UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Improving the measurement of prosociality through aggregation of game behavior

Haesevoets, T.; Reinders Folmer, C.; Van Hiel, A.

DOI
10.1016/j.copsyc.2021.09.018

Publication date
2022

Document Version
Final published version

Published in
Current Opinion in Psychology

License
Article 25fa Dutch Copyright Act (https://www.openaccess.nl/en/policies/open-access-in-dutch-copyright-law-taverne-amendment)

Link to publication

Citation for published version (APA):
Haesevoets, T., Reinders Folmer, C., & Van Hiel, A. (2022). Improving the measurement of prosociality through aggregation of game behavior. *Current Opinion in Psychology*, *44*, 237-244. https://doi.org/10.1016/j.copsyc.2021.09.018

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.
UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:05 Sept 2026

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Psychology

Review

Improving the measurement of prosociality through aggregation of game behavior

Tessa Haesevoets[1], Chris Reinders Folmer[2] and Alain Van Hiel[1]

Abstract

Prior research has found that people's choices in economic games are often only modestly related to their prosocial personality traits and to mundane prosocial behaviors. The present article reviews the recent literature showing that the strength of these relationships depends on the level of aggregation. Specifically, we demonstrate an increase in behavioral consistency after horizontal aggregation (across multiple game types), vertical aggregation (across multiple game variants), and a combination thereof. Moreover, we show that aggregation increases the magnitude of the relationships of game behavior with prosocial personality and mundane prosocial behavior. These findings illustrate that economic games can genuinely capture a core facet of human prosociality — but that their capacity for doing so is greater when multiple game behaviors are considered.

Addresses

[1] Department of Developmental, Personality and Social Psychology, Ghent University, Belgium

[2] Amsterdam Law School, University of Amsterdam, the Netherlands

Corresponding author: Haesevoets, Tessa (tessa.haesevoets@ugent.be)

Current Opinion in Psychology 2022, **44**:237–244

This review comes from a themed issue on **Prosociality (2022)**

Edited by **Stefan Pfattheicher** and **Isabel Thielmann**

For complete overview about the section, refer Prosociality (2022)

Available online 30 September 2021

https://doi.org/10.1016/j.copsyc.2021.09.018

2352-250X/© 2021 Elsevier Ltd. All rights reserved.

Keywords

Social dilemmas, Economic games, Behavioral consistency, Horizontal aggregation, Vertical aggregation, Multiple aggregation, Prosocial personality, Mundane prosocial behavior.

Introduction

Social dilemmas are situations in which collective interests are at odds with selfish interests [1–4]. Such situations include many of the principal challenges of our time, including climate change [5] and pandemic mitigation [6]. To investigate behavior in social dilemmas empirically, scholars have modeled these situations into so-called *economic games* in which people have to choose between cooperative (reflecting a prosocial choice) and defective (reflecting a proself choice) alternatives [7,8]. This approach has resulted in an extensive range of games that are used to model social dilemma situations [9,10], and to understand people's actions within them. Moreover, such games are used to derive possible resolutions through which mundane forms of prosocial behavior may be promoted. Prosocial behavior covers a broad range of actions intended to benefit one or more people other than oneself [11].

Although research on economic games has yielded a wealth of insight into how people act in such experimental settings (e.g. [4,9,10]), several previous studies failed to report large[1] associations between people's behavior in different economic games and their prosocial personality and prosocial behavior in mundane settings (i.e. real-life prosocial behaviors such as donating blood and volunteering). An important limitation of many prior studies in this domain, however, is that they typically relied on individual game behaviors — that is, for the most, they focus on a single, particular game type, presented in one specific version, often played in a one-shot manner [13,14].

In this article, we argue that this particular approach can possibly explain the rather modest and varying associations of game behavior with prosocial personality and real-life prosocial behaviors that have been observed in much prior research. More specifically, we suggest that aggregating choices within and across a broader range of settings increases the reliability of game-based measures and, thereby, also boosts the correlations with both trait prosociality and prosocial behavior outside of the laboratory. We begin this review with a brief discussion of the aggregation principle. Next, we provide a detailed review of recent research on the consistency of people's choices in game-based social dilemmas and their link with prosocial personality and mundane forms of prosocial behavior.

[1] To interpret the magnitudes of correlation coefficients, we use the recent guidelines of Funder and Ozer [12] who consider correlations of 0.05, 0.10, 0.20, 0.30, and 0.40 as very small, small, moderate, large, and very large, respectively.

The aggregation principle

Imagine assessing students' course performance with a single multiple-choice item. Most people would agree that in such a case the use of a single-item measure would probably result in an unreliable measurement. Yet, within the social dilemma literature, it is common practice to measure game behavior with only one single trial of one particular game (e.g. a single trial of a Prisoner's Dilemma). This overlooks the notion; however, that choice behavior in economic games may be strongly shaped not only by the structural differences between different game types (e.g. a Prisoner's Dilemma vs a Public Goods Dilemma) but also by differences in operationalization of the same type of game (e.g. a Prisoner's Dilemma with symmetrical outcomes vs a Prisoner's Dilemma with asymmetrical outcomes). As such, choice behavior in a single game type, a single variant of that game, let alone a single trial of that game, is likely to be severely limited as an indicator of people's tendency to act prosocially, compared with decisions across multiple game types and/or game variants.

The aim of the present review is to illustrate that some of the weaker than expected relationships among game behaviors and between game behavior and third variables reflecting prosociality (such as trait prosociality and mundane forms of prosocial behavior) obtained in prior studies (e.g. [15–17]) can, at least in part, be a consequence of failures to aggregate. According to the aggregation principle, the sum of a set of multiple measurements is a more stable and unbiased estimator than any single measurement from the set [18]. This greater representation occurs because single measures may be unreliable due to measurement errors. By combining multiple measures, such measurement errors are averaged out, thereby resulting in a measure that is 'more reflective of psychological reality' ([18], p. 21, also see [19]).

A well-known illustration of the aggregation principle concerns the observation that the reliability of an instrument increases when the number of items increases [20,21]. For example, single items of the Stanford-Binet Intelligence test correlate only 0.15 with each other, subtests based on four or five items correlate around 0.30 or 0.40, whereas the aggregated battery of items that make up the Performance subscale correlates around 0.80 with the battery of items that forms the Verbal subscale. Particularly interesting in the context of the present research question, however, is that Rushton, Brainerd, and Pressley ([18], p. 19) have demonstrated, for 12 different research domains, that when constructs are assessed with multiple measures 'relationships become more substantial' (for applications of the aggregation principle, see [22,23]; for more information on the pro and cons of single and multiple measures, see [24–28]).

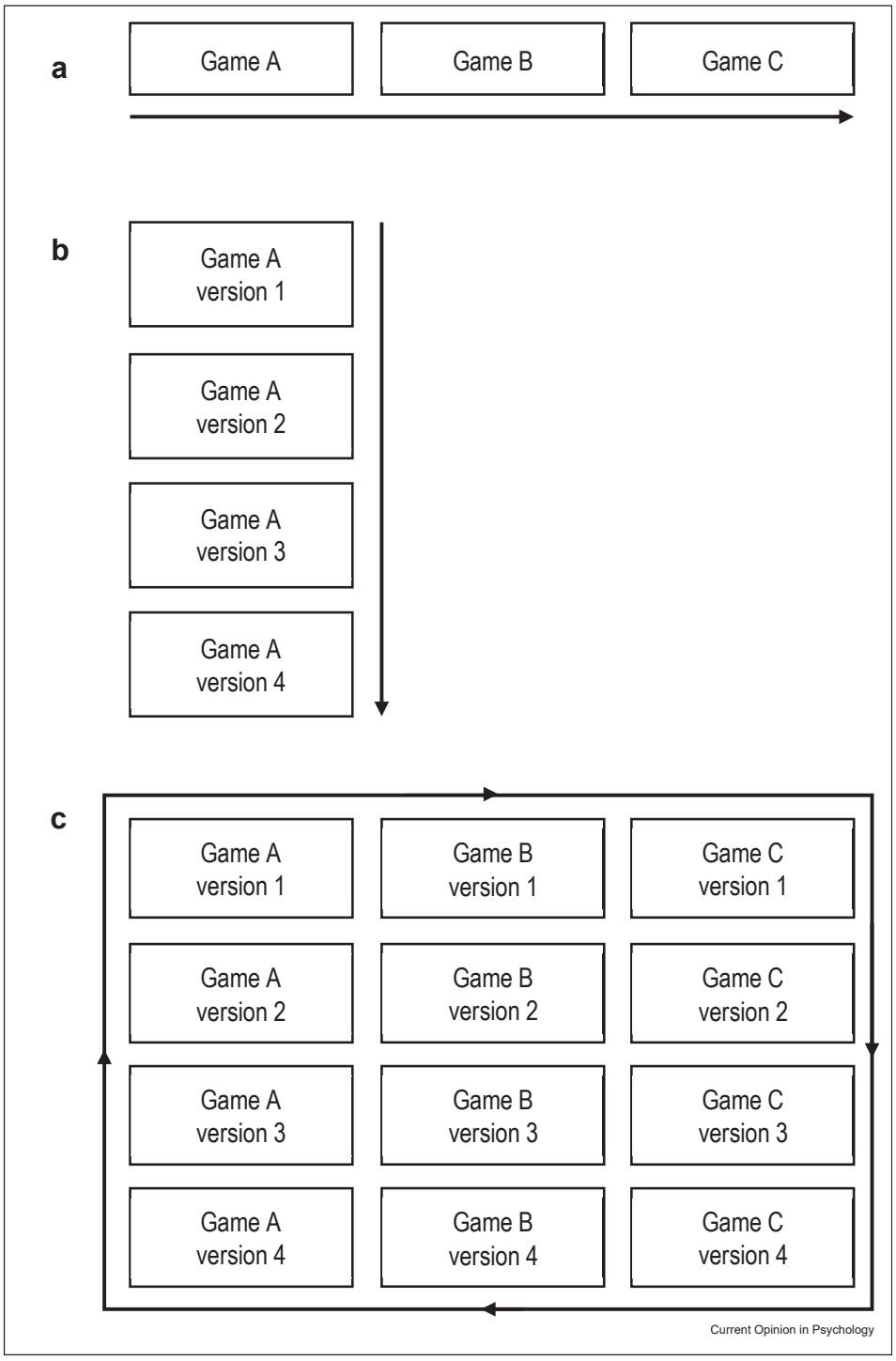
Behavioral consistency in social dilemma games

Generally speaking, this article is a methodological critique of the way in which game behavior has been measured in ample prior social dilemma research. In what follows, evidence is presented that this research domain has been limited by the tendency to focus on choice behavior in isolated games (e.g. within a single game type or game variant; often in just a single interaction or a single game trial), instead of aggregating multiple game behaviors. We distinguish between three types of aggregation. The first type of aggregation concerns aggregation across a series of different games, which we label *horizontal aggregation*. The second type of aggregation refers to aggregating choices across multiple variants of the same game, which we label *vertical aggregation*. Obviously, these two types of aggregation can also be combined, thereby reflecting aggregation over multiple variants of different games (i.e. *multiple aggregation*). Figure 1 visually displays these three different types of aggregation.

To illustrate that some of the weaker than expected statistical relationships observed in prior studies can at least in part be attributed to failures to use such aggregated game-based responses, in the following we provide empirical evidence on the horizontal aggregation principle, by showing consistency of people's choices across a series of different game types (see Horizontal aggregation: behavioral consistency across a series of different games), the vertical aggregation principle, by showing consistency in multiple versions of a single game type (see Vertical aggregation: behavioral consistency across multiple versions of a single game), and, finally, the combination of both these aggregation types, by showing even higher consistency in multiple versions across different game types (see Multiple aggregation: behavioral consistency across different versions of different games). For each of these three steps, it is examined to what extent — discrete and aggregated — game behavior relates with prosocial personality and, where available, with (self-reported) prosocial behavior in mundane settings. If the relationships between game behavior and trait prosociality and/or behavioral prosociality become stronger after aggregation, then this would imply that economic games are indeed 'reflective of psychological reality' [18] and thus measure a common core aspect of human prosociality.

Importantly, the present review exclusively focuses on the class of social dilemma games. Social dilemma games — such as the Prisoner's Dilemma, the Public Goods Dilemma, and the Commons Dilemma — are economic games in which two or more players independently (and usually simultaneously) decide between prosocial and selfish alternatives. The most

Figure 1



Three types of aggregation: (**a**) horizontal aggregation, (**b**) vertical aggregation, and (**c**) multiple aggregation.

defining feature of such games is that acting prosocially increases social welfare (i.e. the sum of all players' outcomes), whereas at the same time, it may decrease a player's individual outcome. As such, social dilemma games are characterized by partial negative interdependence between players' payoffs [1–3,10]. Note that the present review does not include other game classes, such as constant-sum games (which are characterized

by perfect negative interdependence between players' payoffs) or coordination games (which are characterized by positive interdependence between players' payoffs).

Horizontal aggregation: behavioral consistency across a series of different games

Because a conflict between self-interest and the welfare of others is at the basis of all social dilemma games, it would be expected that choice behavior across different games shows substantial consistency. In line with this reasoning, several prior studies have found relatively strong consistency in behavior across different games (e.g. [29,30], also see [31–33]). Other studies, however, have reported only small to moderate correlations between different games (e.g. [15]). Yet, it must be noted that this latter research not only included social dilemma games but also games from other classes (such as a Dictator Game), which may have suppressed the intercorrelations among the game behaviors.

Particularly interesting in the context of the present research question is our own prior research [34**] in which we examined consistency in choice behavior across five social dilemma games, namely, A Prisoner's Dilemma, an Assurance Game, a Public Goods Dilemma, a Commons Dilemma, and a Trust Game in the role of a trustor (see Table 1 for a brief overview of the structural features of these games). We found a large average correlation of $r = 0.31$ across these five different games (see Appendix A of the Supplementary Materials for more information)[2,3]. This large average correlation suggests that 'across-game'consistency is rather high: When a person acts prosocially in one particular social dilemma game (e.g. a Prisoner's Dilemma), it is likely that he or she will also act prosocially in other social dilemma games (e.g. a Public Goods Dilemma and a Commons Dilemma). Yet, it is important to note that the exact strength of these relationships is dependent on which specific games are compared (i.e. aggregation can reasonably be expected to work better for game types with a similar — rather than a dissimilar — game structure).

In this study, we ([34**], also see [35*]) did not just focus on the association between different game types but also on the association between choice behavior and individual differences in prosocial personality, as measured by Social Value Orientation (SVO), Right-Wing Authoritarianism (RWA), and Social Dominance Orientation (SDO). In addition, the HEXACO Agreeableness and Honesty-Humility dimensions were also measured. More importantly, however, we did not just focus on the association between single game types and prosocial personality but also that between prosocial personality and aggregated game behavior. As shown in Table A1 of Appendix A, the correlations between single game behavior and these prosocial personality measures were small to moderate ($0.08 \leq r \leq 0.18$; mean $r = 0.14$); crucially, however, after aggregating choice behavior across the five different games, correlations with prosocial personality were all moderate to large ($0.13 \leq r \leq 0.27$; mean $r = .21$).[4] From these findings, it can thus be derived that the link between game behavior and prosocial personality becomes somewhat stronger after horizontal aggregation is applied (although obviously, this will only apply for personality traits that are conceptually linked to the game behaviors). Accordingly, horizontal aggregation of choices across different game types appears to capture a common core of decision behavior that is more aligned with prosocial personality than are isolated choices in a specific game type.

Vertical aggregation: behavioral consistency across multiple versions of a single game

Besides consistency across different types of games, a further crucial question is whether people show consistency in choice behavior across different versions of the same game. As noted, studies that rely on economic games have tended to focus on decisions within one specific game type (e.g. [36–40]). Yet, even studies that have used the same game type have often been performed in different variants, which differ, for example, in terms of the endowments at stake, the (a)symmetry in outcomes, and/or the level of conflict between players' choices (i.e. the degree to which their outcomes are aligned or opposed, which can vary in different operationalizations of the same game). Although such game variants are rooted in the same game structure (e.g. that of the Prisoner's Dilemma), such game variations may nevertheless have considerable impact on choice behavior. Yet, because these different game variants all pertain to the same game type (e.g. they are all Prisoner's Dilemmas), it would be expected that choice behavior across different variants of the same game also shows substantial consistency. Most prior research, however, does not allow us to verify this, as this research generally did not investigate the consistency of people's choices but instead focused on between-participant differences due to treatments.

A notable exception in this regard is the recent study by Haesevoets, Reinders Folmer, Bostyn, and Van Hiel [41**], which explored the consistency of people's choices across 16 different versions of a Prisoner's

[2] Note that high scores in the Commons Dilemma reflect selfish (instead of prosocial) behavior. To ensure that all game behaviors point in the same direction, throughout this article, for this game, the signs of the correlations were reversed, such that higher scores indicate greater prosocial behavior.

[3] Besides these five social dilemma games, Haesevoets et al. also included a Dictator Game and an Ultimatum Bargaining Game, which are, by definition, no social dilemma games (but instead constant-sum games). Because of this, these two games are not included in the calculation of this average correlation. If these two games are also included in the calculation, the average correlation ($r = 0.22$) is only moderate.

[4] Note that RWA and SDO reflect non-prosocial (instead of prosocial) traits (i.e. high levels of RWA and SDO are linked with lower cooperation). To ensure that all personality traits point in the same direction, throughout this article, for these two traits, the signs of the correlations were also reversed.

Table 1

Summary of the five social dilemma games used by Haesevoets et al. [34].**

Game type	Number of players	Type of interaction	Type of choice	Game structure
Prisoner's Dilemma	Two players	Simultaneous	Dichotomous	Two players decide simultaneously whether to cooperate or defect. The best individual outcome is unilateral defection (DC), followed by mutual cooperation (CC), mutual defection (DD), and unilateral cooperation (CD).
Assurance Game[a]	Two players	Simultaneous	Dichotomous	Two players decide simultaneously whether to cooperate or defect. Here, the best individual outcome is mutual cooperation (CC), followed by unilateral defection (DC), mutual defection (DD), and unilateral cooperation (CD).
Public Goods Dilemma	Multiple players	Simultaneous	Continuous	Multiple players decide simultaneously how much resources to contribute to a public good. The contributed resources are multiplied by a factor larger than one and shared equally across all players, irrespective of their individual contributions.
Commons Dilemma	Multiple players	Simultaneous	Continuous	Multiple players decide simultaneously how much resources to harvest from a shared resource. The resources which are not harvested are multiplied by a factor larger than one and shared equally across all players, irrespective of their individual harvests.
Trust Game (trustor)[b]	Two players	Sequential	Continuous	A trustor decides how much resources to transfer to a trustee. The amount that the trustor transfers is (usually) tripled and added to the trustee's earnings. As a response, the trustee can transfer any amount back to the trustor.

[a] A common misunderstanding is that the Assurance Game (also known as stag hunt) presents no social dilemma and leads inevitably to mutual cooperation. However, if in this game a person believes that the other player will defect, the best this person can do is to defect as well.

[b] The Trust Game is a social dilemma game from the perspective of the trustor, and a constant-sum game from the perspective of the trustee.

Dilemma. The authors obtained these different game variants by systematically manipulating the stakes (low vs high endowments), the (a)symmetry in outcomes (symmetrical vs asymmetrical outcomes), and the level of conflict between the players' interests (low vs medium vs high vs very high conflict). Although the game variants differed on these three crucial dimensions, correlations between choices in each of the 16 variants of this game were all positive and significant, resulting in a large average correlation of $r = .36$ and a high Cronbach's alpha of .90 (see Appendix B for more detailed information). Together, these findings also suggest high 'within game' consistency: When a person acts prosocially in one variant of a particular game type (e.g. a Prisoner's Dilemma with symmetrical outcomes), he or she is likely to also do so in other variants of that game type (e.g. a Prisoner's Dilemma with asymmetrical outcomes). Here too, the exact strength of these relationships depends on which particular game variants are compared (e.g. having symmetrical vs asymmetrical outcomes exerts a more powerful influence on people's choices than having low vs high endowments).

Haesevoets et al. [41**] also examined the link between individual game variants and prosocial personality in terms of SVO, dispositional trust, RWA, and SDO. Moreover, they again examined the association between prosocial personality and aggregated game behavior. The prosocial personality measures displayed small to moderate correlations with choice behavior in individual game variants ($0.14 \leq r \leq 0.19$; mean $r = 0.16$), whereas their associations with aggregated game behavior were moderate to large ($0.22 \leq r \leq 0.30$; mean $r = 0.26$; see Table B1 of Appendix B for more details). As such, these findings illustrate that vertical aggregation across multiple versions of the same game type also results in higher correlations with personality traits, at least when these traits are conceptually linked to the game behaviors.

Multiple aggregation: behavioral consistency across different versions of different games

Several studies have presented participants with multiple trials of multiple games (e.g. [31,42]). Particularly interesting for our research question, however, is a recent study by Haesevoets, Van Hiel, Dierckx, and Reinders Folmer [43**]. These authors presented participants with eight versions of a Prisoner's Dilemma, eight versions of a Public Goods Dilemma, and eight versions of a Commons Dilemma. This 'multi-variant-multi-game-type' approach allows a more direct comparison of consistency within and across different games. When comparing the two game types with the closest similarity in outcome structure (i.e. the Public Goods Dilemma and the Commons Dilemma), the results confirmed that consistency in choice behavior was greater within different variants of the same game type (vertical aggregation; very large average correlation of $r = 0.75$ among the eight Public Goods Dilemmas and very large average correlation of

$r = 0.63$ among the eight Commons Dilemmas)[5] than across similar variants of different game types (horizontal aggregation; very large average correlation of $r = 0.44$ between the Public Goods Dilemmas and the Commons Dilemmas with the same stakes and incentives for choosing a prosocial option). In line with theoretical expectations, these findings illustrate that 'within game' consistency (vertical aggregation) indeed appears to be stronger than 'across game' consistency (horizontal aggregation; see Appendix C for more detailed information).

Here too, the authors [43**] examined the link between game behavior and prosocial personality in terms of SVO, dispositional trust, RWA, and SDO. Again, choices in individual games displayed small to moderate correlations with these prosocial personality measures ($0.09 \leq r \leq 0.22$; mean $r = 0.11$). For each of the three game types, aggregating choices across the eight game variants (vertical aggregation) resulted in somewhat stronger relationships ($0.11 \leq r \leq 0.29$; mean $r = 0.17$). However, the strongest associations between game behavior and prosocial personality were generally obtained when decisions were aggregated to a single index of choice behavior across all game types and all game variants (multiple aggregation; $0.17 \leq r \leq 0.26$; mean $r = 0.20$; see Table C1 of Appendix C for more details). Based on these findings, it can thus be concluded that the link between game behavior and prosocial personality is the strongest after multiple aggregation (i.e. aggregation across multiple game types as well as multiple game variants).

As a further indicator of prosociality, Haesevoets et al. [43**] also included indicators of prosocial behavior in mundane settings, including (self-reported) blood donations, donations to noble causes, volunteering, commuting behaviors, and environmental behaviors. The authors explored how game behavior was associated with such mundane prosocial behaviors and how aggregation impacted these associations (see Table C2 of Appendix C). For each of the three game types, correlations were rather small when comparing choices in individual games with the individual prosocial behaviors (Prisoner's Dilemma: $r = 0.11$; Public Goods Dilemma: $r = 0.15$; Commons Dilemma: $r = 0.07$). For the Prisoner's Dilemma (but not for the Public Goods Dilemma and the Commons Dilemma), the relationships between game behavior and mundane prosocial behavior became stronger when choices were aggregated across game variants (vertical aggregation; Prisoner's Dilemma: $r = 0.17$; Public Goods Dilemma: $r = 0.17$; Commons Dilemma: $r = 0.08$). Interestingly, associations with individual game behavior were also stronger if the mundane prosocial behaviors were aggregated, at least for the Prisoner's Dilemma ($r = 0.18$) and the Public Goods Dilemma ($r = 0.25$) but not for the Commons Dilemma ($r = 0.10$). The strongest associations between game behavior and mundane prosocial behavior were, however, found when aggregation was applied on both the side of the games and on the side of the real-life prosocial behaviors. That is, aggregated mundane prosocial behaviors displayed rather large associations to aggregated choices in the Prisoner's Dilemma ($r = 0.28$) and the Public Goods Dilemma ($r = 0.28$). But again, this was not the case for the Commons Dilemma ($r = 0.12$). These findings thus suggest that the link between game behavior and real-life prosocial behavior also becomes stronger when relying on aggregated measures of game behavior as well as aggregated measures of mundane prosocial behavior; but, here too, it is important to note that the similarity of the situational structure between mundane and game-based situations may be decisive for the exact size of these correlations.

Conclusion

To study people's reactions in social dilemmas as precisely as possible, an extensive range of economic games have been developed, which have been studied in a broad range of variants. Although understanding choice behavior within specific game types or game variants can be useful for understanding people's reactions within that particular setting, behavior within isolated games (i.e. within a single game type, a single game variant, or even a single interaction) can only be modestly related to behavior in other games (or other game variants) and to prosocial personality traits and mundane prosocial behaviors.

Our review investigated behavioral consistency in the context of social dilemma games. The results clarify that, for this particular game class, aggregation across different game types (horizontal aggregation), across multiple variants of the same game type (vertical aggregation), and across multiple variants of different game types (multiple aggregation) all three boost behavioral consistency. These three aggregation effects — and the stronger relationships that aggregation reveals with indicators of prosocial personality and mundane forms of prosocial behavior — illustrate that we can better capture 'human prosociality' (or at least a core facet thereof) when aggregated (instead of single) measures of game behavior are used. As such, these findings also clarify that to strengthen the utility of

[5] In the present study [43**] 'within game' consistency was much larger than in the study of Haesevoets et al. [41**] (in which the authors reported an average correlation of solely $r = 0.36$ among their 16 Prisoner's Dilemmas). This can possibly be explained by the exact nature of the games (solely symmetrical game variants in the former study and both symmetrical and asymmetrical game variants in the latter study). Moreover, participants in the former study had to indicate their choices on a continuous scale, whereas in the latter study, they had to choose between a cooperative and a defective alternative. The inclusion of both symmetrical and asymmetrical game variants and the use of a dichotomous choice format may have suppressed the intercorrelations in the study of Haesevoets et al. [41**].

social dilemma games for studying prosocial behavior, future research should examine game behavior across a broader range of settings — that is, across multiple game trials, multiple game variants and/or multiple game types (depending on which exact question one wants to answer).

From a practical side, our results reveal that vertical aggregation (across game variants) is generally more performant than horizontal aggregation (across game types) and should thus be preferred. Vertical aggregation also allows to strike a better balance between theoretical precision and measurement quality. Nevertheless, horizontal aggregation may provide useful insight into the common core of different game types. Because the employment of aggregated measures of game behavior may help to reduce bias due to measurement error or idiosyncrasies of particular game types or variants, we strongly recommend future social dilemma research to rely on aggregation for the study of prosocial behavior. Moreover, we also encourage researchers to further explore the value of aggregation in the context of other game classes (such as constant-sum games and coordination games).

Funding

This research was supported by Grant BOF.PDO.2017.0017.01 of the Special Research Fund (BOF, Bijzonder Onderzoeksfonds) of Ghent University.

Author contributions

TH, CRF, and AVH developed the research idea together. TH conducted the analyses. TH wrote a first draft of the article; thereafter CRF and AVH revised the first draft.

Conflict of interest statement

Nothing declared.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.copsyc.2021.09.018.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Komorita SS, Parks CD: **Interpersonal relations: mixed-motive interaction**. *Annu Rev Psychol* 1995, **46**:183–207.

2. Weber JM, Kopelman S, Messick DM: **A conceptual review of decision making in social dilemmas: applying a logic of appropriateness**. *Pers Soc Psychol Rev* 2004, **8**:281–307.

3. Kollock P: **Social dilemmas: the anatomy of cooperation**. *Annu Rev Sociol* 1998, **24**:183–214.

4. Van Lange PA, Joireman J, Parks CD, Van Dijk E: **The psychology of social dilemmas: a review**. *Organ Behav Hum Decis Process* 2013, **120**:125–141.

5. Van Lange PA, Joireman J, Milinski M: **Climate change: what psychology can offer in terms of insights and solutions**. *Curr Dir Psychol Sci* 2018, **27**:269–274.

6. Fischer M, Twardawski M, Steindorf L, Thielmann I: **Stockpiling during the COVID-19 pandemic as a real-life social dilemma: a person-situation perspective**. *J Res Pers* 2021, **91**:104075.

7. Dawes RM: **Social dilemmas**. *Annu Rev Psychol* 1980, **31**:169–193.

8. Dawes RM, Messick DM: **Social dilemmas**. *Int J Psychol* 2000, **35**:111–116.

9. Van Dijk E, De Dreu CK: **Experimental games and social decision making**. *Annu Rev Psychol* 2021, **72**:415–438.

10. Thielmann I, Böhm R, Ott M, Hilbig BE: **Economic games: an introduction and guide for research**. *Collabra: Psychology* 2021, **7**:19004.

11. Batson CD, Powell AA: *Altruism and prosocial behavior. Handbook of psychology: personality and social psychology*. John Wiley & Sons; 2003.

12. Funder DC, Ozer DJ: **Evaluating effect size in psychological research: sense and nonsense**. *Adv Methods Pract Psychol Sci* 2019, **2**:156–168.

13. Balliet D, Parks C, Joireman J: **Social value orientation and cooperation in social dilemmas: a meta-analysis**. *Group Process Intergr Relat* 2009, **12**:533–547.

14. Galizzi MM, Navarro-Martinez D: **On the external validity of social preference games: a systematic lab-field study**. *Manag Sci* 2019, **65**:976–1002.

15. Blanco M, Engelmann D, Normann HT: **A within-subject analysis of other-regarding preferences**. *Game Econ Behav* 2011, **72**:321–338.

16. Levitt S, List JA: **What do laboratory experiments measuring social preferences reveal about the real world**. *J Econ Perspect* 2007, **21**:153–174.

17. Voors M, Turley T, Kontoleon A, Bulte E, List JA: **Exploring whether behavior in context-free experiments is predictive of behavior in the field: evidence from lab and field experiments in rural Sierra Leone**. *Econ Lett* 2012, **114**:308–311.

18. Rushton JP, Brainerd CJ, Pressley M: **Behavioral development and construct validity: the principle of aggregation**. *Psychol Bull* 1983, **94**:18–38.

19. Spector PE: Summated rating scale construction: an introduction. London: Sage Publication, 1992.

20. Gulliksen H: *Theory of mental tests*. New York: Wiley; 1950.

21. Lord RM, Novick MR: *Statistical theories of mental test scores*. Reading, Mass: Addison-Wesley; 1968.

22. Fishbein M, Ajzen I: **Attitudes towards objects as predictors of single and multiple behavioral *criteria***. *Psychol Rev* 1974, **81**:59–74.

23. Jaccard JJ: **Predicting social behavior from personality traits**. *J Res Pers* 1974, **7**:358–367.

24. Bergkvist L, Rossiter JR: **The predictive validity of multiple-item versus single-item measures of the same constructs**. *J Mar Res* 2007, **44**:175–184.

25. Gardner DG, Cummings LL, Dunham RB, Pierce JL: **Single-item versus multiple-item measurement scales: an empirical comparison**. *Educ Psychol Meas* 1998, **58**:898–915.

26. Nunnally JC, Bernstein IH: *Psychometric theory*. New York: McGraw-Hill; 1994.

27. Sarstedt M, Wilczynski P: **More for less? A comparison of single-item and multi-item measures**. *Betriebswirtschaft* 2009, **69**:211–627.

28. Wanous JP, Reichers AE: **Estimating the reliability of a single-item measure**. *Psychol Rep* 1996, **78**:631–634.

29. Yamagishi T, Mifune N, Li Y, Shinada M, Hashimoto H, Horita Y, *et al.*: **Is behavioral pro-sociality game-specific? Pro-social preference and expectations of pro-sociality**. *Organ Behav Hum Decis Process* 2013, **120**:260–271.

30. Yamagishi T, Horita Y, Mifune N, Hashimoto H, Li Y, Shinada M, *et al.*: **Rejection of unfair offers in the ultimatum game is no evidence of strong reciprocity**. *Proc Natl Acad Sci Unit States Am* 2012, **109**:20364–20368.

31. McAuliffe WH, Forster DE, Pedersen EJ, McCullough ME: **Does cooperation in the laboratory reflect the operation of a broad trait?** *Eur J Pers* 2019, **33**:89–103.

32. Wilhelm O, Kaltwasser L, Hildebrandt A: **Will the real factors of prosociality please stand up? A comment on Böckler, Tusche, and Singer (2016)**. *Soc Psychol Personal Sci* 2018, **9**:493–499.

33. Peysakhovich A, Nowak M, Rand D: **Humans display a 'cooperative phenotype' that is domain general and temporally stable**. *Nat Commun* 2014, **5**:4939.

34. ** Haesevoets T, Reinders Folmer C, Van Hiel A: **cooperation in mixed–motive games: the role of individual differences in selfish and social orientation**. *Eur J Pers* 2015, **29**:445–458.
These scholars argued that if all economic games appeal to the conflict between concerns for self and others, one would expect strong consistency in behavior across games. Moreover, these authors also examined how game behavior is linked with various personality measures.

35. * Haesevoets T, Van Hiel A, Reinders Folmer C: **the underlying motives of different mixed–motive games**. *Eur J Pers* 2015, **29**:580–584.
In this article, the authors further examined how game behavior is linked with personality traits that are conceptually linked with game behavior.

36. Andreoni J, Miller J: **Giving according to GARP: an experimental test of the consistency of preferences for altruism**. *Econometrica* 2002, **70**:737–753.

37. Brosig J, Riechmann T, Weimann J: **Selfish in the end? An investigation of consistency and stability of individual behavior**. *Working Paper Series* 2007.

38. Fischbacher U, Gächter S: **Social preferences, beliefs, and the dynamics of free riding in public goods experiments**. *Am Econ Rev* 2010, **100**:541–556.

39. Au WT, Lu S, Leung H, Yam P, Fung JM: **Risk and Prisoner's Dilemma: A reinterpretation of coombs' Re-parameterization**. *J Behav Decis Making* 2012, **25**:476–490.

40. Thielmann I, Böhm R: **Who does (not) participate in intergroup conflict?** *Soc Psychol Personal Sci* 2016, **7**:778–787.

41. ** Haesevoets T, Reinders Folmer C, Bostyn DH: **Van Hiel A: behavioural consistency within the prisoner's dilemma game: the role of personality and situation**. *Eur J Pers* 2018, **32**:405–426.
This novel work examined the consistency of people's choices in multiple versions of a single game. Considerable within-game consistency in choices was found. Moreover, these authors also investigated how game behavior is associated with important personality traits.

42. Baumert A, Schlösser T, Schmitt M: **Economic games**. *Eur J Psychol Assess* 2014, **30**:178–192. 2014.

43. ** Haesevoets T, Van Hiel A, Dierckx K: **Reinders Folmer C: do multiple-trial games better reflect prosocial behavior than single-trial games?** *Judgm Decis Mak* 2020, **15**:330–345.
In this work, the authors investigated how the same participants reacted in multiple versions of different games. Their data revealed that participants reacted more consistently across different variants of the same game than across similar variants of the different games. Moreover, they also investigated how game behavior is associated with personality traits and with prosocial behavior in mundane settings.