and Museum Ethnography, University of Oxford, Oxford, United Kingdom. [2] kindness.org, New York, NY, USA. [3] Department of Work and Social Psychology, Maastricht University, Maastricht, The Netherlands. [4] Department of Social Psychology, Tilburg University, Tilburg, The Netherlands.

Social Psychological Bulletin, 2026, Vol. 21, Article e14583, <https://doi.org/10.32872/spb.14583>

Received: 2024-05-07 • Accepted: 2025-04-13 • Published (VoR): 2026-04-09



Handling Editor: Gabriela Czarnek, Jagiellonian University, Krakow, Poland

Corresponding Author: Oliver Scott Curry, School of Anthropology and Museum Ethnography, 51/53 Banbury Road, Oxford, OX2 6PE, United Kingdom. E-mail: oliver.curry@anthro.ox.ac.uk

Supplementary Materials: Code, Data, Materials, Preregistration [see [Index of Supplementary Materials](#)]



Abstract

Is kindness regulated by a cost-benefit ratio? Previous research suggests that money allocation decisions are regulated by a ‘welfare trade-off ratio’ (WTR) that reflects the weight attached to the actor’s welfare relative to the recipient’s welfare. Here we replicate this research, and extend it by creating a new measure—*The Kindness Questionnaire*—which asks which real-world acts of kindness, previously rated for cost and benefit, participants would perform for others. In Study 1 ($n = 6,601$) money allocation (MA) and *Kindness Questionnaire* (KQ) decisions for family, friends, colleagues and strangers were highly consistent with an underlying WTR (~92%); more consistent than would be expected by chance; and generally more consistent than with cost or benefit alone. WTRs were high (~0.81); and, for money allocation, declined with social distance. In Study 2 ($n = 8,492$) MA and KQ decisions for neighbors were highly consistent with an underlying WTR (~89%); more consistent than would be expected by chance; and generally more consistent than with cost or benefit alone. WTRs were high (~0.75). In both studies, *The Kindness Questionnaires* showed good convergent, divergent and incremental validity. These studies corroborate ‘welfare tradeoff ratio’ theory, establish proof of principle for a new way of measuring kindness, and provide new tools for measuring kindness to colleagues, strangers and neighbors.



This is an open access article distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), CC BY 4.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Keywords

kindness, welfare trade-off ratio, prosocial behavior, scale development

Highlights

- Previous research suggests that monetary allocation decisions are guided by a psychological cost-benefit variable called a welfare trade-off ratio (WTR).
- This study introduces and tests a new tool, the Kindness Questionnaire (KQ), to measure WTRs for real-world kind acts.
- In two large U.S. samples, decisions about performing kind acts were highly consistent with an underlying WTR.
- These findings suggest that the same psychological logic applies to both monetary and real-world kind decisions, offering a new way to understand and assess kindness.

Kindness is typically understood as *actions* intended to *benefit* others, at some *cost* to the actor—an ‘ABC’ model of kindness (Curry et al., 2018). Whereas it was once a problem to explain why anyone would pay a cost to benefit others, various theories now exist (kin altruism, mutualism, reciprocal altruism, competitive altruism) that explain why we are kind to family, friends, colleagues and strangers (Curry et al., 2018).

Previous research has suggested that people’s decisions about whether to be kind to others do not depend solely on the cost incurred (‘help if cost is below a certain level’), or the benefit provided (‘help if benefit is above a certain level’), but rather on the *ratio* of the cost to benefit (‘help if the ratio of cost to benefit is below a certain level’). This ratio represents the point at which individuals are indifferent between cost to self and benefit to others, and it is used to make decisions about kind acts. For example, a person who is willing to help another up to a cost-benefit ratio of 0.50 would pay a cost of \$5 (or less) to provide a benefit of \$10 (or more), and would also pay a cost of \$15 (or less) to provide a benefit of \$30 (or more). Individuals differ in the cost-benefit ratios they employ: a ‘kinder’ person is willing to incur a greater cost to provide a given level of benefit. For example, a person who is willing to help another up to a cost-benefit of 0.75 would pay a cost of \$7.50 (or less) to provide a benefit of \$10 (or more), and so on. The ratio can be interpreted as the weight an actor attaches to their own welfare relative to the recipient’s welfare, and for that reason has been referred to as a ‘welfare trade-off ratio’ (WTR) (Delton et al., 2023; Delton & Robertson, 2016).

Empirical evidence in support of such WTRs comes mostly from money-allocation tasks, in which it has been found that decisions are highly consistent with an underlying WTR, and that the value of this ratio declines with social distance. For example, in one study with US students ($n = 167$), decisions were 97% consistent with an underlying WTR; and the ratios were 0.62 for friends, and 0.34 for acquaintances (Delton et al., 2023; Delton & Robertson, 2016). Another study with MTurkers ($n = 3,864$) found median

WTRs of 1.07 for mothers and fathers, 0.93 for romantic partners, 0.80 for siblings and friends, 0.53 for acquaintances, and 0.27 for strangers (Forster et al., 2017). A further study with US students found WTRs of 0.26 to strangers in 'low need', and 0.60 to strangers in 'high need' (Sznycer et al., 2019).

Here we replicate this money allocation research with large samples of the US general public. And we extend it by creating and testing a new measure, *The Kindness Questionnaire* (KQ). The KQ asks people whether they would be willing to perform a series of kind acts previously rated for cost and benefit; this enables us to investigate whether decisions about performing real-world acts of kindness are similarly regulated by a welfare trade-off ratio, and to estimate the value of that ratio. We also investigate the convergent, divergent and incremental validity of *The Kindness Questionnaire*.

Study 1

Do people make decisions about allocating money and performing real-world acts of kindness that are consistent with an underlying WTR? What is the value of this WTR, and does it decline with social distance across family, friends, colleagues and strangers? And do *The Kindness Questionnaires* demonstrate good convergent, divergent and incremental validity?

Method

We designed a survey that asked participants to complete standard money allocation tasks (MAs), and *Kindness Questionnaires* (KQs), to family, friends, colleagues and strangers, and to complete the Multidimensional Measure of Prosocial Behavior (MMPB) (Nielson et al., 2017). Note that all money allocation and kind act decisions were hypothetical, and so technically should be referred to as 'hypothetical money allocations' (HMA) and 'hypothetical kind acts' (HKA); but for ease of exposition we refer to them simply as 'money allocation' (MA) and 'kind acts' (KA) throughout. All materials, data and analysis are available on OSF (see Curry et al., 2022).

The MA measures consisted of 15 items, with cost-benefit ratios running from 0.10 to 1.50 (cost-benefit ratios were calculated by dividing the opportunity cost to the participant by the benefit to the recipient. All MA items, and their ratios, are shown in Table S1a). Participants were asked to name a specific family member, friend, and colleague, and to think of a typical stranger. Then, for each of the items, they were asked whether they could prefer an amount for themselves, or an amount for another person. For example: "Choose the option that you prefer: \$11.77 for YOU or \$13.07 for [family member's name]."

We created *The Kindness Questionnaires* (KQs), by choosing a selection of kind acts suitable for family, friends, colleagues and strangers, from a large pool of items previ-

ously rated for perceived cost and benefit (Curry et al., *in press*). In that paper we investigated the relationship between the benefit, cost, cost-benefit ratio and kindness of a large sample of real-world kind acts, drawn from a variety of popular and professional sources. In Study 1 of that paper, participants ($n = 15,997$) rated 1,692 acts to family, friends, colleagues and strangers on the variables of cost, benefit, and kindness on a 1 (“Not at all [beneficial / costly / kind]”) to 9 (“Extremely [beneficial / costly / kind]”) scale. In Study 2 of that paper, participants ($n = 4,801$) rated 385 acts to a generic ‘someone’. Cost-benefit ratios for the kind acts were calculated by dividing average cost rating for an item by its average benefit rating. By deriving ratios from Likert data, we assume that: costs and benefits have a true zero; and the intervals for kind act costs and benefits (used to create the KQ ratios) are equal, just as they are for the monetary costs and benefits (used to create the MA task).

Different items were chosen from this item pool for different recipients. We would have preferred equal numbers of items, but we were constrained in our choice by: a) the number of items from the pool, that were b) suitable for each recipient (for example, ‘make breakfast in bed’ might be suitable for a family member, but not a colleague), and c) the range and distribution of items’ cost-benefit ratios. The KQ for Family had 23 items, with cost-benefit ratios ranging from 0.21–0.69. The KQ for Friends had 21 items, with cost-benefit ratios ranging from 0.21–0.80.¹ The KQ for Colleagues had 22 items, with cost-benefit ratios ranging from 0.23–0.94. The KQ for Strangers had 19 items, with cost-benefit ratios ranging from 0.25–1.07. All items, and cost-benefit ratios, are shown in Table S1b–e. Participants were asked to name a specific family member, friend, and colleague, and to think of a typical stranger. They were then asked whether they would perform each of the kind acts for each recipient. For example: “Given the opportunity, would you: ‘Help carry heavy bags for [family member’s name]’ – yes or no?” (the average cost rating for this item is 1.99, the average benefit is 7.77, hence the cost-benefit ratio is 0.26).

MAs and KQs were scored using the Kirby method (Kirby, 2000). This method was first developed to measure time discounting, but has subsequently been widely used to measure prosocial decision-making (Delton et al., 2023; Delton & Robertson, 2016; Forster et al., 2017; Sznycer et al., 2019). This method calculates the consistency of responses with each possible welfare trade-off ratio, and reports the maximum consistency, and the ratio with which responses were maximally consistent. For example, if a participant was willing to perform all acts below a ratio of 0.50, and no acts above that ratio, they would be 100% consistent with a ratio of 0.50. Alternatively, if a participant was willing to perform most acts below a ratio of 0.50, and few acts above that ratio, they might be only 90% consistent with a ratio of 0.50. In cases where responses are equally

1) The Friend KQ had an additional unrated item included in error (a total of 22 items); this item was excluded from subsequent analysis.

maximally consistent with two or more ratios—for example, 80% consistent with ratios of 0.50 and 0.70—the method reports the geometric mean of those ratios (for example, 80% consistent with a ratio of 0.59). The R-code used to analyze the data using this method is available on OSF (see [Curry et al., 2022](#)).

Participants then completed a general measure of prosociality, the Multidimensional Measure of Prosocial Behaviour ([Nielson et al., 2017](#)). A sample item reads: “If I see someone being given a hard time, I stand up for that person” (not like me at all 1 - very much like me 5). Participants completed the KQs, then the MAs, then the MMPB. The order of recipients in the KQs and MAs were randomized; the order of items in all scales was randomized.

Participants also completed: one-item measures of religiosity, and political identification; standard demographics (age, sex, ethnicity, education, income) and location (US state); and a series of attention checks.

The study was approved by the Committee on the Use of Human Subjects in Research at Harvard University (IRB19-0070). The survey was hosted on Qualtrics.com, and participants were recruited by PureProfile.com at a cost of £2.60 per participant. Our funding enabled us to aim for samples of approximately 100 from each of the 50 US states ($n \approx 5,000$). Data was collected April 8th, 2021 through May 14th, 2021.

Results

A total of 6,601 people completed the survey (age: 52y mean, 18y sd; 56% female, 43% male, 1% other).²

Descriptive statistics for each recipient-specific MA and KQ are shown in [Table 1](#). MAs tended to be bi- or tri-modal, including peaks at the minimum and maximum values. KQs were found to be highly skewed, especially for family and friends where approximately half of the scores were at ceiling. Descriptive statistics (mean, standard deviation) for each MA and KQ item are displayed in Tables S1a-e.

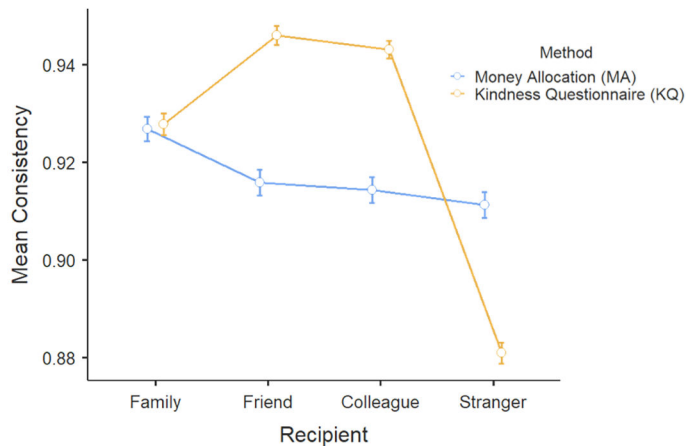
2) 12,232 participants started the study. 1,793 did not complete at least 80% of crucial KQ or MA choices. 3,370 were removed due to speeding through the study ([Greszki et al., 2015](#)). An additional 437 were removed due to not passing attention/quality checks. 31 people were excluded due to reporting ages lower than 18 or greater than 100. 6,601 participants were used in the final analyses.

Table 1*MA and KQ Descriptives (Study 1)*

	Consistency		WTRs		
	Mean	SD	Mean	SD	
MA	Family	93%	10%	1.08	0.44
	Friend	92%	11%	0.94	0.47
	Colleague	91%	11%	0.86	0.49
	Stranger	91%	11%	0.73	0.53
KQ	Family	93%	9%	0.67	0.08
	Friend	95%	8%	0.73	0.13
	Colleague	94%	7%	0.81	0.17
	Stranger	88%	9%	0.66	0.23

Consistencies

MA (91–93%) and KQ (88–95%) choices were highly consistent. A repeated-measures ANOVA showed that the main effect of Method was significant, $F(1, 6600) = 40.39$, $p < .001$, indicating a difference in consistency between MA and KQ across recipient types. The main effect of Recipient was also significant, $F(3, 19800) = 701.14$, $p < .001$, indicating a difference in consistency between recipient groups across the methods. There was a significant interaction between Method and Recipient, $F(3, 19800) = 537.77$, $p < .001$. Post-hoc tests using the Bonferroni correction showed that the consistencies of MA and KQ decisions for family were not significantly different (mean difference = -0.1% , $p = 1.0$), whereas the MA and KQ decisions for friends, colleagues and strangers were significantly ($ps < .001$, $-3\% \leq \text{mean differences} \leq 3\%$, $-0.24 \leq ds \leq 0.23$) different from one another, although the effect size was small (Table S3 in file KQ_S1_data.omv; Figure 1).

Figure 1*MA and KQ Consistencies Compared*

Note. Error bars depict 95% confidence intervals.

One sample t-tests showed that the consistency of all measures was significantly ($ps < 0.001$) and substantially ($ds > 2.09$) greater than would be expected by chance (68%) (Table S2 in file KQ_S1_data.omv).³ Because consistency scores might be inflated by extreme responses (at floor, or at ceiling) we re-ran these analyses after filtering out participants with minimum or maximum WTRs on any measure. Consistencies in this sample ($n = 530$) remained high (MA: 90–91%; KQ: 84–91%), and all were still significantly ($ps < 0.001$) and substantially ($ds > 1.56$) greater than would be expected by chance.

We also investigated whether MA and KQ choices were *more* consistent with a cost-benefit ratio (WTR) than with cost or benefit alone (again, calculated using the Kirby method). A repeated-measures ANOVA showed that the main effect of Method was significant, $F(1, 6600) = 2026.40$, $p < .001$, indicating a difference between MA and KQ across recipient type and measure (cost alone, benefit alone, or cost-benefit ratio). The main effect of Recipient was significant, $F(3, 19800) = 1064.92$, indicating a difference between recipient type across the two methods and three measures. The main effect of Measure was also significant, $F(2, 13200) = 7854.34$, indicating a difference between the cost, benefit, and cost-benefit ratio measures across recipient types and methods.

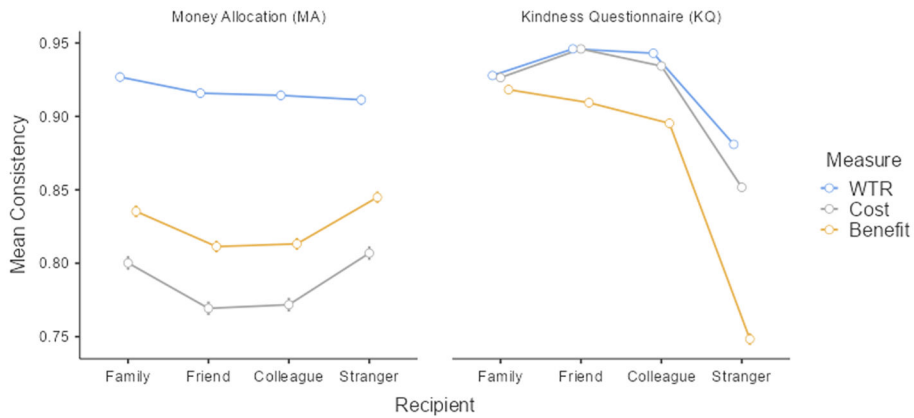
Post-hoc tests using the Bonferroni correction showed that MA choices for all recipients were significantly ($ps < .001$) and substantially more consistent with WTR than with

3) We calculated the level of consistency expected by chance by generating and scoring random responses to the measures, and choosing the highest, most conservative value.

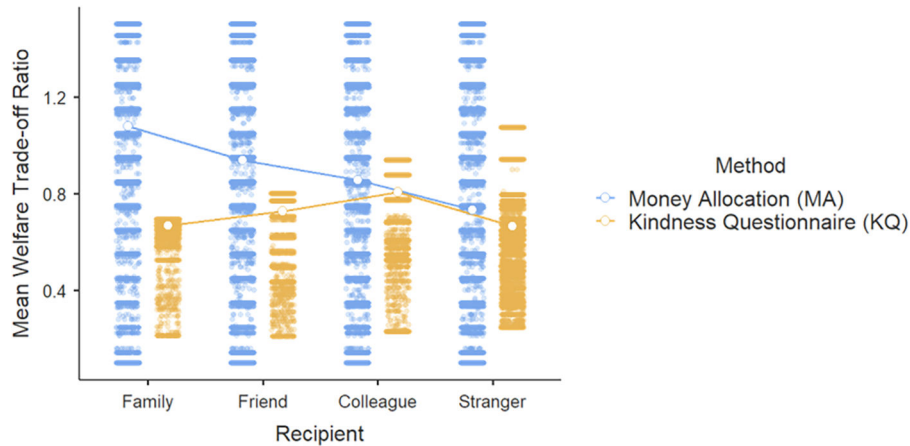
cost ($10\% \leq \text{mean differences} \leq 15\%$, $0.69 \leq ds \leq 0.87$) or benefit ($7\% \leq \text{mean differences} \leq 10\%$, $0.55 \leq ds \leq 0.76$) alone. KQ choices for family, colleagues and strangers were significantly ($ps < .001$) more consistent with WTR than with cost ($0.1\% \leq \text{mean differences} \leq 3\%$, $0.11 \leq ds \leq 0.64$) alone; KQ choices for friends were not significantly more consistent with WTR than with cost (mean difference = 0.004%, $p = 1.0$). KQ choices for all recipients were significantly ($ps < .001$) more consistent with WTR than with benefit ($1\% \leq \text{mean differences} \leq 13\%$, $0.27 \leq ds \leq 1.18$) alone (Figure 2; Table S4 in file KQ_S1_data.omv).

Figure 2

MA and KQ Consistencies Compared (WTR, Cost, Benefit)



Note. Error bars depict 95% confidence intervals.

Figure 3*MA and KQ WTRs Compared (With Observed Values)*

Note. White circles depict mean welfare trade-off ratios, blue and orange marks depict observed values. Error bars depict 95% confidence intervals.

WTRs

WTRs for MA (0.73–1.08) and KQ (0.66–0.81) were high. A repeated-measures ANOVA showed that the main effect for Method was significant, $F(1, 6600) = 1746.29$, indicating a difference between MA and KQ in terms of overall WTR scores. The main effect for Recipient was also significant, $F(3, 19800) = 1034.8$, indicating a difference in WTR between recipient groups. The interaction between Method and Recipient was also significant, $F(3, 19800) = 1700.58$, $p < .001$. Post-hoc tests revealed that the MA WTRs were significantly higher than KQ WTRs, with a mean difference of 0.19, and that MA WTRs declined significantly with greater social distance, in the expected direction (family > friend > colleague > stranger). KQ WTRs did not decline with social distance (family < friend < colleague > stranger), presumably because family and friend scores were at ceiling (Table S5 in file KQ_S1_data.omv; Figure 3).

To test the assumption that kind act ratings had equal (as opposed to logarithmic) intervals, we compared the results of the current untransformed KQs, and those from KQs based on transformed (exponentiated using base 2) cost and benefit data, with the MAs. In all four cases (family, friend, colleague, stranger), the original untransformed KQs were significantly and substantially closer to the corresponding MAs than were the transformed KQs (see Table S9). Exponentiating with a higher base results in even more distant values. This suggests that it is reasonable to assume that the intervals of the kind act cost and benefit ratings used to create the KQs are similar to those of the MAs.

Convergent and Divergent Validity

To test the convergent and divergent validity of the KQ, we investigated whether each recipient-specific KQ WTR predicted the corresponding MA WTR *more than* the MA WTRs to other recipients—for example, whether KQ Family WTR predicts MA Family WTR *more than* MA Friend WTR. Pearson correlations showed that each KQ WTR correlates positively with the corresponding MA WTR, and does so significantly more than with the other MA WTRs (Table 2). Correlation comparisons using ‘cocor’ (<http://comparingcorrelations.org/>) showed that all correlations were significantly different from one another (Steiger’s z s > 7.76 , $ps < .001$; see Table S6 for specific correlation comparisons). A series of multiple regressions also showed that each KQ was a unique, and the best, predictor of the corresponding MA, even when controlling for the other KQs (Tables S7a-d in file KQ_S1_data.omv).

Table 2

Convergent and Divergent Validity of KQ WTRs (Pearson Correlations)

		KQ			
		Family	Friend	Colleague	Stranger
MA	Family	0.26	0.20	0.19	0.19
	Friend	0.16	0.29	0.24	0.27
	Colleague	0.12	0.21	0.36	0.28
	Stranger	0.12	0.20	0.24	0.37

Note. Correlations below the diagonal (for example, MA Friend and KQ Family) involve different measure pairings than correlations above the diagonal (for example, KQ Friend and MA Family), so values differ. All $ps < .001$; largest r in each column and row bolded.

Incremental Validity

To test the incremental predictive validity of the KQ, we investigated the relationship between MAs, KQs and the average Multidimensional Measure of Prosocial Behaviour (MMPB) score ($\alpha = 0.93$).

Pearson correlations showed that each recipient-specific KQ WTR (family: $r = 0.28$; friend: 0.35 ; colleague: 0.37 ; stranger: 0.45 ; $ps < 0.001$) predicted MMPB more than the corresponding MA WTR did (family: $r = 0.24$; friend: 0.30 ; colleague: 0.29 ; stranger: 0.31 ; $ps < .001$). Indeed, correlation comparisons using ‘cocor’ showed that each KQ was significantly more correlated with MMPB than the corresponding MA WTR (Steiger’s z s > 2.79 , $ps < .006$; see Table S9 of the supplementary materials). Furthermore, multiple regressions showed that whereas MAs alone explained 13% of the variance in MMPB (Model 1), the KQs explained an additional 17% (Model 2), and each recipient-specific KQ WTR made a significant, and larger, contribution to MMPB than each corresponding, and all recipient-specific MA WTRs (Table 3).

Table 3*A Multiple Regression to Test the Incremental Validity of KQ Over MA (MMPB)*

		Model 1				Model 2			
		B	SE	<i>p</i>	β	B	SE	<i>p</i>	β
MA	Family	0.11	0.02	< .001	0.07	0.07	0.02	< .001	0.04
	Friend	0.16	0.02	< .001	0.12	0.09	0.02	< .001	0.07
	Colleague	0.11	0.02	< .001	0.08	0.01	0.02	.60	0.01
	Stranger	0.21	0.02	< .001	0.17	0.10	0.02	< .001	0.08
KQ	Family					0.70	0.10	< .001	0.09
	Friend					0.65	0.06	< .001	0.12
	Colleague					0.58	0.05	< .001	0.15
	Stranger					0.79	0.03	< .001	0.28
R2 = 0.13					R2 = 0.30				
ΔR^2					<i>p</i>				
0.18					< .001				

Discussion

People's decisions about how to allocate money, and whether to perform a kind act, are consistent with an underlying cost-benefit or 'welfare tradeoff ratio'. This replicates on a larger scale the results of previous research using money-allocation tasks. And it extends and corroborates previous findings with a new method, *The Kindness Questionnaire*. These results suggest that people employ a welfare trade-off ratio not only when it comes to explicit and precise monetary costs and benefits, but also when it comes to the implicit and imprecise costs and benefits of real-world kind acts.

The study also extends previous research by showing that, as the theory predicts, MA and to a lesser extent KQ choices are more consistent with WTR than cost or benefit alone. This was the case for fifteen of the sixteen predictions that we tested (2 methods \times 4 recipients \times 2 comparisons). In other words, people do not make kind decisions based on cost or benefit alone, but rather on the *ratio* of the cost to benefit.

However, KQ choices were more consistent with cost than MA choices were; and—the sole prediction that was not supported—KQ choices for friends were not significantly more consistent with WTR than with cost. This may be because there is something different about monetary and kind act choices, or something different about kind act choices for friends; or it could be a methodological issue arising from the high correlation between cost and cost-benefit ratio in this sample of kind acts. Further research will be needed to test these possibilities.

The study found that the US public places a much higher value on the welfare of others (MA: \sim 0.90; KQ: \sim 0.72) than previous research with US college students (MA:

~0.48) (Delton et al., 2023), but on a par with research using larger MTurk samples (MA: ~0.78) (Forster et al., 2017), and comparable to previous research looking specifically at WTRs to strangers in need (0.60) (Sznycer et al., 2019).

The study also found that, as expected, for monetary choices, WTR declines with greater social distance. People allocated more money to family, than to friends, colleagues and strangers. This was not the case for kind act choices, presumably because the range of possible scores for KQs, especially to family and friends, was too low, and hence responses were at ceiling. Nevertheless, the decline in KQ and MA WTRs for colleagues and strangers was more or less identical.

Finally, the study found that the KQs showed good convergent and divergent validity with the MA tasks, and good incremental validity with a general measure of prosociality (MMPB). However, given that the KQ measures were at ceiling for family and friends, this version of *The Kindness Questionnaire* should be used to measure kindness to colleagues and strangers only.

In summary, the results of Study 1 show that people make consistent decisions about kindness, the level of kindness is high, and *The Kindness Questionnaires* provide a promising new way to measure kindness.

Study 2

Do people make decisions about allocating money and performing real-world acts of kindness for neighbors that are consistent with a cost-benefit ratio? What is the value of this ratio? And does *The Kindness Questionnaire* demonstrate good convergent, divergent and incremental validity? Study 2 provides a conceptual replication of Study 1, and tests its methodological generalizability (with new measures and targets), with a larger sample.

Method

We designed a survey that asked participants to complete a money allocation task (MA), and *The Kindness Questionnaire* (KQ), to neighbors (participants were not given any instructions about who constitutes a neighbor—so recipients may have ranged from 'next door neighbor' to 'someone in your neighborhood'), and to complete the Prosocialness Scale for Adults (PSA) (Caprara et al., 2005). All materials, data and pre-registered analysis are available on OSF (see Curry et al., 2022).

The MA measure was the same as in Study 1. All items, and cost-benefit ratios, are shown in Table S8a.

As before, we created *The Kindness Questionnaires* for neighbors by selecting suitable kind acts from previous research (Study 2, Curry et al., in press), although this time we made a point of including items with a larger range of cost-benefit ratios than in Study 1.

The KQ for neighbors had 22 items, with cost-benefit ratios ranging from 0.20 – 1.06. All items, and cost-benefit ratios, are shown in Table S8b. Participants were asked how likely it was that they would perform each of the kind acts for each recipient. For example: “Given the opportunity, would you: ‘Leave a note after damaging a neighbor’s car’? (1–4; Very unlikely, Unlikely, Likely, Very likely)”. The average cost rating for this item is 4.48, the average benefit is 7.89, hence and the cost-benefit ratio is 0.57. We changed the response scale from ‘yes/no’ in Study 1, to the 4-point measure here, in order to: a) test and ensure generalizability across measures (after, Yarkoni, 2022), and b) to give us the option of analyzing the data like a normal Likert scale (at a later date). However, for this study, the responses were dichotomized (1 or 2 = 0 [‘No’]; 3 or 4 = 1 [‘Yes’]) so that they could still be scored using the Kirby method, along with the MA responses (Kirby, 2000).

Participants then completed: a 1-item measure of general happiness ‘How happy are you, in general’ (1–5) (Dunn et al., 2008); the Prosocialness Scale for Adults (PSA) (Caprara et al., 2005). For example: “I am available for volunteer activities to help those who are in need” (1–5; never/almost never true, occasionally true, sometimes true, often true, and almost always/always true); a 6-item interdependence (with neighbors) scale (Ayers et al., 2023); a 1-item measure of social identification with one’s neighborhood (Postmes et al., 2013), and a 1-item pictorial measure of sense of inclusion of community in self (Mashek et al., 2007). Participants completed the KQ, then the MA, then the PSA. The order of items in all scales was randomized.

Participants also completed: a range of demographic questions, including age, gender, income, education, marital status, number of children, state of residence, rural/suburban/urban region, duration of residence, political identity, religion and religiosity, ethnicity, and similarity to neighbors (Parker et al., 2018); and a series of attention checks.⁴

The study was approved by the Committee on the Use of Human Subjects in Research at Harvard University (IRB19-0070). The survey was hosted on—and participants were recruited by—SurveyMonkey.com. Participants were compensated \$0.60 for completing the survey. Our funding enabled us to aim for samples of approximately 200 from each of the 50 US states ($n \approx 10,000$). Data was collected in the week of 26 September 2022.

Results

A total of 8,492 people completed the survey (age: median age group 45–54y; 54% female, 45% male, 1% other).⁵

4) Participants also completed: the same KQ scale but *from* neighbors; a range of questions about their emotional reactions to their neighbors (“How does your neighbor make you feel? Angry, Happy, Sad, Regretful, Disappointed, Grateful, Guilty, Proud, Ashamed, Afraid, Envious” (1–5; Not at all, A little, A moderate amount, A lot, A great deal); a single-item measure of satisfaction with one’s neighborhood (Neal, 2021); and two open-ended questions: “What is the kindest thing a neighbor has ever done for you?” and “What one thing could your neighbors do to be kinder to you?”. We plan to report these results in future publications.

The mean consistency for the Money Allocation (MA) measure was 91% ($SD = 11\%$). For the Kindness Questionnaire (KQ) measure, the mean consistency was 87% ($SD = 8\%$). The mean welfare-tradeoff ratio (WTR) for MA was 0.76 ($SD = 0.49$) and the mean WTR for the KQ was 0.74 ($SD = 0.25$). MA WTRs were tri-modal, including peaks at the minimum and maximum values. KQ WTRs were highly skewed, with approximately one third of the scores at ceiling. Descriptive statistics (mean, standard deviation) for each MA and KQ item are displayed in Tables S7a–b.

Consistencies

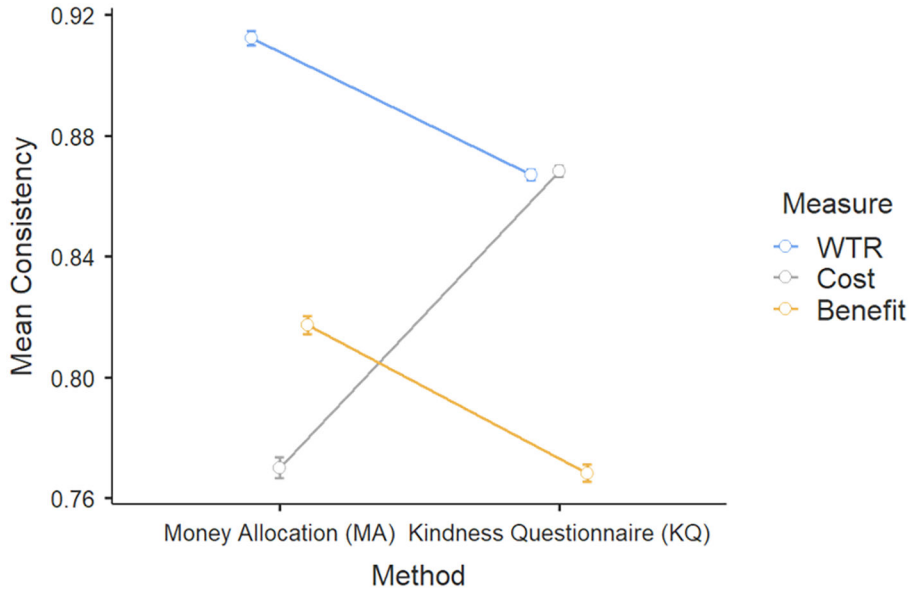
MA (91%) and KQ (87%) choices were highly consistent. A paired sample t -test showed that the MA consistency ($M = 91\%$, $SD = 11\%$) was significantly higher than the KQ consistency ($M = 87\%$, $SD = 11\%$), $t(8491) = -31.54$, $p < .001$; $d = -0.34$) (Table S12 in file KQ_S2_data.omv). One sample t -tests showed that the consistency of both the MA ($t(8491) = 195.29$, $p < .001$, $d = 2.12$) and the KQ ($t(8491) = 217.59$, $p < .001$, $d = 2.36$) were significantly and substantially greater than would be expected by chance (68%) (Table S9 in file KQ_S2_data.omv). Again, because consistency scores might be inflated by extreme WTRs (at floor, or at ceiling) we re-ran these analyses after filtering out participants with minimum or maximum scores on any measure. Consistencies in this sample ($n = 4,047$) remained high (MS: 90%; KQ: 85%), and both were still significantly ($ps < .001$) and substantially ($ds > 2.13$) greater than would be expected by chance.

Again, we investigated whether MA and KQ choices were *more* consistent with a cost-benefit ratio (WTR) than they were with cost or benefit alone. A repeated-measures ANOVA showed that the main effect of Method (MA and KQ) was not significant, $F(1,8491) = 0.62$, $p = .431$. However, the main effect of Measure was significant, $F(2, 16982) = 7401.01$, $p < .001$. And the interaction between Measure and Method was significant, $F(2, 16982) = 5075.53$, $p < .001$. Post-hoc tests using the Bonferroni correction showed that MA choices were significantly ($ps < .001$) and substantially more consistent with WTR than with cost (mean difference = 14%, $d = 1.44$) or benefit (mean difference = 9%, $d = 1.08$) alone. KQ choices were significantly ($ps < .001$) and substantially more consistent with WTR than with benefit (mean difference = 10%, $d = 1.57$) alone; but significantly ($p < .001$) *less* consistent with WTR than with cost (mean difference = -0.1%, $d = -0.12$) alone, although the effect size was small (Figure 4; Table S14).

5) 10,078 people began the study. 27 people did not complete at least 80% of crucial MA and KQ choices. 1,357 people were excluded due to speeding through the study (Greszki et al., 2015). 202 additional people did not pass the attention check, leaving 8,492 participants included in the analyses.

Figure 4

MA and KQ Consistencies Compared (WTR, Cost, Benefit)



Note. Error bars depict 95% confidence intervals.

WTRs

WTRs for MA (0.76) and KQ (0.74) were high. A paired sample *t*-test showed that the MA WTR ($M = 0.76$, $SD = 0.49$) was significantly higher than the KQ WTR ($M = 0.74$, $SD = 0.25$), $t(8491) = -4.76$, $p < .001$; $d = -0.05$) although the effect size was very small (Table S12).

Again, the original untransformed KQ was much closer to the MA than the transformed KQ was (based on exponentiated cost and benefit data; see Table S10).

Convergent and Divergent Validity

To test convergent and divergent validity, we investigated whether the KQ WTR predicted the MA WTR, and whether it did so more than it predicted some other positive personal quality: general happiness. Pearson correlations showed that the KQ WTR correlates positively with the MA WTR ($r = 0.33$; $p < .001$), and does so more than it correlates with general happiness ($r = 0.20$; $p < .001$; Table S13 in file KQ_S2_data.omv). Again, correlation comparisons using ‘cocor’ showed that the correlation between KQ to Neighbors and PSA was significantly higher than the correlation between KQ to Neighbors and general happiness (Steiger’s $z = 9.65$, $p < .001$).

Incremental Validity

To test the incremental predictive validity of the KQ, we investigated the relationship between MA, KQ and the total PSA score ($\alpha = 0.94$).

A multiple regression showed that whereas the MA WTR alone explained 8% of the variance in PSA (Model 1), the KQ WTR explained an additional 14% (Model 2), and the KQ WTR ($\beta = 0.40$) made a significant, and larger, contribution to PSA than the MA WTR ($\beta = 0.15$; Table 4).⁶

Table 4

A Multiple Regression to Test the Incremental Validity of KQ Over MA (PSA)

		Model 1				Model 2			
		B	SE	<i>p</i>	β	B	SE	<i>p</i>	β
MA	Neighbor	0.49	0.02	< .001	0.29	0.26	0.02	< .001	0.15
KQ	Neighbor					1.37	0.03	< .001	0.40
R2 = 0.08					R2 = 0.22				
ΔR^2					p				
0.14					< .001				

Discussion

Again, Study 2 found that people's decisions about how to allocate money to, and whether to perform a kind act for, a neighbor are consistent with an underlying psychological variable that represents a cost-benefit or 'welfare tradeoff ratio'.

The study also replicates and extends the findings of Study 1, by showing that MA and to a lesser extent KQ choices are more consistent with WTR than cost or benefit alone. This was the case for three of the four predictions we tested (2 methods x 1 recipient x 2 comparisons). In other words, for the most part, people do not make kind decisions based on cost or benefit alone, but rather on the *ratio* of the cost to benefit. However, KQ choices were more consistent with cost than MA choices were; and, contrary to our prediction, KQ choices were not significantly more consistent with WTR than with cost. Again, this may be because there is something different about monetary and kind act choices, or it could be a methodological issue arising from the high correlation between cost and cost-benefit ratio in this sample of kind acts. Further research will be needed to test between these possibilities.

⁶ The results were much the same when we regressed interdependence, identity and inclusion onto MA and KQ. See Tables S14a–c.

These findings further corroborate the theory that people employ a welfare trade-off ratio not only when it comes to explicit and precise monetary costs and benefits, but also, to some extent, to the implicit and imprecise intuited costs and benefits of real-world kind acts.

The study also found that people place a high value on the welfare of neighbors (~0.87), comparable to the value placed on the welfare of colleagues in Study 1. And the KQ showed convergent validity with MA, divergent validity with general happiness, and good incremental validity with a general measure of prosociality (PSA).

In summary, the results of Study 2 show that people make consistent decisions about kindness, the level of kindness is high, and the KQ provides a promising new way to measure kindness.

General Discussion

The results of two large studies, using two different measures (MA and KQ), found that people's decisions about whether to be kind to others are largely consistent with an underlying variable that represents the ratio of the costs an individual is willing to pay to benefit others, a welfare trade-off ratio. The level of consistency was high, substantially higher than would be expected by chance. And the results from the MAs and KQs were comparable, suggesting that people apply the same logic to explicit and precise monetary costs and benefits, and the more implicit and imprecise costs and benefits of real-world kind acts.

The *level* of kindness (the WTR) was also high—substantially higher than previous research with US student samples, but comparable to larger samples of MTurk users. The results of Study 1 also show that, for money allocation at least, the level of kindness declines with social distance, which is in keeping with previous research (Forster et al., 2017).

These two studies also demonstrated the convergent, divergent and incremental validity of *The Kindness Questionnaires*. This new measure provides a promising alternative to—and combines the best features of—previous monetary and standard general measures of kindness and related constructs (Baumsteiger & Siegel, 2019; Boxer et al., 2004; Büssing et al., 2013; Canter et al., 2017; Carlo & Randall, 2002; Comunian, 1998; Gherghel et al., 2021; Johnson et al., 1989; Nickell, 1998; Pommier et al., 2020; Rushton et al., 1981; Seligman et al., 2004; Strauss et al., 2016). *The Kindness Questionnaires* may be more accessible than monetary measures, in that they can be used with innumerate populations, such as children. And they may have more predictive validity than standard general measures, in that estimating an individual's WTR on a sample of acts (for example, 0.50) could be used to predict their willingness to perform any and all acts from the larger item pool. *The Kindness Questionnaires* can also be tailored to specific recipients.

However, the studies had a number of limitations. First, all measures were self-report, and involved hypothetical money-allocation and kind acts. Thus these findings might exaggerate the true value of kindness—although it should be noted that previous research suggests hypothetical money-allocation tasks predict real money-allocation (Delton, 2010). Second, *The Kindness Questionnaires* relied on previously normed cost and benefit data, rather than the estimates of the participants themselves. This may reduce the accuracy of the KQ measures—although the similarity of the WTRs obtained from the MAs and KQs suggest that this may not be a substantial problem. Third, ceiling and floor effects may have contributed to anomalous ratings of costs and benefits, limiting the range of available cost-benefit ratios. Fourth, the Study 1 KQs for different recipients used different items, and a different range of cost-benefit ratios, making the results difficult to compare. Fifth, the cost and cost-benefit ratio of the KQ items were highly correlated, possibly masking the degree to which KQ choices were more consistent with WTR than with cost alone. And sixth, as noted above, the Study 1 KQs for family and friends were at ceiling, which limits their utility and interpretability, and suggests that the current versions of *The Kindness Questionnaire* should be used to measure kindness to colleagues, strangers and neighbors only.

Future research should overcome these limitations by: first, using and comparing real-world behavioral measures; second, scoring KQ on participant provided estimates of costs and benefits, and/or using estimates based on a wider range of more objective currencies (perhaps monetary value, to provide even more direct comparison with MA); third, prefacing ratings of the costs and benefits of kind acts with more practice items, in order to better calibrate the scale by anchoring its upper and lower limits; fourth, creating KQs with a uniform set of items that can be used across different types of recipients; fifth, choosing KQ items that vary more on cost; and sixth, rating the costs and benefits of kind acts on a more expansive scale (for example, 1–100), in order to provide more fine-grained response options, and generate items with a wider and higher range of ratios.

Further work along these lines will advance our understanding of the psychology of kindness.

Funding: Thanks to Verizon, Simple Skincare, and Nextdoor for the generous grants that made this research possible.

Acknowledgments: Thanks to Ryan McManus for statistical advice. And thanks to Gabriel Lima and Eva Jahan for research assistance.

Competing Interests: All authors are, or were at one time, employed by kindness.org. Compensation was not dependent on the research findings. No individuals or organizations outside the research team were involved in the study design, data collection, analysis, or interpretation of the results.

Author Contributions: *Oliver Scott Curry*—Conceptualization | Methodology | Formal analysis | Writing | Supervision. *Chloe San Miguel*—Methodology | Software | Formal analysis | Data curation. *James Wilkinson*—Conceptualization | Investigation. *Mehmet Necip Tunç*—Conceptualization | Investigation | Formal analysis.

Ethics Statement: This research was conducted in accordance with relevant ethical guidelines and regulations. Study 1 was reviewed and deemed exempt under 45 CFR 46.104(d)(3) by the Harvard University–Area Committee on the Use of Human Subjects (IRB Protocol #IRB19-0070). Study 2 was reviewed and determined to be exempt under 45 CFR 46.104(d)(2)(i) & (ii) by Ethical & Independent Review Services (Study ID: 22157). For both studies, participants were presented with an online consent form detailing the study’s purpose, procedures, risks, and their rights. Informed consent was provided by choosing to proceed with the survey. No identifying information was collected, and all data were fully anonymous.

Preregistration Statement: Only Study 2 was preregistered (see [Curry et al., 2022](#)); deviations from the preregistered protocol are documented in the supplementary materials.

Data Availability: For this article, data are available (see [Curry et al., 2022](#)).

Supplementary Materials

For this article, the following Supplementary Materials are available:

- Materials, data, code, and analysis (see [Curry et al., 2022](#))

Index of Supplementary Materials

Curry, O. S., San Miguel, C., Wilkinson, J., & Tunç, M. N. (2022). *Consistent Kindness: Money allocation and kind act decisions are regulated by a ‘welfare trade-off ratio’* [Materials, data, code, analysis, and preregistration of Study 2]. OSF. <https://osf.io/zuh59>

References

- Ayers, J. D., Sznycer, D., Sullivan, D., Guevara Beltrán, D., van den Akker, O. R., Muñoz, A. E., Hruschka, D. J., Cronk, L., & Aktipis, A. (2023). Fitness interdependence as indexed by shared fate: Factor structure and validity of a new measure. *Evolutionary Behavioral Sciences*, 17(3), 259–284. <https://doi.org/10.1037/ebs0000300>
- Baumsteiger, R., & Siegel, J. T. (2019). Measuring prosociality: The development of a Prosocial Behavioral Intentions Scale. *Journal of Personality Assessment*, 101(3), 305–314. <https://doi.org/10.1080/00223891.2017.1411918>
- Boxer, P., Tisak, M. S., & Goldstein, S. E. (2004). Is it bad to be good? An exploration of aggressive and prosocial behavior subtypes in adolescence. *Journal of Youth and Adolescence*, 33(2), 91–100. <https://doi.org/10.1023/B:JOYO.0000013421.02015.ef>
- Büssing, A., Kerkisiek, P., Günther, A., & Baumann, K. (2013). Altruism in adolescents and young adults: Validation of an instrument to measure generative altruism with structural equation modeling. *International Journal of Children's Spirituality*, 18(4), 335–350. <https://doi.org/10.1080/1364436X.2013.849661>
- Canter, D., Youngs, D., & Yaneva, M. (2017). Towards a measure of kindness: An exploration of a neglected interpersonal trait. *Personality and Individual Differences*, 106, 15–20. <https://doi.org/10.1016/j.paid.2016.10.019>
- Caprara, G. V., Steca, P., Zelli, A., & Capanna, C. (2005). A new scale for measuring adults' prosocialness. *European Journal of Psychological Assessment*, 21(2), 77–89. <https://doi.org/10.1027/1015-5759.21.2.77>
- Carlo, G., & Randall, B. A. (2002). The development of a measure of prosocial behaviors for late adolescents. *Journal of Youth and Adolescence*, 31(1), 31–44. <https://doi.org/10.1023/A:1014033032440>
- Comunian, A. L. (1998). The Kindness Scale. *Psychological Reports*, 83(3, suppl), 1351–1361. <https://doi.org/10.2466/pr0.1998.83.3f.1351>
- Curry, O. S., Rowland, L. A., Van Lissa, C. J., Zlotowitz, S., McAlaney, J., & Whitehouse, H. (2018). Happy to help? A systematic review and meta-analysis of the effects of performing acts of kindness on the well-being of the actor. *Journal of Experimental Social Psychology*, 76, 320–329. <https://doi.org/10.1016/j.jesp.2018.02.014>
- Curry, O. S., San Miguel, C., Wilkinson, J., & Tunç, M. N. (in press). The costs and benefits of kindness. *Journal of Positive Psychology*. <https://osf.io/gvfdw/>
- Delton, A. W. (2010). *A psychological calculus for welfare tradeoffs*.
- Delton, A. W., Jaeggi, A. V., Lim, J., Sznycer, D., Gurven, M., Robertson, T. E., Sugiyama, L. S., Cosmides, L., & Tooby, J. (2023). Cognitive foundations for helping and harming others: Making welfare tradeoffs in industrialized and small-scale societies. *Evolution and Human Behavior*, 44(5), 485–501. <https://doi.org/10.1016/j.evolhumbehav.2023.01.013>
- Delton, A. W., & Robertson, T. E. (2016). How the mind makes welfare tradeoffs: Evolution, computation, and emotion. *Current Opinion in Psychology*, 7, 12–16. <https://doi.org/10.1016/j.copsyc.2015.06.006>

- Dunn, E. W., Aknin, L. B., & Norton, M. I. (2008). Spending money on others promotes happiness. *Science*, 319(5870), 1687–1688. <https://doi.org/10.1126/science.1150952>
- Forster, D. E., Pedersen, E. J., Smith, A., McCullough, M. E., & Lieberman, D. (2017). Benefit valuation predicts gratitude. *Evolution and Human Behavior*, 38(1), 18–26. <https://doi.org/10.1016/j.evolhumbehav.2016.06.003>
- Gherghel, C., Nastas, D., Hashimoto, T., & Takai, J. (2021). The relationship between frequency of performing acts of kindness and subjective well-being: A mediation model in three cultures. *Current Psychology*, 40(9), 4446–4459. <https://doi.org/10.1007/s12144-019-00391-x>
- Greszki, R., Meyer, M., & Schoen, H. (2015). Exploring the effects of removing “too fast” responses and respondents from web surveys. *Public Opinion Quarterly*, 79(2), 471–503. <https://doi.org/10.1093/poq/nfu058>
- Johnson, R. C., Danko, G. P., Darvill, T. J., Bochner, S., Bowers, J. K., Huang, Y.-H., Park, J. Y., Pecjak, V., Rahim, A. R. A., & Pennington, D. (1989). Cross-cultural assessment of altruism and its correlates. *Personality and Individual Differences*, 10(8), 855–868. [https://doi.org/10.1016/0191-8869\(89\)90021-4](https://doi.org/10.1016/0191-8869(89)90021-4)
- Kirby, K. (2000). *Instructions for inferring discount rates from choices between immediate and delayed rewards* [Unpublished manuscript]. Williams College.
- Mashek, D., Cannaday, L. W., & Tangney, J. P. (2007). Inclusion of community in self scale: A single-item pictorial measure of community connectedness. *Journal of Community Psychology*, 35(2), 257–275. <https://doi.org/10.1002/jcop.20146>
- Neal, Z. (2021). Does the neighbourhood matter for neighbourhood satisfaction? A meta-analysis. *Urban Studies*, 58(9), 1775–1791. <https://doi.org/10.1177/0042098020926091>
- Nickell, G. S. (1998). The helping attitude scale. *106th Annual Convention of the American Psychological Association at San Francisco*, 1–10. <https://scholar.google.com/scholar?cluster=3907632416234758741&hl=en&oi=scholar>
- Nielson, M. G., Padilla-Walker, L., & Holmes, E. K. (2017). How do men and women help? Validation of a multidimensional measure of prosocial behavior. *Journal of Adolescence*, 56, 91–106. <https://doi.org/10.1016/j.adolescence.2017.02.006>
- Parker, K., Horowitz, J. M., Brown, A., Fry, R., Cohn, D., & Igielnik, R. (2018, May 22). *What unites and divides urban, suburban and rural communities*. Pew Research Center’s Social & Demographic Trends Project. <https://www.pewresearch.org/social-trends/2018/05/22/what-unites-and-divides-urban-suburban-and-rural-communities/>
- Pommier, E., Neff, K. D., & Tóth-Király, I. (2020). The development and validation of the compassion scale. *Assessment*, 27(1), 21–39. <https://doi.org/10.1177/1073191119874108>
- Postmes, T., Haslam, S. A., & Jans, L. (2013). A single-item measure of social identification: Reliability, validity, and utility. *The British Journal of Social Psychology*, 52(4), 597–617. <https://doi.org/10.1111/bjso.12006>

- Rushton, J. P., Chrisjohn, R. D., & Cynthia Fekken, G. (1981). The altruistic personality and the self-report altruism scale. *Personality and Individual Differences*, 2(4), 293–302.
[https://doi.org/10.1016/0191-8869\(81\)90084-2](https://doi.org/10.1016/0191-8869(81)90084-2)
- Seligman, M. E. P., Park, N., & Peterson, C. (2004). The Values In Action (VIA) classification of character strengths. *Ricerche di Psicologia*, 27, 63–78.
- Strauss, C., Lever Taylor, B., Gu, J., Kuyken, W., Baer, R., Jones, F., & Cavanagh, K. (2016). What is compassion and how can we measure it? A review of definitions and measures. *Clinical Psychology Review*, 47, 15–27. <https://doi.org/10.1016/j.cpr.2016.05.004>
- Sznycer, D., Delton, A. W., Robertson, T. E., Cosmides, L., & Tooby, J. (2019). The ecological rationality of helping others: Potential helpers integrate cues of recipients' need and willingness to sacrifice. *Evolution and Human Behavior*, 40(1), 34–45.
<https://doi.org/10.1016/j.evolhumbehav.2018.07.005>
- Yarkoni, T. (2022). The generalizability crisis. *The Behavioral and Brain Sciences*, 45, Article e1.
<https://doi.org/10.1017/S0140525X20001685>

PSPS Polish
Social
Psychological
Society

Social Psychological Bulletin (SPB) is an official journal of the Polish Social Psychological Society (PSPS).

zpid ^{LP}

Leibniz-Institut für
Psychologie

PsychOpen GOLD is a publishing service provided by the Leibniz Institute for Psychology (ZPID), Germany.