CENTRE FOR DECISION RESEARCH & EXPERIMENTAL ECONOMICS

University of Nottingham
UK | CHINA | MALAYSIA

Discussion Paper No. 2023-10

Malte Baader, Chris Starmer,
Fabio Tufano and
Simon Gächter

October 2023

‘IOS$_{11}$’: A new, extended, interactive version of the ‘Inclusion of Other in the Self’ scale

CeDEx Discussion Paper Series
ISSN 1749 - 3293

CENTRE FOR DECISION RESEARCH & EXPERIMENTAL ECONOMICS

The Centre for Decision Research and Experimental Economics was founded in 2000, and is based in the School of Economics at the University of Nottingham.

The focus for the Centre is research into individual and strategic decision-making using a combination of theoretical and experimental methods. On the theory side, members of the Centre investigate individual choice under uncertainty, cooperative and non-cooperative game theory, as well as theories of psychology, bounded rationality and evolutionary game theory. Members of the Centre have applied experimental methods in the fields of public economics, individual choice under risk and uncertainty, strategic interaction, and the performance of auctions, markets and other economic institutions. Much of the Centre's research involves collaborative projects with researchers from other departments in the UK and overseas.

Please visit http://www.nottingham.ac.uk/cedex for more information about the Centre or contact

Samantha Stapleford-Allen
Centre for Decision Research and Experimental Economics
School of Economics
University of Nottingham
University Park
Nottingham
NG7 2RD
Tel: +44 (0)115 74 86214
Samantha.Stapleford-Allen@nottingham.ac.uk

The full list of CeDEx Discussion Papers is available at

http://www.nottingham.ac.uk/cedex/publications/discussion-papers/index.aspx

‘IOS$_{11}$’: A new, extended, interactive version of the ‘Inclusion of Other in the Self’ scale

Malte Baader[1], Chris Starmer[2], Fabio Tufano[3] & Simon Gächter[2]

[1]Department of Banking and Finance, University of Zurich

[2]School of Economics, University of Nottingham

[3]School of Business, University of Leicester

13 October 2023

Author Note

We thank Gary Charness, Anna Hochleitner and Martin Sefton for very helpful comments. We thank also colleagues from CeDEx (University of Nottingham) and participants at the 2023 ESA European Meeting for their very valuable input.

Correspondence concerning this article should be addressed to Malte Baader (University of Zurich, Plattenstrasse 14, 8032 Zurich, Switzerland, Email: malte.baader@bf.uzh.ch) or Fabio Tufano (University of Leicester, Brookfield, 266 London Road, Leicester LE2 1RQ, United Kingdom, +44 (0)116 252 2901, Email: fabio.tufano@leicester.ac.uk).

Abstract

We introduce and test a new tool designed to measure "relationship closeness". Studying relationship closeness has a long history in psychology and is currently expanding in other fields including economics. Our new measurement tool is a refinement of the widely used 'Inclusion of Other in the Self' scale (IOS for short) of Aron et al. (1992) and is designed to embody three key features. First, it retains attractive attributes of the standard IOS tool including being an effective measurement technique which is easy to implement and understand. Second, we enhance the scope for convenient use of the tool via the development of a portable interactive interface that will be particularly useful in online studies. Thirdly and crucially, through extensive pre-registered experimental testing, we demonstrate that our enhanced tool – IOS$_{11}$ which features an 11-point response scale – outperforms previous versions of IOS in better proxying features of relationships captured by a range of more complex survey tools; the performance of IOS$_{11}$ is also indistinguishable from that of the more complex 'Oneness' measure of Cialdini et al. (1997) which uses the standard IOS as one of its two-item inputs.

Keywords: Inclusion of other in the self scale; oneness; relationship closeness; psychometric evaluation; replication

Introduction

In this paper, we present and test a new tool designed to measure "relationship closeness". Broadly speaking, relationship closeness refers to the subjectively perceived *quality* of a relationship between two agents. As such, it is a natural object of study in the human and social sciences and there are extensive literatures studying the determinants of relationship closeness and investigating its impact on wide-ranging dimensions of human well-being including, for example, health, the incidence and resolution of conflict and economic productivity. Based on existing research, the study of relationship closeness can offer important insights into the human condition and contribute to public understanding of pressing contemporary issues such as how to build healthier, more resilient, productive, and inclusive societies (see Mashek & Aron, 2004; Reis et al., 2000 for reviews). Our current contribution lies in providing a new and improved technique for measuring relationship closeness that is low cost to implement and well-suited to a wide range of applications.

Influential work in psychology dating back several decades has developed a range of techniques for quantifying relationship closeness. Prominent examples include: the *Relationship Closeness Inventory* (RCI), the *Subjective Closeness Index (SCI)* (both by Berscheid et al., 1989), the *Love and Liking scale* (LLS) (Rubin, 1970) as well as the *Personal Acquaintance Measure (PAM)* (Starzyk et al., 2006).[1] While these methods focus on different types or aspects of relationships and differ in their conceptual foundations, they share the common feature that their implementation requires responses to, sometimes quite extensive, multi-item questionnaires.

Our primary concern is with an offshoot from this literature, which has sought to develop more compact tools for measuring relationship closeness which are both effective in

[1] Other related measurement approaches are the degree of acquaintanceship (Paunonen, 1989), and measures such as time spent together (Gager & Sanchez, 2003; Guldner & Swensen, 1995) or relationship levels (Laursen et al., 2000; Wei & Lo, 2006).

capturing relevant dimensions of relationships while being quick and easy to implement. Two prominent tools developed in this spirit are the *Inclusion of Other in the Self* (*IOS*) scale and the *Oneness* scale which are both described in detail in Section 2. Both techniques are well-known and the two key papers that introduced and popularized them – Aron et al. (1992) and Cialdini et al. (1997), respectively – had, at the time of writing, accumulated almost 9000 citations between them.[2] Both tools are quick and easy to implement and both have been shown to do a remarkably good job of proxying relationship closeness across a wide range of relationship classes, from acquaintances to close friends (Gächter et al., 2015).[3] The two tools have been widely used across the social and behavioral sciences especially in the disciplines of psychology and sociology (see e.g. Kastendieck et al., 2022; Markowitz & Slovic, 2020; Tropp & Wright, 2001; Waugh & Fredrickson, 2006) and in various applied fields such as health (e.g. Bartz et al., 2010; Calbi et al., 2021); there is also growing interest in new areas of application (e.g. research in economics including Gächter et al. (2022); Gächter et al. (2023); Goette and Tripodi (2021) or computer science, see Westlund et al. (2018); Miller et al. (2019)) where, until fairly recently, these tools had barely been used at all.

To date, however, researchers considering using one of these tools have faced a tradeoff. Specifically, the IOS scale is more "convenient" to implement (it requires measurement of just one scale instead of two) but comparative testing has shown Oneness to be the more "predictive" tool in that it correlates more strongly with other, more complex, measures of relationship closeness (Gächter et al., 2015). Since the publication of Gächter et al. (2015), several studies have relied on their evidence to motivate use of IOS as a good predictor of relationship closeness despite the fact that it is not the best available tool in this

[2] At the time of writing, Aron et al. (1992) have 6469 Google Scholar citations while Cialdini et al. (1997) have 2191. Interestingly, only a minority of papers cite both articles.

[3] Gächter et al. (2015), building on and extending previous work including Aron et al (1992) and Starzyk et al (2006), show that IOS and Oneness correlate strongly with the principal component of several other measures of relationship closeness.

respect (see, for example, Bicchieri et al., 2022; Castillo, 2021; Dimant, 2023; Molleman & Gächter, 2018; Parisi et al., 2021; Pellencin et al., 2018; Robson, 2021; Tarr et al., 2016 among others). While we do not dispute that sacrificing accuracy for simplicity or convenience may have been a perfectly defensible trade-off, as we demonstrate below, it is no longer necessary.

In this paper, we propose a novel measurement instrument which builds closely on the original IOS scale. A key feature is that we extend the tools' answer range (from a 7-point) to an 11-point scale. Based on this feature, we refer to our tool as "IOS$_{11}$". The primary motivation for extending the response scale in this way is that it provides a more nuanced measurement tool, with its degree of granularity more comparable to that of the two-item Oneness scale.[4] We do not presume that finer granularity is the only plausible explanation of the differential performance between IOS and Oneness, however. Other contending possibilities, for example, are that the two items of the Oneness measure pick up somewhat distinct aspects of relationship closeness or that two-item measurement is inherently less noisy (see Neuberg et al. (1997) and Gillen et al. (2019) for further discussion on these issues). While our data allow us to shed some light on what factors may be at play, our primary objective was to test the conjecture that finer granularity might reduce the gap between the predictive performance of Oneness and our refined IOS tool.

Minded by the important growth of, often very large-scale, data collection in online environments (Mason & Suri, 2012; Snowberg & Yariv, 2021), a second innovative feature of IOS$_{11}$ is that we implement it via an interactive, computerized, interface. The result is a simple and intuitive task suited to a range of computerized environments from lab to online participant pools such as Amazon MTurk or Prolific.

[4] To see why, consider a subject who responds with scores of, say, 3 and 4 on the two items measured for the Oneness scale. This subject receives a Oneness score of 3.5, a value not measurable on the 7-point scale of the standard IOS. To the extent that the advantage of the Oneness measure derives from this finer implied scale, IOS$_{11}$ should substantially close that gap.

Following Gächter et al. (2015), we test the performance of IOS$_{11}$ by examining its correlation with a set of other well-established but more elaborate measures of relationship closeness (RCI, SCI, LLS, and PAM) and we benchmark the performance of our new tool against Oneness and the original IOS tool. We also contribute to the literature by adding a pre-registered replication of Gächter et al. (2015, Study 3) alongside our validation of IOS$_{11}$. We find that IOS$_{11}$ elicits relationship closeness more accurately than IOS and just as well as the more complex Oneness measure. As such, we conclude that our tool with its combination of high accuracy and cost-effectiveness is now the most attractive available approach for fast, convenient, and effective measurement of relationship closeness.

The remainder of the paper is organized as follows. In Section 2, we explain the details of IOS$_{11}$. In Section 3, we outline how we validate it relative to the original IOS and Oneness. In Section 4, we present results with concluding remarks in Section 5.

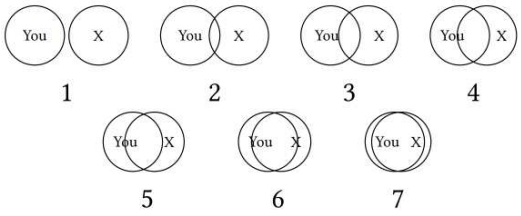
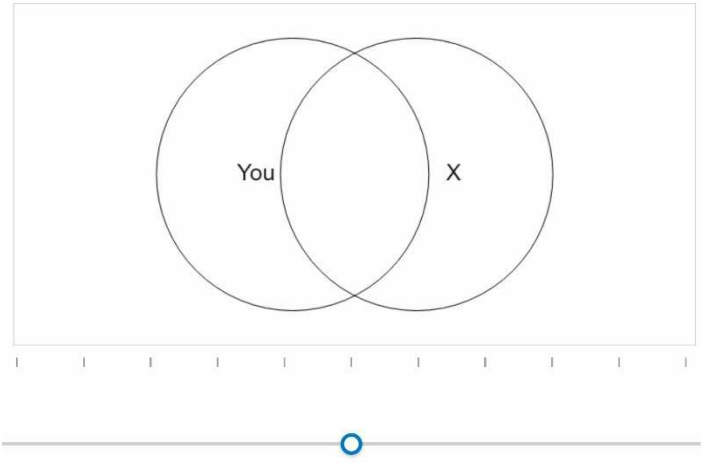
The IOS$_{11}$ scale

The top left panel of Fig. 1 presents the original IOS task, adapted from Aron et al. (1992). A respondent is required to say which of the seven pairs of circles best represents their relationship with another identified individual. As noted in the introduction, responses to this simple task correlate very highly (Spearman's $\rho \in [0.514, 0.820]$, p-values < 0.001) with scales based on considerably more complex measurement approaches. However, the two-item Oneness measure which takes the average of responses on two items – the IOS score and the We scale score (Brewer & Gardner, 1996) (top right panel) – has been shown to outperform the basic IOS measure in its correlation with other measures of relationship closeness.[5]

[5] Using a regression framework Gächter et al. (2015) find that an *Index of Relationship Closeness* is more strongly related to Oneness ($\hat{\beta}$ = 0.929***) than IOS alone ($\hat{\beta}$ = 0.870***).

Figure 1

Graphical comparison of the interfaces of the IOS and our IOS$_{11}$ scale

IOS	**We**
"In the following figure we ask you to consider which of these pairs of circles best represents your relationship with X. By selecting the appropriate number, please indicate to what extent you and X are connected."	*"Please, select the appropriate number below to indicate to what extent you would use the term "WE" to characterize you and X."*
	Not at all Very much so 1 2 3 4 5 6 7
Oneness = $(IOS + We)/2$	
IOS$_{11}$	
"Once you move the slider below, a pair of circles will appear in the box. The position of the slider will determine the extent to which the circles overlap. When the slider is all the way to the left, the circles will look like this  *. When the slider is near the middle, the circles look like this*  *. With it all the way to the right the circles look like this*  *. You should interpret the degree of overlap as representing the relationship between you and X.* *Please position the slider so that the circles indicate to what extent you and X are connected."*	
	

Note. The initial screen participants see when entering the elicitation is blank. For illustration purposes we are depicting the slider at a central position in this figure.

In developing IOS$_{11}$, and for reasons already explained, we conjectured that extending the 7-point response scale of the original IOS tool might enhance its predictive accuracy.

Extending the number of pairs of circles from which participants can choose, however, creates two obvious challenges. The first is how to visualize an increased number of overlapping circles without their presentation becoming too cluttered, complicated, or confusing. Secondly, we needed to decide by how many options the answer range should be extended.

We addressed the first of these challenges by developing our new tool as a computerized version of IOS using an interactive screen that allows participants to intuitively adjust the degree to which circles overlap. Our layout is displayed in the bottom panel of Fig. 1. Participants move a slider below the circle diagram to adjust the degree to which the circles overlap. These changes to the original scale do not affect the portability, ease of explanation or the time it takes to complete the scale. The resulting tool also has the obvious attraction that IOS$_{11}$ can be implemented in a wide range of computerized environments supporting easy use in online surveys and online or lab experiments.[6]

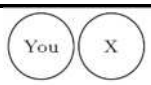
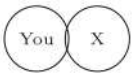
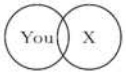
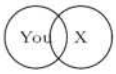
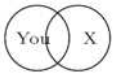
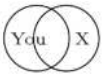
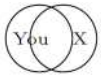
This leads us to the second consideration of how many degrees of overlap to offer. The move to a computerized environment would, in principle, allow the implementation of a very fine-grained (quasi-continuous) scale. However, Simms et al. (2019) have suggested that using a continuous or 'visual-analogue' scale, can be a source of noise if respondents "*[are] unable to reliably make meaningful and valid fine-grained distinctions"* (Simms et al., 2019, p. 9).[7]

Moved by this consideration, we stick with a discrete version of the task. To enhance comparability to previous studies we kept the maximum and minimum overlap of circles identical to IOS by Aron et al. (1992).

[6] Interested researchers can access the interface and code using the following link: https:/doi.org/10.17605/OSF.IO/9DBR6.
[7] Other researchers (Beranek & Castillo, 2022; Kamphorst et al., 2017; Le et al., 2007) have proposed using continuous scales.

Table 1

Comparison of IOS and IOS$_{11}$

	IOS	IOS$_{11}$	IOS$_{11}$ Recoded
	1	1	1
		2	1.5
	2	3	2
		4	2.5
	3	5	3
	4	6	4
	5	7	5
		8	5.5
	6	9	6
		10	6.5
	7	11	7

Note. 'X' serves as a placeholder for the initial of the person considered. The original scale does not reduce the distance between circles linearly. Thus, we extend our scale in the range [1,3] and [5,7] to yield an almost linear change in overlap. 'IOS$_{11}$ recoded' is a recoding of 'IOS$_{11}$' that retains the 1-7 scale.

We then chose the number of levels such that the change in distance between the centers of the circles is approximately linear and so that the original IOS levels form a subset

of the extended version.[8] This leads to a setup with the 11 relationship closeness levels as shown in the middle column of Table 1. The left-hand column of Table 1 shows how scores on the original IOS map into a subset of scores on the new tool. Additionally, the rightmost column of Table 1 shows how IOS_{11} can be recoded to a 7-point scale with endpoints matching the original IOS scale for comparability.

Method

We test the performance of IOS_{11} by examining how well it correlates with a range of other measures of relationship closeness and we benchmark its performance against the original IOS and Oneness scales. We employ a between-participant design, where each either performs the two tasks necessary to measure Oneness (i.e., IOS and We scales) or completes our new IOS_{11} measure. We then explore the within-participant correlation of each of IOS, Oneness and IOS_{11} to a series of well-established survey instruments designed to capture relationship closeness. Following Gächter et al. (2015), the different scales that we use are the *Relationship Closeness Inventory* (RCI), the *Subjective Closeness Index (SCI)* (both by Berscheid et al., 1989), the *Love* and *Liking scales* (LLS) (Rubin, 1970) as well as the *Personal Acquaintance Measure (PAM)* (Starzyk et al., 2006).[9]

Note that some of these measures were constructed to capture different specific degrees of relationships (e.g., the RCI explicitly refers to romantic relationships, whereas the PAM was designed for acquaintances). However, from a behavioral scientist's perspective, it would be useful to have a general-purpose and portable measurement tool that could be applicable to a range of relationships. For that reason and, again, following Gächter et al. (2015), we employed a between-subject variation where participants were asked to either consider a very

[8] We could not achieve perfect linearity together with incorporating the original IOS overlaps. See online Appendix A.2 for more details.

[9] The scales we use as benchmarks have received substantial attention in the psychological literature over the years, amounting to 3,778 citations on Google Scholar at the point of writing this paper. Using them also embeds replications of Gächter et al. (2015) as well as Starzyk et al. (2006), providing more evidence for the robustness of their findings.

close person; a friend; or an acquaintance across all of the core questions within the study.[10] Hence, our main experiment can be considered a two-by-three treatment design varying Oneness and IOS$_{11}$ on the one hand and the type of relationship considered on the other.[11]

We presented the instruments eliciting relationship closeness in random order, followed by questions regarding demographics and other individual attributes. The full instructions and details of the various measures of relationship closeness employed as benchmarks can be found in the online Appendix B.

Procedures

We pre-registered our study and collected data online in July 2021 using Qualtrics (Qualtrics, 2021).[12] The study was approved by the Nottingham School of Economics' Research Ethics Committee and 751 individuals participated with $N \approx 125$ per treatment using Prolific's UK sample.[13] All participants completed an informed consent form at the start of the study. The mean age of our participants is 35.22 years (SD=13.86, Min=17, Max=75) with 501 (67%) identifying as female, 242 (32%) identifying as male and 10 participants not revealing their gender. The sample includes 29% students and 56% of the participants are either in full- or part-time employment. Using an online participant pool such as Prolific therefore provided us with a more heterogeneous demographic than utilizing a student sample. We also obtained additional survey data of other demographics directly from Prolific including age, gender, education levels and details about the participant's household. We paid a flat fee of £1.20 per participant and the study took about 15 minutes to complete.

[10] To ensure salience of the considered person throughout the study, we ask participants in the beginning of the experiment to provide the initials of the person they are thinking of. These initials are then inserted in all parts where the instructions explicitly refer to another person. For example, in the IOS and 'We' scale we substitute the 'X' with the provided initial.
[11] In addition, we asked each participant to rate a stranger via either Oneness or IOS$_{11}$, using whichever tool they encountered in their treatment arm. This was included to examine individual-level variation in interpretation of the scale. While this produced limited evidence of any consistent demographic determinants, for completeness, this analysis is presented in online Appendix A.1.
[12] Registration number AEARCTR-0007947. See https://www.socialscienceregistry.org/trials/7947
[13] Peer et al. (2022) provide evidence on high data quality using Prolific. See www.prolific.co for more information on their services. The exact number of participants in each treatment can be found in Fig. 2.

Results

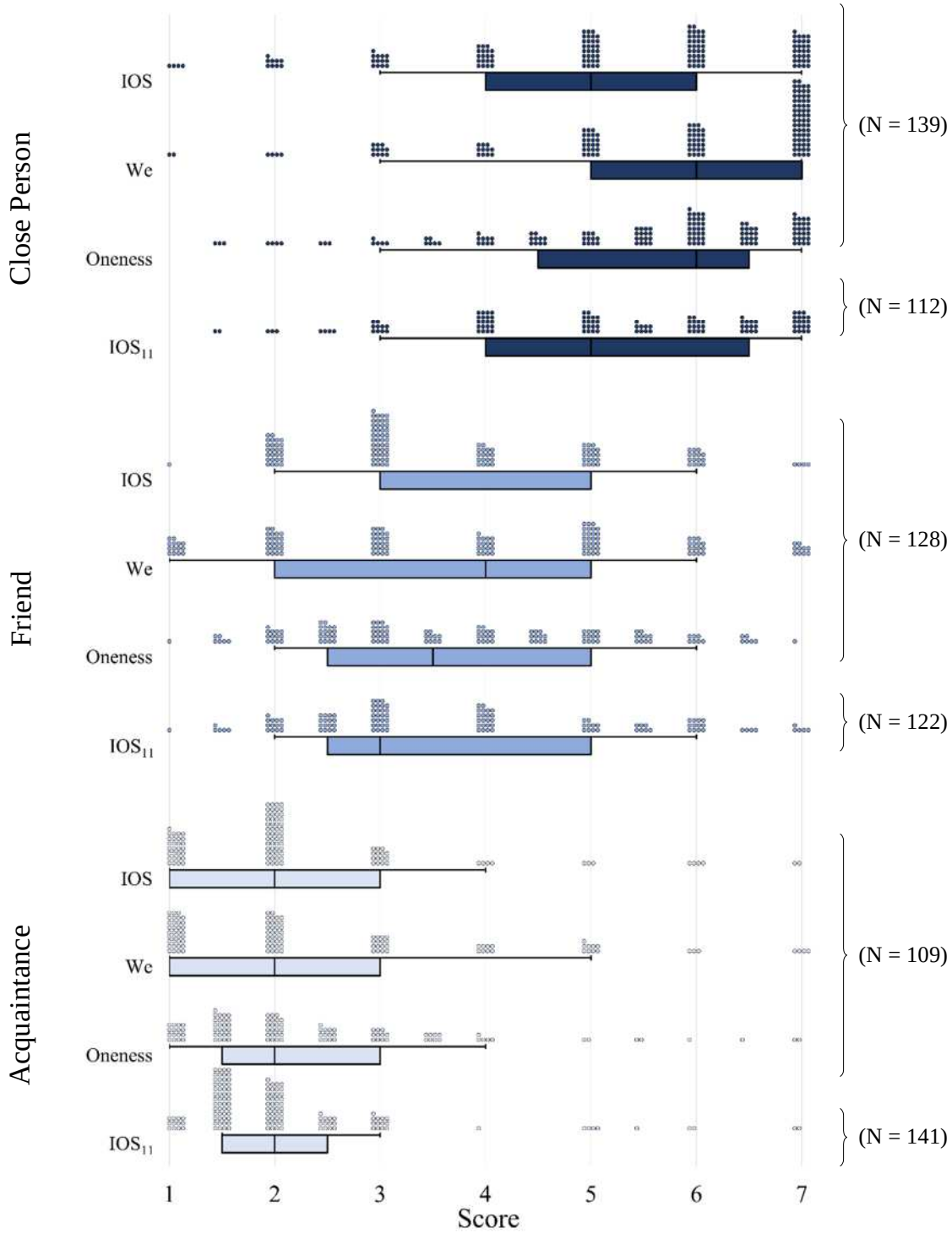
As a first descriptive benchmarking of IOS_{11} against IOS and Oneness we examine the reported relationship closeness scores across different treatments. All analyses below utilize the recoded scores for IOS_{11} (as per final column of Table 1) to allow for direct comparisons between methodologies. Fig. 2 plots scores of IOS, We, Oneness (the arithmetic mean of IOS plus We), and IOS_{11} for each level of relationship. The box plots capture the interquartile range for each measure and the underlying distributions are indicated by the circles above the boxes. The different colors indicate whether the person thought of was a close person (dark blue), a friend (blue) or an acquaintance (light blue). The different scales (IOS, We, Oneness, and IOS_{11}) for each relationship level are then presented in separate bars from top to bottom.

Fig. 2 shows that for all four instruments, there is clear and coherent variation in reported closeness comparing different relationship levels. Participants who considered a close person reported significantly higher scores than those who considered a friend (p-values < 0.001) and scores for those considering an acquaintance were lower still (p-values < 0.001).[14] Moreover, the figure also shows that reported levels of closeness are quite similar across methods. Notwithstanding this general coherence, Fig. 2 reveals some apparent differences across the distributions of scores for different methods, in the comparison of IOS and We scale scores. Notice that for ratings of a close person, the interquartile range and median value for We scale lie to the right of that for IOS reflecting, in part, a markedly stronger tendency for participants to record maximum values on the We scale, relative to IOS (p-value = 0.001 for Kolmogorov-Smirnov (KS) test comparing the two distributions).

[14] Test statistics are based on comparing the underlying distributions using Kolmogorov-Smirnov tests and here we report the highest p-value of pairwise comparisons of all measurement tools across relationship levels. Further details are in the online Appendix A, Table A1.

Figure 2

Relationship levels, elicitation tools and recorded scores



Note. Different relationship levels are presented in three distinct panels. Close person (top; dark blue), friend (middle; blue), and acquaintance (bottom; light blue). In each panel, we present scores of *IOS*, *We scale*, *Oneness* and *IOS$_{11}$* from top to bottom. Oneness is the arithmetic mean of IOS and We scale. IOS$_{11}$ scores are recoded as defined in Table 2. The boxplots capture the median and the interquartile range. The whiskers range from the $10th$ to $90th$ percentile. Each circle in the distribution plot captures a unique observation.

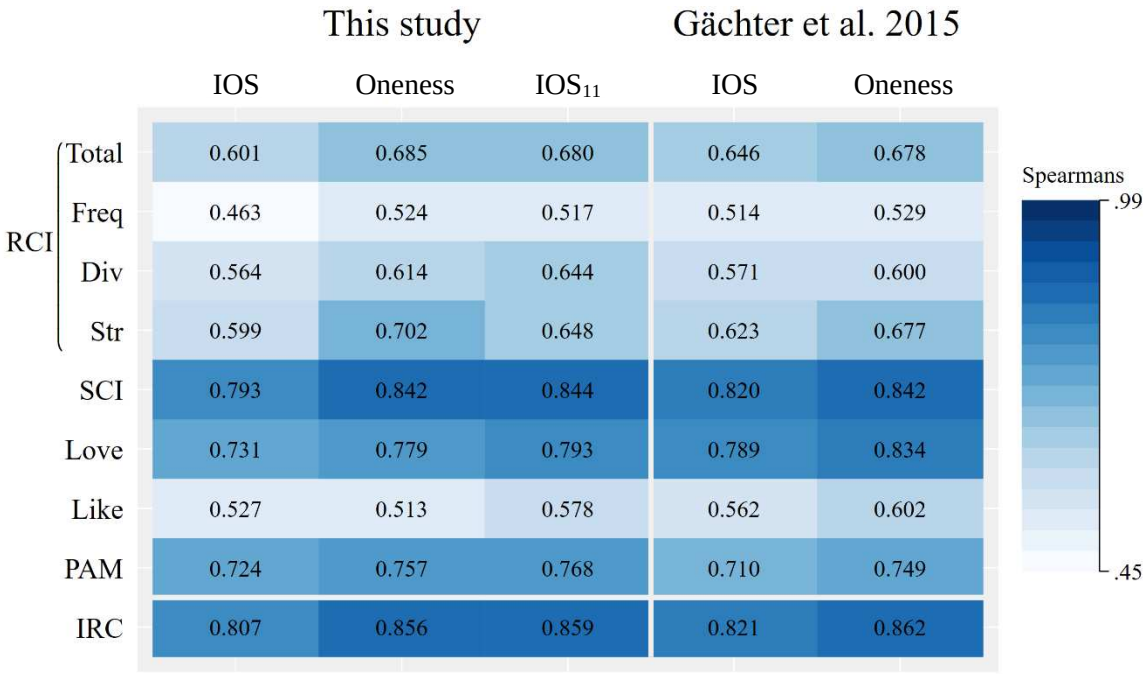
This is suggestive evidence that IOS and We scales may, to some extent, be capturing different aspects of relationship closeness and, if they are, this could be part of the explanation for why Oneness, which combines the two measures, has tended to psychometrically outperform IOS, alone. Notice, however, that relative to IOS, at the eyeball level the distribution of IOS_{11} more closely resembles that of Oneness.

Indeed, based on KS tests, IOS_{11} and Oneness are statistically indistinguishable from each other for a close person (p-value = 0.164), a friend (p-value = 0.966) and an acquaintance (p-value = 0.095). To the extent that Oneness outperforms IOS in tracking other measures of relationship closeness, these results suggest the possibility that IOS_{11} might close some of that performance gap. So, we now turn to examine exactly that issue.

Table 2 reports within-participant Spearman's rank correlations between IOS, Oneness and IOS_{11} (columns) and a set of nine benchmark scales (rows). Columns 1 to 3 display the results for IOS, Oneness, and IOS_{11} from our study, whereas columns 4 and 5 reproduce results for IOS and Oneness from Gächter et al. (2015) for comparison. The first row reports correlations with the overall RCI benchmark and the next three rows report correlations with its three sub-components (frequency, diversity, and strength). "Love" and "Like" are two elements of LLS. The final row reports correlations with an Index of relationship closeness (IRC); this is a single index but derived from the set of other benchmarks using a principal components analysis, following the approach of Gächter et al. (2015). Across the table, we find moderately-strong to strong correlations throughout and all are statistically significant at the 1% level. A companion Table 3 reports pairwise tests of differences between correlation coefficients (IOS vs Oneness; IOS vs IOS_{11}; IOS_{11} vs Oneness).

Table 2

Correlations across relationship scales



Note. All cells in the figure represent Spearman's rank correlations that are significant at the 1% level. Benchmark scales are in the rows and the measures of relationship closeness in the different columns. RCI is the *Relationship Closeness Inventory* with its subdomains *Frequency*, *Diversity* and *Strength*. SCI indicates the *Subjective Closeness Index (SCI)*, Love the *Love scale* and *Like the Liking scale*. PAM refers to the *Personal Acquaintance Measure* and IRC to the *Index of Relationship Closeness*.

Eyeballing of Table 2 combined with the tests presented in Table 3, reveals three broad patterns. First, correlations between Oneness and the various benchmarks tend to be systematically higher than those between the benchmarks and original IOS (in Table 3, comparing IOS vs Oneness, there are two cases where the correlation is significantly higher for Oneness, at the 5% level or higher, and none in the opposite direction). Second, our new IOS$_{11}$ tool outperforms the original IOS (in Table 3, there are three cases where IOS$_{11}$ has a significantly higher correlation with a comparator benchmark, at the 5% level or better, and no cases where IOS performs better).

Table 3

Pairwise comparisons of correlation coefficients

	IOS vs. Oneness	IOS vs IOS$_{11}$	Oneness vs IOS$_{11}$
RCI Total	0.059	0.078	0.894
RCI Frequency	0.284	0.346	0.890
RCI Diversity	0.315	0.097	0.518
RCI Strength	0.019	0.285	0.197
SCI	0.054	0.043	0.926
Love	0.140	0.049	0.630
Like	0.795	0.336	0.221
PAM	0.345	0.194	0.726
IRC	0.037	0.024	0.876

Note. P-values of a test of equality of correlation coefficients described in Cohen et al. (2013). The table rows correspond with Table 2 by presenting benchmark scales. The three columns, respectively, present results for comparisons of: IOS vs Oneness; IOS vs IOS$_{11}$ and Oneness vs IOS$_{11}$. RCI is the *Relationship Closeness Inventory* with its subdomains *Frequency, Diversity* and *Strength*. SCI is the *Subjective Closeness Index (SCI), the next two rows are Love* and *Liking scales.* PAM is the *Personal Acquaintance Measure* and IRC is the *Index of Relationship Closeness*.

Thirdly, we find no significant differences at all comparing the correlations of Oneness and IOS$_{11}$ for each of the nine benchmarks (in Table 2, across the nine benchmarks, differences go in both directions, but they are never significantly different at the 5% level and few of the p-values in the final column of Table 3 are close to significance at any conventional level).

The three broad patterns just identified each hold for IRC: this is meaningful because the IRC is arguably the most informative of the benchmarks (by virtue of being distilled from the larger set of measures via principal components analysis). More specifically, based on results reported in the final row of Table 3: we replicate the finding of Gächter et al. (2015) that Oneness outperforms IOS in terms of its correlation with IRC (p-value = 0.037 in Table 3); we see that the correlation of IOS$_{11}$ with IRC is stronger than that for the original IOS measure (p-value = 0.024); and it is statistically indistinguishable from Oneness (p-value =

0.876).

It is also worth noting that we replicate the findings from Gächter et al. (2015) remarkably well in finding correlation coefficients that very closely mimic the original results. This is particularly noteworthy as we utilized a different study population (US vs. UK), and a substantive amount of time has passed since the original data collection (2014 vs 2021).[15]

Based on these results we summarize our main finding as follows: our new tool, IOS_{11}, matches the performance of Oneness in terms of its correlation with a set of established measures of relationship closeness but it does so whilst maintaining the simplicity of the single-item IOS scale.

Conclusion

In this paper, we have proposed IOS_{11} as a new tool for eliciting relationship closeness. Previous research has established that very simple and portable tools such as the original IOS and Oneness provide remarkably effective tools for assessing relationship closeness. And, the volume of literature using them and the current growth of interest in novel applications – in particular in ever more frequent online studies – attests to their usefulness for a wide range of research purposes, old and new.

Our work complements ongoing research developing measurement techniques for relationship closeness such as Le et al. (2007), Kamphorst et al. (2017) and Beranek and Castillo (2022). The first two of these studies develop online versions of IOS and like us, both conjecture that a more fine-grained measurement tool may increase precision but, unlike us, neither study tests this. Beranek and Castillo (2022) compare scores obtained from the standard IOS tool with a continuous version and a step-choice version. Unlike us, however,

[15] We also find that based on KS tests Oneness scores in Gächter et al. (2015) and our study are statistically identical for a close person ($\mu_{Gächter} = 5.367$, $\mu_{Study} = 5.658$, p-value = 0.236) and an acquaintance ($\mu_{Gächter} = 2.390$, $\mu_{Study} = 2.641$, p-value = 0.255), but differ significantly for a friend ($\mu_{Gächter} = 3.762$, $\mu_{Study} = 4.227$, p-value = 0.020).

they do not benchmark to Oneness or other more complex tools for assessing relationship closeness.

Our primary contribution lies in addressing the issue that, until now, researchers considering using IOS-like tools have faced a tradeoff between the simplicity of the single-item IOS measure and the added accuracy of the two-item Oneness measure. Our new tool resolves this tension by offering a new 11-point version of the IOS scale which, according to our results, is indistinguishable from the Oneness measure in terms of its ability to track a range of more complex measures of relationship closeness. As such, for those considering the use of some IOS-style measurement tools, IOS$_{11}$ provides a convenient, highly portable, and efficient method for elicitation in any computerized environment.

Declarations

Supplementary Materials

Additional analyses and experimental instructions are available in the online supplementary materials (https://dl.dropbox.com/s/6ywrg4yai2r903e/IOS11_BSTG_2023-10-13_OnlineAppendix.pdf)

Funding

This work was supported by the British Academy/Leverhulme Trust [grant number SRG1819\190917], the European Research Council [grant number ERC-AdG 101020453 PRINCIPLES] and the University of Nottingham (UK).

Competing interests

The authors declare that there were no conflicts of interest with respect to the authorship or the publication of this article.

Ethics approval

This research was approved by the Research Ethics Committee in the School of Economics at the University of Nottingham.

Consent to participate

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

Consent for publication

Consent for publication was obtained from all individual participants included in the study.

Data and code availability and pre-registration

The data, the IOS$_{11}$ software, and the analysis files are available via the Open Science Framework (https:/doi.org/10.17605/OSF.IO/9DBR6)

Open science practices

The experiments were pre-registered at the American Economic Association's registry for randomized controlled trials (Registration number AEARCTR-0007947; see https://www.socialscienceregistry.org/trials/7947).

Authors' contributions

M.B., S.G., C.S. and F.T. developed the study concept. M.B. developed the IOS$_{11}$ software and conducted the experiments. M.B. and F.T. analyzed the data. M.B., F.T. and C.S. drafted the manuscript with S.G. providing critical revisions. All authors approved the final version for submission.

References

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of personality and social psychology*, *63*(4), 596. https://doi.org/10.1037/0022-3514.63.4.596

Bartz, J. A., Zaki, J., Ochsner, K. N., Bolger, N., Kolevzon, A., Ludwig, N., & Lydon, J. E. (2010). Effects of oxytocin on recollections of maternal care and closeness. *Proceedings of the National Academy of Sciences*, *107*(50), 21371-21375. https://doi.org/10.1073/pnas.1012669107

Beranek, B., & Castillo, G. (2022). *Continuous Inclusion of Other in the Self*.

Berscheid, E., Snyder, M., & Omoto, A. M. (1989). The Relationship Closeness Inventory: Assessing the closeness of interpersonal relationships. *Journal of personality and social psychology*, *57*(5), 792. https://doi.org/10.1037/0022-3514.57.5.792

Bicchieri, C., Dimant, E., Gächter, S., & Nosenzo, D. (2022). Social proximity and the erosion of norm compliance. *Games and Economic Behavior*, *132*, 59-72. https://doi.org/10.1016/j.geb.2021.11.012

Brewer, M. B., & Gardner, W. (1996). Who is this" We"? Levels of collective identity and self representations. *Journal of personality and social psychology*, *71*(1), 83. https://doi.org/10.1037/0022-3514.71.1.83

Calbi, M., Langiulli, N., Ferroni, F., Montalti, M., Kolesnikov, A., Gallese, V., & Umilta, M. A. (2021). The consequences of COVID-19 on social interactions: an online study on face covering. *Scientific reports*, *11*(1), 2601. https://doi.org/10.1038/s41598-021-81780-w

Castillo, G. (2021). Preference reversals with social distances. *Journal of Economic Psychology*, *86*, 102410. https://doi.org/10.1016/j.joep.2021.102410

Cialdini, R. B., Brown, S. L., Lewis, B. P., Luce, C., & Neuberg, S. L. (1997). Reinterpreting the empathy–altruism relationship: When one into one equals oneness. *Journal of personality and social psychology*, *73*(3), 481. https://doi.org/10.1037/0022-3514.73.3.481

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge.

Dimant, E. (2023). Hate trumps love: The impact of political polarization on social preferences. *Management Science*. https://doi.org/10.1287/mnsc.2023.4701

Gächter, S., Starmer, C., Thöni, C., Tufano, F., & Weber, T. O. (2022). Social closeness can help, harm and be irrelevant in solving pure coordination problems. *Economics Letters*, *216*, 110552. https://doi.org/10.1016/j.econlet.2022.110552

Gächter, S., Starmer, C., & Tufano, F. (2015). Measuring the closeness of relationships: a comprehensive evaluation of the'inclusion of the other in the self'scale. *PloS one*, *10*(6), e0129478. https://doi.org/10.1371/journal.pone.0129478

Gächter, S., Starmer, C., & Tufano, F. (2023). Measuring “group cohesion” to reveal the power of social relationships in team production. *Review of Economics and Statistics*, 1-45. https://doi.org/10.1162/rest_a_01283

Gager, C. T., & Sanchez, L. (2003). Two as one? Couples’ perceptions of time spent together, marital quality, and the risk of divorce. *Journal of Family issues*, *24*(1), 21-50. https://doi.org/10.1177/0192513X02238519

Gillen, B., Snowberg, E., & Yariv, L. (2019). Experimenting with measurement error: Techniques with applications to the caltech cohort study. *Journal of Political Economy*, *127*(4), 1826-1863. https://doi.org/10.1086/701681

Goette, L., & Tripodi, E. (2021). Social influence in prosocial behavior: evidence from a large-scale experiment. *Journal of the European Economic Association*, *19*(4), 2373-2398. https://doi.org/10.1093/jeea/jvaa054

Guldner, G. T., & Swensen, C. H. (1995). Time spent together and relationship quality: Long-distance relationships as a test case. *Journal of social and Personal Relationships*, *12*(2), 313-320. https://doi.org/10.1177/0265407595122010

Kamphorst, B. A., Nauts, S., & Blouin-Hudon, E.-M. (2017). Introducing a continuous measure of future self-continuity. *Social Science Computer Review*, *35*(3), 417-421. https://doi.org/10.1177/0894439316653513

Kastendieck, T., Zillmer, S., & Hess, U. (2022). (Un) mask yourself! Effects of face masks on facial mimicry and emotion perception during the COVID-19 pandemic. *Cognition and Emotion*, *36*(1), 59-69. https://doi.org/10.1080/02699931.2021.1950639

Laursen, B., Noack, P., Wilder, D., & Williams, V. (2000). Adolescent perceptions of reciprocity, authority, and closeness in relationships with mothers, fathers, and friends. *International Journal of Behavioral Development*, *24*(4), 464-471. https://doi.org/10.1080/016502500750038017

Le, B., Moss, W. B., & Mashek, D. (2007). Assessing relationship closeness online: Moving from an interval-scaled to continuous measure of including others in the self. *Social Science Computer Review*, *25*(3), 405-409. https://doi.org/10.1177/0894439307297693

Markowitz, D. M., & Slovic, P. (2020). Social, psychological, and demographic characteristics of dehumanization toward immigrants. *Proceedings of the National Academy of Sciences*, *117*(17), 9260-9269. https://doi.org/10.1073/pnas.1921790117

Mashek, D. J., & Aron, A. (2004). *Handbook of closeness and intimacy*. Psychology Press.

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods*, *44*(1), 1-23. https://doi.org/10.3758/s13428-011-0124-6

Miller, M. R., Jun, H., Herrera, F., Yu Villa, J., Welch, G., & Bailenson, J. N. (2019). Social interaction in augmented reality. *PloS one*, *14*(5), e0216290. https://doi.org/10.1371/journal.pone.0216290

Molleman, L., & Gächter, S. (2018). Societal background influences social learning in cooperative decision making. *Evolution and Human Behavior*, *39*(5), 547-555. https://doi.org/10.1016/j.evolhumbehav.2018.05.007

Neuberg, S. L., Cialdini, R. B., Brown, S. L., Luce, C., Sagarin, B. J., & Lewis, B. P. (1997). Does empathy lead to anything more than superficial helping? Comment on Batson et al.(1997). https://doi.org/10.1037/0022-3514.73.3.510

Parisi, R., Lagomarsino, F., Rania, N., & Coppola, I. (2021). Women face to fear and safety devices during the COVID-19 pandemic in Italy: Impact of physical distancing on individual responsibility, intimate, and social relationship. *Frontiers in public health*, *9*, 622155. https://doi.org/10.3389/fpubh.2021.622155

Paunonen, S. V. (1989). Consensus in personality judgments: Moderating effects of target-rater acquaintanceship and behavior observability. *Journal of personality and social psychology*, *56*(5), 823. https://doi.org/10.1037/0022-3514.56.5.823

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. *Behavior research methods*, 1. https://doi.org/10.3758/s13428-021-01694-3

Pellencin, E., Paladino, M. P., Herbelin, B., & Serino, A. (2018). Social perception of others shapes one's own multisensory peripersonal space. *Cortex*, *104*, 163-179. https://doi.org/10.1016/j.cortex.2017.08.033

Qualtrics. (2021). *The code for this paper was generated using Qualtrics software. Copyright © 2021 Qualtrics. Qualtrics and all other Qualtrics product or service names are registered trademarks or trademarks of Qualtrics*. In https://www.qualtrics.com

Reis, H. T., Collins, W. A., & Berscheid, E. (2000). The relationship context of human behavior and development. *Psychological bulletin*, *126*(6), 844. https://doi.org/10.1037/0033-2909.126.6.844

Robson, M. (2021). Inequality aversion, self-interest and social connectedness. *Journal of Economic Behavior & Organization*, *183*, 744-772. https://doi.org/10.1016/j.jebo.2020.12.029

Rubin, Z. (1970). Measurement of romantic love. *Journal of personality and social psychology*, *16*(2), 265. https://doi.org/10.1037/h0029841

Simms, L. J., Zelazny, K., Williams, T. F., & Bernstein, L. (2019). Does the number of response options matter? Psychometric perspectives using personality questionnaire data. *Psychological assessment*, *31*(4), 557. https://doi.org/10.1037/pas0000648

Snowberg, E., & Yariv, L. (2021). Testing the waters: Behavior across participant pools. *American Economic Review*, *111*(2), 687-719. https://doi.org/10.1257/aer.20181065

Starzyk, K. B., Holden, R. R., Fabrigar, L. R., & MacDonald, T. K. (2006). The Personal Acquaintance Measure: a tool for appraising one's acquaintance with any person. *Journal of personality and social psychology*, *90*(5), 833. https://doi.org/10.1037/0022-3514.90.5.833

Tarr, B., Launay, J., & Dunbar, R. I. (2016). Silent disco: dancing in synchrony leads to elevated pain thresholds and social closeness. *Evolution and Human Behavior*, *37*(5), 343-349. https://doi.org/10.1016/j.evolhumbehav.2016.02.004

Tropp, L. R., & Wright, S. C. (2001). Ingroup identification as the inclusion of ingroup in the self. *Personality and Social Psychology Bulletin*, *27*(5), 585-600. https://doi.org/10.1177/0146167201275007

Waugh, C. E., & Fredrickson, B. L. (2006). Nice to know you: Positive emotions, self–other overlap, and complex understanding in the formation of a new relationship. *The journal of positive psychology*, *1*(2), 93-106. https://doi.org/10.1080/17439760500510569

Wei, R., & Lo, V.-H. (2006). Staying connected while on the move: Cell phone use and social connectedness. *New media & society*, *8*(1), 53-72. https://doi.org/10.1177/1461444806059870

Westlund, J. M. K., Park, H. W., Williams, R., & Breazeal, C. (2018). Measuring young children's long-term relationships with social robots. Proceedings of the 17th ACM conference on interaction design and children,