

Psychological Bulletin

Ordinal Effects of Ostracism: A Meta-Analysis of Cyberball Studies

--Manuscript Draft--

Manuscript Number:	
Full Title:	Ordinal Effects of Ostracism: A Meta-Analysis of Cyberball Studies
Abstract:	Cyberball is widely used to study ostracism and related phenomena. We examine 120 studies (combined N = 11,869) to determine the overall effect size, and conditions under which the effect may be reversed, eliminated, or small. Our analyses showed that (1) the average ostracism effect is large ($d > 1 $) and (2) the structural aspects of the Cyberball game (i.e., number of players, gender composition of sample, origin of study, sample age, duration of ostracism, type of needs scale) are inconsequential. Further, we test a proposition by Williams (2009) that the immediate impact of ostracism is resistant to moderation, but that moderation is more likely to be observed in delayed measures. Our findings suggest that (3) both the first and last measure are moderated, and (4) time passed since being ostracized does not predict the effect sizes of the last measure. Thus, support for this proposition is tenuous, and we suggest modifications to the temporal need-threat model of ostracism.
Article Type:	Article
Keywords:	Cyberball; meta-analysis; ordinal; ostracism
Corresponding Author:	Chris Hubertus Joseph Hartgerink, BSc. Tilburg University Tilburg, NETHERLANDS
Corresponding Author E-Mail:	c.hartgerink@gmail.com
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Tilburg University
Other Authors:	Ilja van Beest
	Jelte Michiel Wicherts
	Kipling D. Williams
Corresponding Author's Secondary Institution:	
First Author:	Chris Hubertus Joseph Hartgerink, BSc.
Order of Authors Secondary Information:	
Manuscript Region of Origin:	NETHERLANDS
Suggested Reviewers:	Michael Bernstein PennState Abington mjb70@psu.edu Has published an earlier review on this topic.
	Ginette Blackhart East Tennessee State University blackhar@etsu.edu Has published review on this topic.
	Jonathan Gerber Gordon College jonathan.gerber@gordon.edu Has published earlier review on this topic.
Opposed Reviewers:	
Order of Authors:	Chris Hubertus Joseph Hartgerink, BSc.
	Ilja van Beest

	Jelte Michiel Wicherts
	Kipling D. Williams
Manuscript Classifications:	interpersonal relationships; meta-analysis; psychology

March 15th, 2014

Dolores Albarracín, Incoming Editor
University of Pennsylvania
3620 Walnut Street, Office 346
Philadelphia, PA 19104

Dear Dr. Albarracín,

We hereby submit our manuscript 'Ordinal Effects of Ostracism: A Meta-Analysis of Cyberball Studies'. We would appreciate it if you could consider our work for publication in *Psychological Bulletin*. This is an original manuscript, and is not under consideration elsewhere. The main text of the manuscript is 9473 words long, and is accompanied by 3 figures and 5 tables.

Our manuscript reports a comprehensive meta-analysis of studies of ostracism using the Cyberball paradigm. Our work provides several valuable theoretical insights on social ostracism. We also expect it to spur debate, as it both corroborates and conflicts with current theorizing. Additionally, we provide practical insights into power estimation and effect moderation when designing studies using the Cyberball game. We declare no conflicts of interest, and potential reviewers that come to mind are Michael Bernstein, Ginette Blackhart and Jonathan Gerber, who have published earlier reviews on this topic.

We note that all our research files are available on the Open Science Framework (OSF), which includes author names. This OSF page is linked to in the manuscript, and reviewers can easily discover us as authors. We do not consider this a problem, personally. The availability of our research files supersedes our anonymity during peer review.

We look forward to your reply, and hope you will find our research intriguing for review.

Kind regards,

Chris H.J. Hartgerink
also on behalf of, Ilja van Beest, Jelte M. Wicherts, Kipling D. Williams

Prof. de Moorplein 264
5037DN, Tilburg, the Netherlands
+31651621779
c.hartgerink@gmail.com

Ordinal Effects of Ostracism:
A Meta-Analysis of Cyberball Studies

Chris H.J. Hartgerink¹

Ilja Van Beest¹

Jelte M. Wicherts¹

Kipling D. Williams²

¹ Tilburg University, the Netherlands

² Purdue University, United States of America

Word count (main text): 9473 words

Word count (abstract): 159 words

Word count (references, tables, footnotes, figure notes): 4224 words

Author note

We would like to kindly thank all researchers who sent in their unpublished studies and helped us collect all the data we needed. We would also like to thank Andrew Hales for his constructive comments and feedback. The preparation of this article was supported by grant number 016-125-385 from the Netherlands Organization for Scientific Research (NWO) awarded to Jelte M. Wicherts and by the NSF under Grant #BCS-1339160 awarded to Kipling D. Williams.

Ordinal Effects of Ostracism:
A Meta-Analysis of Cyberball Studies

Word count (main text): 9473 words

Word count (abstract): 159 words

Word count (references, tables, footnotes, figure notes): 4224 words

Abstract

Cyberball is widely used to study ostracism and related phenomena. We examine 120 studies (combined $N = 11,869$) to determine the overall effect size, and conditions under which the effect may be reversed, eliminated, or small. Our analyses showed that (1) the average ostracism effect is large ($d > |1|$) and (2) the structural aspects of the Cyberball game (i.e., number of players, gender composition of sample, origin of study, sample age, duration of ostracism, type of needs scale) are inconsequential. Further, we test a proposition by Williams (2009) that the immediate impact of ostracism is resistant to moderation, but that moderation is more likely to be observed in delayed measures. Our findings suggest that (3) both the first and last measure are moderated, and (4) time passed since being ostracized does not predict the effect sizes of the last measure. Thus, support for this proposition is tenuous, and we suggest modifications to the temporal need-threat model of ostracism.

Keywords: Cyberball, meta-analysis, ordinal, ostracism

Ordinal Effects of Ostracism: a Meta-Analysis of Cyberball Studies

Cyberball (Williams, Cheung, & Choi, 2000; Williams & Jarvis, 2006) is a virtual ball-tossing game that is used to manipulate the degree of social inclusion, or ostracism, in social psychological experiments. In this game the participant supposedly plays with two (or three) other participants, who are in fact part of the computer program. The program varies the degree to which the other players are passing participants the ball; ostracized participants are not passed the ball after two initial tosses, whereas included participants are passed the ball repeatedly. Figure 1 shows a still from the game. In the study of the psychological effects of ostracism and exclusion, this methodological paradigm has been widely used in parallel with other paradigms, such as the future life rejection (see Baumeister, Twenge, & Nuss, 2002), the get-acquainted paradigm (Nezlek, Kowalski, Leary, Blevings, & Holgate, 1997), and the autobiographical memory manipulation (i.e., remember a time when you were excluded; Craighead, Kimball, & Rehak, 1979). Our literature search showed that at least 200 published papers involved the use of the Cyberball paradigm to study ostracism, and that over 19,500 participants have played the game thus far. Thus, the Cyberball paradigm has received much traction in experimental studies on ostracism.

Historical background

Since its introduction, Cyberball research has informed theory on ostracism, exclusion, and rejection (e.g., Williams, 2009), has been the topic of several meta-analyses (Blackhart, Nelson, Knowles, & Baumeister, 2009; Cacioppo, Frum, Asp, Weiss, Lewis, & Cacioppo, 2013; Gerber & Wheeler, 2009), and has received growing interest even outside the domain of social psychology (Williams, 2009). Because everybody gets excluded sometimes (approximately once a day according to Nezlek et al., 1997), and because ostracism research is socially relevant in many domains (e.g., workplace, church, schools, families, and even in the context of school shootings; Leary, Kowalski, Smith, & Phillips, 2003). Through

experimental work, it has been repeatedly shown that being ostracized has an effect on people—either on their psychological functioning (e.g., decreases in positive mood; Lustenberger & Jagacinski, 2010) or on certain interpersonal behaviors (e.g., increases either social susceptibility or aggressive behaviors; Carter-Sowell, Chen, & Williams, 2008; Van Beest, Carter-Sowell, Van Dijk, & Williams, 2012). These experiments have highlighted the (mostly negative) impact of ostracism on fundamental needs (e.g., belonging; Baumeister, & Leary, 1995), mood, physiology (e.g., body temperature; IJzerman, Galucci, Pouw, Weißgerber, Van Doesum, & Williams, 2012), and various other constructs, including those measured with behavioral measures (e.g., conformity, compliance, aggression). In the current paper, we refer to the general effect of being ostracized compared to being included in Cyberball as the *ostracism effect*.

Williams (2009) proposed a temporal need-threat model of ostracism, in which he suggested three stages in the ostracism effect, namely: (1) a *reflexive* stage, (2) a *reflective* stage, and (3) a *resignation* stage. In the reflexive stage, the response to the ostracism sequence is immediate and occurs like a reflex. This initial response is theorized to be socially painful and threatening (Baumeister & Leary, 1995), and easily detectable due to evolutionary over-sensitivity to cues of ostracism (Haselton, & Buss, 2000). Such a reflex would not take into account situational specifics, and provides little room for coping. The reflex is proposed to affect primarily pain, fundamental needs, and emotional reactions (e.g., increased anger and sadness). The affected fundamental needs are (1) belonging, (2) self-esteem, (3) control, and (4) meaningful existence, measured by a need satisfaction scale (Williams, 2009). According to Williams, measures of reflexive responses must occur during, or in the case of self-report measures, immediately following Cyberball (with the wording of the questions referring to how participants felt *during the game*). The *reflective* (or delayed) stage, which follows this immediate response, is subject to more rational thought

and coping with the threats. Part of such coping is need fortification of the threatened fundamental needs. Coping can be measured both in terms of speed of recovery (higher levels of need satisfaction approaching the levels of included participants), and emotional, cognitive, and behavioral choices. The *resignation* stage occurs after prolonged ostracism, causing prolonged periods of pain and more fundamental need threat. If one is not able to fortify the fundamental needs, a prolonged ostracism sequence leads to feelings of helplessness, alienation, depression, and unworthiness. Because the resignation stage is hypothesized to occur only after prolonged and repeated exposure to ostracism (as in months or years), it is not feasible (and even unethical) to study resignation responses in laboratory experiments. Hence, in this paper we limit ourselves to studying the reflexive and reflective stages. For these stages, Williams asserts that moderation and variation of need satisfaction effects by individual differences and socially relevant factors (e.g., type of group from which one is excluded) will be less likely to occur for reflexive measures than for reflective measures.

Previous meta-analyses on the topic of social exclusion have focused on estimating the effect of exclusion¹ on different constructs (e.g., belonging, self-esteem, etc.; Gerber & Wheeler, 2009), neurophysiological effects in fMRI studies (Cacioppo et al., 2013), impact on affective responses (Blackhart et al., 2009), and moderation of (non-Cyberball) effects by types of manipulation (Blackhart et al., 2009). Gerber and Wheeler (2009) found that rejection has a medium to large effect on mood and fundamental needs, which results in either anti-social responses if control *can* be restored, or pro-social responses if control *cannot* be restored. Blackhart and colleagues (2009) found that exclusion moved individuals to a more neutral mood and/or emotional state, and that it negatively affected self-esteem. They also found indications of moderation by type of rejection manipulation used. Cacioppo et al. (2013) provided a more nuanced account of pain overlap theory (for reviews see

Eisenberger & Lieberman, 2004; Iannetti, Salomons, Moayedi, Mouraux, & Davis, 2013; MacDonald & Leary, 2005) by showing that social and physical pain activates similar but also distinct brain areas.

Whereas these meta-analyses focused on the social ostracism effect within different constructs (e.g., fundamental needs), in different exclusion paradigms, and in fMRI studies, the current meta-analysis limits the paradigm to that of between-subjects experiments using Cyberball and looks to test more general ideas of social ostracism. Here we focus on the workings of the reflexive and reflective stages (rather than on the resignation stage). Of all Cyberball studies, we included only those that used a between-subjects experimental design with random assignment to assure a minimum level of methodological quality. The criteria were also broad enough to include a sufficiently large number of studies for testing our hypotheses.

Goals of Meta-analysis

The goal of our meta-analysis is to provide a comprehensive understanding of the Cyberball-induced inclusion versus ostracism effect size. Under what conditions, if any, is the effect size negative, zero or especially small? Under what conditions is it especially large? Given that Cyberball is a popular experimental paradigm, it is worthwhile to examine whether specific changes in procedure matter in the resulting effect. Thus, to shed light on the generalizability of the ostracism effect, we assessed whether structural aspects of Cyberball studies moderate effects on the first and last measures taken during the experiment. We inspected whether (1) number of players in the game, (2) gender composition of the sample, (3) origin of study, (4) average age of the sample, (5) total number of ball tosses, (6) the number of throws in the game and (7) type of needs scale affected the found effect size.

To provide estimate these effects, we focused on the first and last reported measure of the selected experiments. That is, we did not include all potential measures of every study in the meta-analysis as this would make it very difficult to make meaningful comparisons across studies. Additionally, this allowed us to model the ostracism effect in an ordinal fashion. In Williams's (2009) temporal need-threat model, ostracism effects in the reflexive stage are more resistant to moderation than effects in the reflective stage, and coping diminishes effects only during the reflective stage. As reflexes typically occur immediately following the stimulus, this implies that on the last measure (compared to the first measure), the ostracism effect would be smaller, while *moderation* of the ostracism effect would be larger. These hypotheses were registered a priori on the Open Science Framework (OSF).²

Our operationalization of moderation of the ostracism effect is straightforward: many experimental designs included a second factor besides the factor related to ostracism, and so interactions in the ANOVA realm would indicate moderation of the ostracism effect. For example, in a 2 (ostracized vs. included) by 2 (in-group vs. out-group) between-subjects design, the ostracism effect could be expected to be larger for the in-group level than for the out-group level. Such moderation can be numerically seen as an interaction effect. For instance, the difference between simple effects of ostracism for the in- and out-group conditions reflects this interaction (specific calculations are reported in the methods section and formulae in the Appendix).

We will use meta-regressions with study-level indicators to study effects of the composition of the manipulation and the samples. This is conceptually similar to a (multiple) linear regression, but with estimated effects as the dependent variable. Our predictors for these meta-regressions are the study-level indicators. First, because collectivism might influence the degree to which belonging is important (see Hofstede, 1980), we used a crude categorization of continents (i.e., U.S., other western countries, Asian countries, and remaining countries).

Second, because social aspects may be less evolutionarily relevant for males than for females (Hawes et al., 2012), we included proportion of male participants. Third, given that exclusion may be more relevant for younger people, we included mean age. Fourth, because it may matter by how many people one gets ostracized, we included the number of players in the game. Fifth, as the length of the exclusion may matter, we included duration of exclusion. Sixth, the number of throws in the Cyberball game were taken as predictor, as more throws might exacerbate the ostracism effect. Seventh, we considered potential differences between types of scales used to measure fundamental needs.

Additionally, we inspected robustness of the findings across different subsets of dependent variables. Overall, the dependent variables included in the meta-analysis were only subject to the criterion that they were expected to be affected by ostracism, which does not limit the measures to just one type. In other words, we included multiple types of dependent variables with varying psychometric properties in the primary studies. We considered measures that speak to both how the participant interacts with others (i.e., interpersonal) and how they experience the situation themselves (i.e., intrapersonal). We define interpersonal measures as measures relating to others, and intrapersonal measures as measures relating only to the self. Examples of interpersonal measures are donation behavior, aggression. Examples of intrapersonal measures are self-reported anger, self-esteem, control, but also physiological measures such as body temperature, galvanic skin response. Finally, given that most Cyberball studies specifically use *fundamental needs* (i.e., belonging, self-esteem, control and meaningful existence) questionnaires, we also tested these as a separate type of intrapersonal measure (see Van Beest & Williams, 2006; Williams et al., 2000; Zadro, Williams, & Richardson, 2004). We also included other measures relating to the fundamental needs, such as the Rosenberg Self Esteem Scale. Overall, these fundamental

needs measures are particularly important for testing Williams's (2009) prediction concerning moderation of ostracism effects over time.

Besides using different subsets of measures for sensitivity analyses, we also coded whether the first- and last measure included was immediate (i.e., variables relating to *during* the game) or delayed (i.e., variables relating to *after* the game). This ensures model correspondence for the included measures.

In sum, the hypotheses are subdivided into two primary and several exploratory questions. The two main questions were (1) is there an ordinal decrease of the ostracism effect across time points? and (2) is there an ordinal difference in the interaction effect across time points? Additional questions were: do study characteristics affect the estimated average effect? In other words, are the results robust across different subsets that substantively could yield different results? These questions will be answered with random and mixed-effects meta-analytic models applied to 120 studies.

Method

Study inclusion criteria

First, experiments were required to have a factor that manipulated number of virtual ball tosses obtained by the participants. For this ostracism factor we only considered the condition in which participants were ostracized by all other participants and the condition in which participants were equally included by all other players. Studies that used other (between-subjects) factors alongside the ostracism factor were included as well. Limiting the study designs included is presumed to decrease variability due to design characteristics, which increases power for moderator analyses. Note that we collapsed effect sizes across irrelevant factors if primary authors expressed no expectations concerning the potential moderating effect of that crossed factor (i.e., non-moderating factors). Moreover, continuous variables that were dichotomized into factorial levels were also collapsed due to the many problems

dichotomization can cause (e.g., underestimation of effect size, spurious effects; see Hunter & Schmidt, 1990; MacCallum, Zhang, Preacher, & Rucker, 2002). For example, when participants were grouped into high- and low neuroticism groups based on a continuous measure of neuroticism (Boyes, & French, 2009), we used pooled means and standard deviations across these two groups, reducing the design to an ostracism/inclusion design.

Second, we only considered experiments that incorporated a between-subjects design with random assignment. Within-subject designs were excluded, because most within-subjects designs regard high-dimensional neurophysiological measurements such as fMRI that are beyond the scope of this meta-analysis (see Cacioppo et al., 2013). Also, meta-analyses of effects of within-subjects designs require the correlations between measures in primary studies, and we did not expect these to be reliably reported in the papers.

Reasons for these inclusion criteria are threefold. (1) Most Cyberball experiments take place in such a format, making it an encompassing criterion for the purposes of this meta-analysis. (2) The choice to limit the meta-analysis to between-subject designs rendered computational aspects more feasible based on reported statistics in papers. (3) These criteria guarantee a minimum standard of experimental rigor, which eliminates the need for subjective quality assessments of individual studies. For the dependent measures the criterion was that they were (expected to be) affected by the ostracism manipulation. We considered the measures that immediately followed the manipulation (first measure) and the measure at the end of the study (last measure), while excluding manipulation checks in this assessment.

Literature search

To have a comprehensive meta-analysis of Cyberball studies, we used seven search strategies in the period of November 2012 through April 2013. These search strategies included database searches, a call for data, cross-reference with Kip Williams's online list of

Cyberball studies, Google Scholar alerts, citation records, SPSP conference abstracts, and personal communications.

The databases searched included Web of Knowledge, PubMed, ScienceDirect and Worldcat using all sources from the Tilburg University library. The first three cover only published articles, whereas Worldcat also covers books and dissertations as well as the PsycINFO database. All these databases were searched with the keywords *cyberball*, *ball-tossing* and *ball AND ostraci**. Web of Knowledge was the first database searched. For this database, an additional search term (i.e., *ball AND exclu**) was used, but this yielded zero relevant hits on a total of 501 hits. Across all these searches, results included 1927 hits of which 109 were saved for coding. Within Web of Knowledge, we looked through all citation records of the seminal papers by Williams et al. (2000); Williams and Jarvis (2006). These papers were cited 332 times (as of 5th of November, 2012), of which 43 papers were saved for coding. The entire literature search provided 2259 initial hits (including possible duplicates across searches), of which 152 were selected to be included in the coding.

The call for data was put on the list servers or forums of Society for Personality and Social Psychology (SPSP), European Association of Social Psychology (EASP), and Social Psychology Network (SPN; all on 3rd of December, 2012). This resulted in nine replies, from which three useful studies, which were included in the coding procedure.

Kip Williams keeps a list of Cyberball studies on his website. This list was used to check for extra articles that did not turn up in the initial searches on November 15th, 2012.³ The list included 93 papers, of which 9 papers were included to be coded.

The final searches included Google Scholar alerts, SPSP conference abstracts and personal communication. The Google Scholar alerts were used to keep up to date with new literature. These alerts notify a user when new hits for a search term occur, and were used for *cyberball* and *ball-tossing*, which yielded 85 hits of which 25 were saved for coding. SPSP

conference abstracts from 2006 through 2013 were searched for Cyberball studies. This led to personal communications with the authors of the conference abstracts, leading to additional studies. Pooled, the personal communication and the conference abstracts yielded 21 hits, of which 20 were saved for coding. The seminal paper by Williams et al. (2000) was added separately.

In sum, the literature search spanned 2468 hits, resulting in 205 that were saved for coding. During coding, papers were assessed to fit the inclusion criteria. Of the 205 papers, 107 papers were excluded for a variety of reasons. Several involved the use of a within-subjects design (52 papers). Some papers could not be accessed (5 papers) or could not be included because we did not receive the required data on request (7 papers). Some were excluded for other reasons (43 papers), such as not involving new data (e.g., a dissertation study that was later published). All included papers were published between 2000 (after the introduction of Cyberball) and April 2013. This resulted in a final, fully coded sample of 98 papers containing 120 studies, with mean sample size 98.9 and median sample size 74.⁴ There were a total of 11,869 Cyberball participants.

Coding procedure

The first author coded all the studies and conducted all the analyses. The third author double-checked a subset of the entire database, while the second author double-checked all 52 studies that entailed a full two-by-two design. The third author checked and reran the R code of all analyses. Finally, an extensive account of all coding decisions is publicly available via Open Science Framework on a paper-by-paper basis (see Footnote 2 for the direct link).

Group means and standard deviations were retrieved for both the first and last relevant measure in each study for effect size calculation. Relevant measures were defined as constructs that were expected by primary authors to show an ostracism effect (e.g., fundamental needs, mood, pro-social helping behavior, etc.). Coding that was crucial for

testing the confirmatory hypotheses concerned the amount of items from the first through last measure plus any additional time in between (e.g., rest period). This made up the estimation of time from the first to last measure, where each item was counted as lasting six seconds (the six-second rule was based on a longstanding practice used to estimate average completion time in the freshmen testing program of the University of Amsterdam; e.g., Smits, Dolan, Vorst, Wicherts, & Timmerman, 2011). Any additional time reported in the procedure was also included. Note that some measures are variable on time (e.g., persistence tasks) and that these were arbitrarily estimated in a conservative manner to at least take these measures into account at some level.

The type of measure used was coded for in the following general terms: (1) interpersonal, (2) intrapersonal, and (3) fundamental needs. Interpersonal measures were defined as measuring constructs that relate to (the self and) others (e.g., *how angry do you feel towards person X?*, donations to charity, etc.). Intrapersonal measures were defined as measuring constructs that relate only to the self (e.g., *how angry do you feel?*, physiological measures, etc.). For the exploratory analyses, we coded sample characteristics (e.g., age, gender composition), Cyberball characteristics (e.g., amount of players, length of game), measure properties (e.g., intra- or interpersonal), and whether the first- and last measure fit the definition of immediate (i.e., during the game) or delayed (i.e., after the game/now), respectively.

Because relevant measures were defined broadly we included different kinds of measures that are expected to show different directions of an ostracism effect. For example, when compared to included participants, belongingness scores are expected to be lower for ostracized participants, whereas retaliation scores are expected to be higher for ostracized participants. To counteract computational problems (i.e., cancellation of effects) being caused by this bidirectionality of ostracism effects, we coded the direction of the ostracism effect for

each specific measure, such that negative effect sizes depict negative psychological effects. Moreover, in two-by-two designs in which the ostracism effect was crossed with another factor (i.e., a moderator), we coded for expected direction of that moderator. For example, in Table 1, we show hypothetical data for the four study designs that are possible when crossing direction of the effect and direction of the moderation. The relevant effect sizes should be corrected to attain comparable effect sizes across studies. Effect sizes for the simple ostracism effect (column wise) were corrected only for the type of measure. For instance, for panels (a) (involving, e.g., need threat) and (c) (involving, e.g., need satisfaction), the corrections entailed a multiplication with -1 or +1, respectively. Simple moderator effects (row wise comparisons) are interesting for understanding the effect of the moderator under either ostracism or inclusion. These simple moderator effects were corrected for both the type of measure *and* the expected moderation (i.e., exacerbation, -1, or minimization, +1). For example in panel (c), the 5 and 8 on the right are used to compute the *standard ostracism effect* (as in Williams et al., 2000), whereas the 3 and 8 in the left column represent an ostracism effect that is thought to be exacerbated. For example, in a given ostracism study with a two-by-two design, adolescents are expected to show stronger ostracism effects, compared to young adults (Pharo, Gross, Richardson, & Hayne, 2011). The 5 and 8 would subsequently represent the scores for the young adults, whereas the 3 and 8 would represent the scores for the young adolescents. In panel (d) we depict a study in which the *moderated* column is thought to lead to a minimal ostracism effect, as could be expected when Cyberball is played with members of a despised out-group (Gonsalkorale & Williams, 2007). The margins (greyed out) denote the simple effects, which are after correction comparable across all panels (a) through (d), indicating that this correction did what we intended it to.

Relevant information that was missing in the papers was requested from the authors via e-mail. In case of non-response, we sent three follow-up e-mails. All this communication

was documented and can be found on the OSF page for this project. In case of non-response or non-willingness to send data, studies were either eliminated if the information was crucial (i.e., means and standard deviations of the measures per group), computed if possible (i.e., cell sizes), or assumed if deemed reasonable on the basis of additional information. For instance, when no information was given we considered the Cyberball manipulation characteristics to be similar to previous studies in the same paper or in earlier papers referred to in the paper.

Statistical analyses

For the analyses, we used the *metafor* package (Viechtbauer, 2010) in the R statistical environment (R Core Team, 2013).

Effect size metric. We used Hedges's *g* version of the standardized mean differences as the effect size. Hedges' *g* corrects for the slightly biased estimate given by Cohen's *d* (Hedges, 1981). Standardized simple effects were calculated across the ostracism factor and the interaction effect was calculated by taking the standardized difference between the unstandardized main effects (see the Appendix for the exact formulae used). This was done for both the first and last dependent variable in each experiment. For example, in a 2 (ostracized vs. included) by 2 (moderator present vs. moderator absent) design with multiple measures, we calculated two simple ostracism effects (Hypothesis 1) and two interaction effects (Hypothesis 2). Non-factorial studies delivered only simple effects for the first and last measure and no interactions.

Meta-analytic model. We used random- and mixed-effects models, since heterogeneity in the effect sizes is expected due to both the inclusion of different measures and additional unknown methodological and substantive factors. The meta-regression element in some of the analyses is the variable time as predictor of the ostracism effect. Analyses without this study-level predictor reduce to a random-effects model. We used Restricted Maximum Likelihood (REML) to estimate tau-squared (i.e., the residual variance), as

recommended by Viechtbauer (2005). Note that when estimating a mixed- or random effects model, one does not estimate a single *true* effect, but rather the mean and variance of underlying effects (Viechtbauer, 2005).

Statistical sensitivity analyses. To test for robustness of the effects, we incorporated several statistical sensitivity analyses. We flagged possibly problematic outliers on the basis of studentized deleted residuals, Q-Q plots, and Cook's distance values. Subsequently, we inspected the effect of these outliers on substantial results in statistical sensitivity analyses in which these outliers were excluded. Another statistical sensitivity analysis entailed fitting of the mixed-effects model with tau-squared fit at the upper bound value of the 95% confidence interval.

Funnel plot asymmetry. A funnel plot depicts each study's effect size against its standard error (Light & Pillemer, 1984). Larger studies have smaller standard errors, and vice versa for smaller studies. Following from a theoretical fluctuation of the population effect size due to sampling variance, a funnel plot should be symmetrical around the estimated mean effect size. If there are no methodological or substantive reasons to expect a link between effect sizes and standard errors, funnel plot *asymmetry* can indicate publication bias (e.g., Bakker et al, 2012). To test funnel plot asymmetry, we used Egger's regression test (Egger, Smith, Schneider, & Minder, 1997) for mixed-effects models (Sterne & Egger, 2005). This tests whether the distribution of effect sizes is equal on both sides of the average effect, when accounting for true heterogeneity. Funnel plot asymmetry thus indicates bias in the estimated mean effect size, and possibly publication bias.

Results

In our reporting of the effect sizes, d indicates a main effect and Δd indicates an interaction effect. Even though we used Hedges's g , we maintained the notation of d , since g

is only a minor correction to Cohen's d . Statistical sensitivity analyses are only reported if they showed different effects (all statistical sensitivity analyses can be found on OSF).

Primary hypotheses

The two confirmatory hypotheses are tested in four meta-analyses, of which the study level effects are reported in Table 2. The table includes effect sizes used in the estimation of the average simple effect of ostracism on the first measure, the average simple effect on the last measure and the estimation of the average interaction effect on both the first and last measure.

Simple ostracism effect (Hypothesis 1). In a random-effects model on the main effect of ostracism ($k = 120$), residual heterogeneity was significant, $Q(119) = 1395, p < .001, I^2 = 92.99\%$ and estimated at $\tau^2 = 0.90, 95\% \text{ CI } [0.70, 1.24]$. The heterogeneity measure τ^2 includes both the estimated proportion of variance explained by moderators at the study level and unexplained variance in the distribution of underlying effect sizes (i.e., τ_{res}^2). The analysis yielded an estimated average effect of $d = -1.36, p < .001, 95\% \text{ CI } [-1.54, -1.18]$. A random-effects version of the Egger's test (Sterne & Egger, 2005) indicated funnel plot asymmetry, $Z = -6.14, p < .001$. Due to the size of the average effect, and hence large power to acquire significant outcomes in primary studies, we do not suspect publication bias to explain this asymmetry. In other words, immediately after being ostracized, the average ostracism effect is estimated at -1.36 standard deviation units, which entails a large effect (Cohen, 1988).

Next, we fitted a mixed-effects regression model for the ostracism effect on the last measure ($k = 95$), including estimated time in seconds since completing the Cyberball game as predictor. Residual heterogeneity was significant, $Q_E(93) = 803, p < .001$ and estimated at $\tau_{\text{res}}^2 = 0.38, 95\% \text{ CI } [0.27, 0.54]$. The intercept was estimated at $d_{\text{intercept}} = -0.76, p < .001, 95\% \text{ CI } [-0.91, -0.61]$. Moreover, the estimated time in seconds failed to moderate the average effect, $b = 0.0001, p = .187, 95\% \text{ CI } [-0.0001, 0.0003]$. However, we have to take into

consideration the low power of the moderation analyses due to the large (residual) heterogeneity in effect sizes (Hedges & Pigott, 2004). A regression test for mixed-effects model with moderator (i.e., including both the time and *SE* as predictor) showed no funnel plot asymmetry, $Z = -0.72$, $p = .474$. In short, long after ostracism has occurred ($M_{time} = 291.2$ seconds), ostracized participants on average scored around -0.73 standard deviation units lower when compared with included participants, an effect that does not appear to be moderated further by time passed since the ostracism occurrence.

Thus, results show a clear effect of ostracism on both the first and last measures, of which the latter is *not* predicted by our operationalization of time. The ostracism effect over time can also be inspected via confidence intervals. Comparing the 95% confidence intervals for the average ostracism effect on the first measure (i.e., [-1.54, -1.18]) and on the last measure (i.e., [-0.91, -0.61]) showed no overlap. Although the difference in average effect sizes between first and last measure cannot be formally tested (because of a lack of information on the correlation between measures in the primary studies), the mean difference is sizeable and CIs suggest that the average ostracism effect is smaller for the last measure, as expected. Given the expected positive correlation between effects for first and last measures, the comparison of CIs is likely to be conservative (Schenker & Gentleman, 2001). Additionally, we noted that estimated residual heterogeneity was larger on the first- than on the last measure. We conclude that the average ostracism effects decreases from the first- to last measures, and that study-level effects are more similar on the last measure.

Moderation of ostracism (Hypothesis 2). To test moderation of the ostracism effect, we selected the factorial experiments that manipulated ostracism and another independent variable in between-subjects designs. A random-effects model on the interaction effect (Δd) on the first measure ($k = 52$) showed heterogeneity in underlying effects, $Q(51) = 103.24$, $p < .001$, $I^2 = 50.60\%$ and an estimated $\tau^2 = 0.19$, 95% CI [0.07, 0.41]. The average interaction

effect equaled $\Delta d = -0.46$, $p < .001$, 95% CI [-0.64, -0.28], indicating a change in the ostracism effect due to the moderator level and vice versa (i.e., moderation of the ostracism effect). There was indication of funnel plot asymmetry in this analysis, $Z = -2.43$, $p = .015$. Thus, the data indicate that, across the board, the ostracism effect *can* be moderated on the first measure following the ostracism sequence, but it is possible that publication bias may have affected the interaction estimates.

On the last measure ($k = 46$), the mixed-effects model (with estimated time as predictor) for the interaction effect again showed residual heterogeneity, $Q_E(44) = 100.82$, $p < .001$ and estimated $\tau_{\text{res}}^2 = 0.21$, 95% CI [0.10, 0.55]. The intercept of the interaction effect was estimated at $\Delta d_{\text{intercept}} = -0.20$, $p = .052$, 95% CI [-0.402, 0.002] and no significant moderation of time was found, $b = 0.0002$, $p = .159$, 95% CI [-0.0001, 0.0004]. The regression test with the time and SE as predictors showed no funnel plot asymmetry, $Z = -0.68$, $p = .495$. These results indicate that moderation of the average ostracism effect is *not* found at a later time-point in the included studies, and time itself does not moderate the computed interaction effects. However, *statistical* sensitivity analyses showed that this interaction *was* significant when we removed three outliers based on studentized residuals, $\Delta d_{\text{intercept}} = -0.32$, $p = .029$, 95% CI [-0.60, -0.03], whereas the regression coefficient time continued to be non-significant, $b = 0.0002$, $p = .207$, 95% CI [-0.0001, 0.0006]. On the last measure, this indicates that the non-significant interaction effect is sensitive to outliers in the data.

To see whether the interaction effects decreased from the first to the last measure, we again compared confidence intervals. On the first measure, the 95% CI was [-0.64, -0.28] whereas for the last measure, the 95% CI was [-0.402, 0.002]. Considering the overlap of these CIs, there is not enough evidence for an average reduction in the moderation across the measures examined.

Exploratory analyses

In addition to the simple effects over all studies, we analyzed subsets of studies that differ in terms of measure type to study robustness of the effects. We also inspected whether sample composition, scale composition, and Cyberball specifics could predict the estimated effect size. Finally, we selected a homogeneous subset of studies to come to grips with the relatively large heterogeneity of simple main effects found in the confirmatory analyses.

Measures. To inspect the robustness of the estimates, we studied simple effects across several subsets of measures. These subsets encompassed interpersonal measures (i.e., measures that relate to others or the self in the context of others), intrapersonal measures (i.e., measures that relate only to the self), fundamental needs (single- and composite needs), and measures that were coded by the first two authors as fitting the description of being immediate or delayed (i.e., questions related to during- or after the game, respectively; shown in Figure 2 as *model*). We ran the analyses for the different measures for the two time points separately (i.e., first and last measure).

The different panels in Figure 2 show the results for the different simple effects per subset and overall; Table 3 summarizes the estimated interaction effects. A comparison of the results within each panel shows whether the overall results are robust and representative of all subsets, or whether there are nuances per type of measure. The main differences are notable in panels (1), (2) and (5). The first and second panels indicate that the effect of ostracism within both moderator levels is stronger for the subset of fundamental needs measures, and weaker for interpersonal measures. This indicates that in a similar factorial design, fundamental measures show stronger effects and interpersonal measures weaker effects. Panel 5 indicates that the moderation of interpersonal measures is stronger compared to the other subsets. This suggests that interpersonal measures are more subject to moderation, whereas the effects of ostracism on fundamental needs are larger initially. Additionally, for the subset of fundamental needs, we noted that the point estimated interactions (Table 3) follow the pattern

predicted by the need-threat model (Williams, 2009): i.e., the first measures are moderated less than the last measures. These sensitivity analyses indicate that the results concerning interactions are sensitive to measures being fundamental needs or interpersonal. Because fundamental needs showed effects in the theorized direction, we explored this further by overlapping the subset of fundamental need measures with the model definition of immediate and delayed (i.e., whether the measures related to feelings during or after the Cyberball game). Thus, we inspected the fundamental needs measures on both time points, which were coded as being truly immediate or delayed. Estimated interactions for this selection were $\Delta d = -0.37$, 95% CI [-0.60, -0.14] ($k = 29$) and $\Delta d = -0.13$, 95% CI [-0.53, 0.27] ($k = 8$) for the first and last measure, respectively. So in this particular subset of studies that use immediate or delayed fundamental needs measures, results are in line with Williams's (2009) prediction.⁵

Composition. We ran mixed-effects models on the ostracism effect (as in Hypothesis 1) inspecting for composition effects, on both the first and the last measure. The predictors in the mixed effects model were (1) country (US, other Western country, Asian, other), (2) proportion of males in the study, (3) mean age of the sample, (4) number of players in the game, (5) length of the game (≤ 5 min, 5-10 min or > 10 min), (6) the number of throws in the game and (7) type of needs scale referenced (by assigning unique values for every unique reference). This model ($k = 45$) showed clear residual heterogeneity, $Q_E(33) = 450$, $p < .001$, estimated $\tau_{\text{res}}^2 = 0.90$, 95% CI [0.54, 1.59], but no overall moderation, $Q_M(11) = 10.75$, $p = .465$. Inspecting the predictors individually also showed no indication for moderation ($ps > .137$; see Table 4). The different types of need scales (e.g., Van Beest & Williams, 2006; Williams, 2009; Zadro et al., 2004) did not significantly moderate effect sizes, showing psychometric convergence among the three scales. On the last measure ($k = 41$), no overall moderation was found, $Q_M(11) = 6.00$, $p = .873$, but the number of players in the game did significantly predict the effects, $b = 1.55$, $p = .047$, 95% CI [0.02; 3.07]. The significance of

this individual predictor should be interpreted carefully, as the omnibus moderation test showed no systematic decrease in heterogeneity. We found no indication for such moderation due to study composition.

Homogeneity? The analysis of the simple ostracism effect on the first measure showed that differences of underlying effects made up 93% of the variability in study outcomes. We performed an additional exploratory analysis in a more homogenous subset of studies to better understand this heterogeneity. This subset only included typical Cyberball studies that involved three players in the game, 30 throws, and lasted less than five minutes. In addition, the homogeneous subset of typical Cyberball studies only involved measures of immediate fundamental needs (single or composite). Performing a meta-analysis on this homogeneous subset of 19 studies showed an I^2 value of 83%, indicating that 83% of the total variability is ascribable to heterogeneity in the effect sizes. We noted that the mean simple ostracism effect in these 19 studies was relatively strong and estimated at $d = -2.05$, 95% CI [-2.44, -1.65]. In other words, the heterogeneity found in the overall analyses does not appear to be an artifact from the inclusion of different measures and the use of alternative Cyberball setups.

In sum, the exploratory analyses indicate several things. First, the effect of ostracism appears relatively large for fundamental needs, and somewhat weaker for the interpersonal measures. Second, interpersonal measures are also more prone to moderation of the ostracism effect. Third, the estimated interaction effect showed more moderation on the first measure, when compared to the second. The subset of fundamental needs showed interaction effects in line with the predictions of the theoretical model. Considering the relatively small size of the interaction effects, more powerful studies are needed to address this core prediction of the need-threat model (Williams, 2009). Fourth, there is substantial heterogeneity in the effect

sizes, even when considering a homogeneous subset of studies. This suggests that responses to ostracism are more variable than they have been construed until now.

Discussion

In this comprehensive meta-analysis of experimental ostracism studies with the Cyberball game, we focused on two confirmatory hypotheses based on theory as well as several exploratory hypotheses. Our results showed that the ostracism effect is quite large on average, but that it varies according to different factors. For our confirmatory hypotheses, two questions were central: (1) *is there an ordinal decrease of the ostracism effect across time points?* and (2) *is there an ordinal difference in the interaction effect across time points?* The results indicated a decrease in the effect from the first to the last measure ($d = -1.36$ and $d = -.76$, respectively), but that this is not moderated by the estimated time between first and last measure. Additionally, our analyses also showed an ordinal increase in the variability of the ostracism effect, where variability was larger on the first measure ($\tau^2 = 0.90$, 95% CI [0.70, 1.24]) then on the last measure ($\tau^2 = 0.38$, 95% CI [0.27, 0.54]; see Figure 3 for visual depiction). Because reflexes are expected to be fairly homogeneous, one would expect relatively less study-level variability (instead of more) at a reflexive time point, compared to a reflective time point. So this difference in variability may suggest that the simple ostracism effect is less of a reflex than has been theorized previously (Williams, 2009). Nevertheless, the reduction in the effect size itself is in accordance Williams's (2009) theory.

The expected interaction effect was found on both the first and last measure, but on both occasions the effect was relatively weak. Simple effects indicated that, on average, the overall ostracism effect operates similarly on both levels of the moderator factor and that both effects are relatively large. At the same time, the moderator factor in the primary studies showed a mean positive effect within the ostracism level and a mean negative effect within the inclusion level (see Figure 2). Both these effects were relatively weak ($d = 0.31$ and $d = -$

0.16, respectively). To understand the implication of this overall pattern it may be fruitful to consider an example in which participants are ostracized or included by either an outgroup or an ingroup. In such a setting, our findings would thus suggest that the relative effect of ostracism compared to inclusion, is similar for both outgroup *and* ingroup conditions. However, if one compares the effect of outgroup and ingroup, one would predict that those ostracized by outgroup members would slightly benefit whereas those included by ingroup members would slightly be harmed.

Structural Aspects of Cyberball and Different Dependent Variables

Cyberball studies can vary widely — they can differ in the amount of players, how the players are portrayed, etc. We investigated several structural elements and analyzed whether these reliably predicted the found effect size. These are summarized in Table 3. We investigated six specific elements (i.e., culture/country, proportion of males, sample age, number of players, length of exclusion session, and need scale used) and found no indication that any of these reliably predicted the effect size found in the studies. We collate and provide our analyses here, but other potentially relevant questions (e.g., does composition of other players matter?) warrant further research.

Does number of players matter? In the studies included in this meta-analysis, approximately 89% of the studies used the three-player version of Cyberball and 11% used the four-player version of Cyberball. Average ostracism effects did not differ between these two types of setups.

Does gender of participants matter? Results indicated that, contrary to earlier findings that study the effects of gender directly (Hawes et al., 2012), proportions of males and females did not significantly predict the mean effect size. In our coded studies, the mean proportion of males was approximately 39%.

Does culture or country matter? We found no indication that culture predicted the average effect size. In our coded studies, approximately 52% were from the United States, 45% from other Western countries (e.g., Australia, the Netherlands, Germany) and 3% from Asian countries. Our analyses used the United States as reference category. We note that the low prevalence of Asian countries might cause a lack of power, and that we cannot definitively state there is no difference between Western and Asian responses to ostracism. We *can* state that there is no systematic difference in the ostracism response for Western countries and the United States.

Does age of participants matter? We failed to find moderation of ostracism effects by mean age of the study samples. Coded studies had a mean sample age ranging from 10 through 32.5 years, with an average of approximately 20.5 years. This suggests that most of the research with Cyberball is done on young adults, with relatively few or no studies investigating children, middle-aged participants, or senior citizens. More research could focus on specific (individual-level) age moderation of ostracism.

Does number of throws matter? Of the coded studies, 60% used 30 throws, 11% used 40 throws, 8% used 20 throws, 4% used 60 throws, and 2% for both 15 and 24 throws. Other categories ranging from 10 through 200 make up the remaining percentages, each making up 1%. Our results indicated the mean ostracism effect was *not* predicted to be different across the different amounts of throws.

Does type of dependent variable matter? Exploratory analyses also showed that the majority of the results were robust across subsets of dependent measures and the overall set of dependent measures (see Figure 2). Exceptions were interpersonal measures showing relatively weak ostracism effects, while fundamental need measures showed somewhat stronger ostracism effects on the first measure. This suggests that psychological effects of ostracism are large, but that this effect is smaller for interpersonal behaviors. On top of this,

interpersonal measures also show more moderation, suggesting that interpersonal behaviors caused by ostracism are more easily moderated. Additionally, we estimated interactions for the measure subsets interpersonal (i.e., measures relating to others), intrapersonal (measures relating to the self), fundamental needs, model (i.e., first measure is reflexive and last measure is reflective) and an overlap of the latter two subsets. For all but two, these subsets showed that measures taken at the first time point were moderated more than the measures taken last. However, the analyses including only fundamental needs showed that moderation was larger at the last time point, when compared to the first time point. This result is crucial, as Williams (2009) specifically predicted this pattern for fundamental needs. Our overall test of ordinal effects of moderation that included many different types of measures failed to show the predicted pattern. Hence, we can reliably state there are interactions on both time points, but cannot make any general conclusions as to how they relate, as results are crucially dependent on the measures taken. In sum, exploratory analyses showed the effects of ostracism and its moderation are robust, except for three deviations, where interpersonal measures show smaller- and fundamental needs larger ostracism effects, and fundamental needs also show less moderation on the first time point than on the last.

Williams's Model of Ostracism: Supported or Not?

Regarding the test of Williams's (2009) model, there are several important observations and limitations. First, Williams proposed fundamental need threat as a result of even a brief episode of ostracism. This was supported by the meta-analysis. The model asserts that negative emotional reactions (i.e., sadness and anger) are also induced by ostracism, and this proposition was contested by Blackhart et al.'s (2009) meta-analysis in which they argued for affective numbness. We did not explicitly test this in the present analysis. That moderation is predicted to occur in the reflective stage, when the context and meaning of the ostracism event can be appraised, is tested and supported by the present meta-analysis. The final stage of

Williams's model—resignation—is outside the aims of the present meta-analysis, because it requires long-term exposure to ostracism. Thus, most propositions set forth in Williams's model that were tested within this meta-analysis, were supported.

The proposition that appears to lack support from this meta-analysis is that reflexive reactions to ostracism are more resistant to moderation than reflective reactions. Across the board, our results indicate there is more moderation on the first time point than on the last time point. However, there are two limitations to this conclusion. Firstly, Williams specifically refers to physiological, online, or immediate retrospective reports to assess reflexive reactions. In many instances in this meta-analysis, the *first* reaction is not isomorphic with reflexive measures. Anything taken after the game, or assessed by wording indicating present state (rather than the participants' state during the game), is not assumed to be reflexive, nor predicted to be resistant to moderation. Secondly, Williams's proposition is restricted to fundamental needs only. Indeed, our specific analyses involving only studies that employed measures of immediate and delayed fundamental need satisfaction corroborated the model prediction that there is more moderation on the last time point, than on the first time point.

Because of this quantitative difference in moderation across measures, we encourage direct testing of this time difference in moderation as predicted by Williams (2009), just as Bernstein and Claypool (2012) was a direct, experimental test of a finding by Gerber and Wheeler (2009). Using our estimated interaction effects to determine sample size under a power of .8, a sample size of 2186 would be necessary to have sufficient power on both time points.⁶ Note that the mean sample size in full factorial designs in our meta-analysis is 110, showing that the mean post-hoc power in these studies is .08 to detect an *interaction* at the last time point (notably, power for the standard ostracism effect is highly sufficient in the included studies, due to the large effect). One possibility is for an interuniversity consortium to conduct

such a study. Based on our estimates, such a powerful study is a necessity to learn more about how moderation between time points compare.

Changes to the need-threat model of ostracism in reaction to the results of our meta-analysis. As a result of our findings, we suggest that the temporal need-threat model of ostracism should be modified. Firstly, it should be recognized that there is potential for moderation in Stage 1 in which researchers take immediate measures of impact that tap into participants' reactions during the game. If factors can reduce physical pain and distress, like for instance, acetaminophen (DeWall et al, 2010)⁷ or transcranial magnetic stimulation (Riva et al., 2012, or if certain populations are less likely to feel pain (e.g., those higher in psychopathy; Wirth et al, 2010), then we would also expect moderation of immediate measures of distress. Secondly, *how long after the ostracism episode occurred* is likely to be crucial as to whether they show more or less moderation. For example, if researchers wait long enough before administering the delayed need satisfaction measures, it becomes more likely that *all* participants will have recovered from the negative impact of ostracism, thus resulting in a homogeneous (and highly satisfied) between-group result. The opportunity to discover moderation has, essentially, passed, leaving a very homogenous self-report of high need-satisfaction. Thus, differences in recovery from ostracism based upon social-situational factors and/or personality differences, if any, occur somewhere between initial pain and final recovery. It is difficult to predict exactly when that time period is. Zadro et al. (2006) report delayed recovery by those high in social anxiety 45-minutes later. Other studies show full recovery within 5-10 minutes. Future research needs to examine the time course more carefully, to determine if and when moderation occurs in delayed measures.

Limitations

Within the current meta-analysis there are several limitations. First, our test of differences between the first and last measure was indirect. In its current setting, the meta-

analysis makes comparisons between the first and last measures based upon the confidence intervals of these estimates. This is an indirect and informal test of whether the effects differ. A direct test would provide more conclusive evidence on whether or not the effect is equal across the first and last measurements. However, such a direct test requires correlations between the measurements per study, per cell, which are (usually) not reported in papers. This would thus require a direct request for data from each paper, which would possibly yield low response rates (Wicherts, Borsboom, Kats, & Molenaar, 2006), lowering the sample size of the meta-analysis overall.⁸ This lack of direct testing was thus chosen as a way of retaining sample size within the meta-analysis.

Second, not all measures were included and tested in a repeated-measures meta-analysis. Initially, a pre-test was run including all measures, but this showed that many papers did not include all statistics required for all measures. Requesting all of this information from the authors yielded a limitation that was similar to the first: a trade-off between retaining a sufficiently large set of studies and comprehensiveness. Another reason for only including the first and last measures was that every measure would require two separate meta-analyses to test both the main- and interaction effect (increasing Type I error rates) if a similar analytical model was used. If all measures were included, it would increase the importance of including a statistical correction due to correlations between measures, to facilitate repeated-measures analyses to minimize Type I error rates. In other words, the failure to include all measures was to prevent the problem of multiple testing and nonresponse to data requests, which would lead to a smaller set of useful studies and hence less powerful analyses.

Third, random (non-systematic) heterogeneity in the effect sizes poses a problem for the power of finding moderator effects (Hedges & Pigott, 2004). This could pose the problem that several of the non-effects found are actually there, but not detected (Type II errors). However, the subset of typical Cyberball studies still showed substantial variability in the

effect sizes: $I^2 = 83\%$. This indicates that the effects are quite variable to begin with, and makes it unlikely that the effects are misrepresented.

Additionally, the specific null-effect of time as a predictor could be due to one of three reasons. First, the (random) heterogeneity in the effect sizes was too large to find moderation by time. Second, imprecise reporting of the measures in the papers led to inaccurate time estimations. Third, the difference in the effect size was not due to time but differences between the types of measures administered at the different time points. For the imprecise reporting of the measures, authors could be contacted, but this also poses new problems (i.e., nonresponse, or authors might not be willing to admit that measures were left out in the paper; LeBel et al., 2013). The difference in measures can be inspected by creating a difference index between the types of measures and regressing the effect sizes on that index. Doing this for the standard ostracism effect on the last measure, showed no significant predictive effect of this difference ($b = -0.03$, $p = .531$), indicating that the effect is not driven by difference in measures on the first and last time point. In short, there are some limitations of the analyses with time as a moderator, but these limitations are either hard to address (i.e., imprecise reporting or heterogeneity), or the data indicates the opposite (i.e., difference in measures). Inspecting whether the types of measures used across all studies are different, and not the difference within a study, shows that these are similarly distributed across time-points (maximum discrepancy of 4.9 percentage points). Substantive differences in proportions of measures across time points are minimal, and form an unlikely driving force for our findings. In sum, we conclude that the findings are not an artifact of selecting the first and last measures.

Fourth, the current meta-analysis only examined between-subjects designs. Possibly there is a difference for the ostracism effect in between- and within-subjects designs, something that we have not directly investigated. Also, the within-subjects designs often used

fMRI data or other physiological data such as EEG (27 out of 49 at least), which pose an interesting avenue for further research in a meta-analytic domain of neurophysiological measures to add to the work of Cacioppo et al. (2013) within the physiological framework. These references can easily be retrieved from the database of examined papers, as is available on the OSF page of this paper. Additionally, the fact that we only included between-subjects designs did not allow for inspection of the moderating effects of personality, as these are continuous measures that should be modeled linearly instead of in a factorial design, meta-analysis

A final note is that this paper only summarized the results of the measures included in the studies. However obvious this might be, it should be pointed out, because the validity of the conclusions are reliant on the validity of the measures. Most prominently represented in the current meta-analysis are the fundamental need measures, which have no proper psychometric validation up-to-date, notwithstanding their wide use. Other kinds of included measures possibly have the same, and one has been openly criticized (e.g., the Hot Sauce aggression paradigm; Ritter & Eslea, 2005). We note that results in this paper are conditional on that these measures *are* valid.

Conclusion

Our results show that the ostracism effect is large, with a mean effect size approaching 1.5 standard deviation units. Moreover, the results show that the impact of ostracism decreases from first to last measure. We estimated a relatively weak interaction effect, which has been studied in underpowered settings until now. More powerful studies are required to reach more direct and definitive conclusions, with respect to these interactions. In addition, the large heterogeneity in the study effect sizes (even in a homogeneous subset of studies) highlights that there are more potentially relevant moderators of ostracism in need of further study, and raises the question whether the immediate ostracism effect (as operationalized in

our analyses) is actually as reflexive as has been previously proposed. These findings support and extend the need-threat model (Williams, 2009), which has played a major role in ostracism research. We invite fellow researchers to think and test ideas that might provide some explanation for these findings, including reanalysis of our data.

References

References marked with an asterisk indicate studies included in the meta-analysis.

- Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, 7, 543-554. doi: 10.1177/1745691612459060
- Baumeister, R. F., & Leary, M. R. (1995). The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, 117, 497-529. doi: 10.1037/0033-2909.117.3.497
- Baumeister, R. F., Twenge, J. M., & Nuss, C. K. (2002). Effects of social exclusion on cognitive processes: Anticipated aloneness reduces intelligent thought. *Journal of Personality and Social Psychology*, 83, 817-827. doi: 10.1037//0022-3514.83.4.817
- Bernstein, M. J., & Claypool, H. M. (2012). Not all social exclusions are created equal: Emotional distress following social exclusion is moderated by exclusion paradigm. *Social Influence*, 7, 113-130. doi: 10.1080/15534510.2012.664326
- Blackhart, G. C., Nelson, B. C., Knowles, M. L., & Baumeister, R. F. (2009). Rejection elicits emotional reactions but neither causes immediate distress nor lowers self-esteem: a meta-analytic review of 192 studies on social exclusion. *Personality and Social Psychology Review*, 13, 269-309. doi: 10.1177/1088868309346065
- Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed.). New York, NY: Russell Sage Foundation.

- *Boyes, M. E., & French, D. J. (2009). Having a Cyberball: Using a ball-throwing game as an experimental social stressor to examine the relationship between neuroticism and coping. *Personality and Individual Differences*, 47, 396–401. doi: 10.1016/j.paid.2009.04.005
- Cacioppo, S., Frum, C., Asp, E., Weiss, R. M., Lewis, J. W., & Cacioppo, J. T. (2013). A Quantitative Meta-Analysis of Functional Imaging Studies of Social Rejection. *Scientific Reports*, 3. doi: 10.1038/srep02027
- *Carter - Sowell, A. R., Chen, Z., & Williams, K. D. (2008). Ostracism increases social susceptibility. *Social Influence*, 3, 143-153. doi: 10.1080/15534510802204868
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd Ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Craighead, W. E., Kimball, W. H., & Rehak, P. J. (1979). Mood changes, physiological responses, and self-statements during social rejection imagery. *Journal of Consulting and Clinical Psychology*, 47, 385-396. doi: 10.1037/0022-006X.47.2.385
- DeWall, C. N., MacDonald, G., Webster, G. D., Masten, C. L., Baumeister, R. F., Powell, C., ... & Eisenberger, N. I. (2010). Acetaminophen reduces social pain behavioral and neural evidence. *Psychological Science*, 21, 931-937. doi: 10.1177/0956797610374741
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 315, 629–634. doi: 10.1136/bmj.315.7109.629
- Eisenberger, N. I., & Lieberman, M. D. (2004). Why rejection hurts: a common neural alarm system for physical and social pain. *Trends in Cognitive Sciences*, 8, 294-300. doi: 10.1016/j.tics.2004.05.010

Gerber, J., & Wheeler, L. (2009). On Being Rejected: A Meta-Analysis of Experimental Research on Rejection. *Perspectives on Psychological Science*, 4, 468–488. doi: 10.1111/j.1745-6924.2009.01158.x

*Gonsalkorale, K., & Williams, K. D. (2007). The KKK won't let me play: ostracism even by a despised outgroup hurts. *European Journal of Social Psychology*, 37, 1176–1186. doi: 10.1002/ejsp.392

*Goodwin, S. A., Williams, K. D., & Carter-Sowell, A. R. (2010). The psychological sting of stigma: The costs of attributing ostracism to racism. *Journal of Experimental Social Psychology*, 46, 612–618. doi: 10.1016/j.jesp.2010.02.002

Haselton, M. G., & Buss, D. M. (2000). Error management theory: a new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology*, 78, 81. doi: 10.1037/0022-3514.78.1.81

*Hawes, D. J., Zadro, L., Fink, E., Richardson, R., O'Moore, K., Griffiths, B., ... Williams, K. D. (2012). The effects of peer ostracism on children's cognitive processes. *European Journal of Developmental Psychology*, 9, 599–613. doi: 10.1080/17405629.2011.638815

Hedges, L.V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational and Behavioral Statistics*, 6, 107–128. doi: 10.3102/10769986006002107

Hedges, L. V., & Pigott, T. D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods*, 9, 426–445. doi: 10.1037/1082-989X.9.4.426

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. London, UK: Sage.

- Hunter, J. E., & Schmidt, F. L. (1990). Dichotomization of Continuous Variables: The Implications for Meta-Analysis. *Journal of Applied Psychology*, 75, 334–349. doi: 10.1037/0021-9010.75.3.334
- Iannetti, G. D., Salomons, T. V., Moayedi, M., Mouraux, A., & Davis, K. D. (2013). Broken hearts and broken bones: contrasting mechanisms of social and physical pain. *Trends in Cognitive Sciences*, 17, 371-378. doi: 10.1016/j.tics.2013.06.002
- *IJzerman, H., Gallucci, M., Pouw, W. T., Weißgerber, S. C., Van Doesum, N. J., & Williams, K. D. (2012). Cold-blooded loneliness: Social exclusion leads to lower skin temperatures. *Acta Psychologica*, 140, 283-288. doi: 10.1016/j.actpsy.2012.05.002
- Leary, M. R., Kowalski, R. M., Smith, L., & Phillips, S. (2003). Teasing, rejection, and violence: Case studies of the school shootings. *Aggressive Behavior*, 29, 202–214. doi: 10.1002/ab.10061
- LeBel, E. P., Borsboom, D., Giner-Sorolla, R., Hasselman, F., Peters, K. R., Ratliff, K. A., & Smith, C. T. (2013). PsychDisclosure.org: Grassroots Support for Reforming Reporting Standards in Psychology. *Perspectives on Psychological Science*, 8, 424–432. doi: 10.1177/1745691613491437
- Light, R. J., & Pillemer, D. B. (1984). *Summing up: the science of reviewing research*. Cambridge, MA: Harvard University Press.
- *Lustenberger, D. E., & Jagacinski, C. M. (2010). Exploring the Effects of Ostracism on Performance and Intrinsic Motivation. *Human Performance*, 23, 283–304. doi: 10.1080/08959285.2010.501046

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7, 19–40. doi: 10.1037//1082-989X.7.1.19

MacDonald, G., & Leary, M. R. (2005). Why does social exclusion hurt? The relationship between social and physical pain. *Psychological Bulletin*, 131, 202–223. doi: 10.1037/0033-2909.131.2.202

Nezlek, J. B., Kowalski, R. M., Leary, M. R., Blevins, T., & Holgate, S. (1997). Personality moderators of reactions to interpersonal rejection: Depression and trait self-esteem. *Personality and Social Psychology Bulletin*, 23, 1235–1244. doi: 10.1177/01461672972312001

Oaten, M., Williams, K. D., Jones, A., & Zadro, L. (2008). The effects of ostracism on self-regulation in the socially anxious. *Journal of Social and Clinical Psychology*, 27, 471–504. doi: 10.1521/jscp.2008.27.5.471

*Pharo, H., Gross, J., Richardson, R., & Hayne, H. (2011). Age-related changes in the effect of ostracism. *Social Influence*, 6, 22–38. doi: 10.1080/15534510.2010.525852

R Core Team. (2013). R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from <http://www.r-project.org/>

Ritter, D., & Eslea, M. (2005). Hot Sauce, toy guns, and graffiti: A critical account of current laboratory aggression paradigms. *Aggressive Behavior*, 31, 407–419. doi: 10.1002/ab.20066

Riva, P., Lauro, L. J. R., DeWall, C. N., & Bushman, B. J. (2012). Buffer the Pain Away
Stimulating the Right Ventrolateral Prefrontal Cortex Reduces Pain Following Social
Exclusion. *Psychological Science*, 23, 1473-1475. doi: 10.1177/0956797612450894

*Schaafsma, J., & Williams, K. D. (2012). Exclusion, intergroup hostility, and religious
fundamentalism. *Journal of Experimental Social Psychology*, 48, 829–837. doi:
10.1016/j.jesp.2012.02.015

Schenker, N., & Gentleman, J. F. (2001). On judging the significance of differences by
examining the overlap between confidence intervals. *The American Statistician*, 55, 182-
186. doi: 10.1198/000313001317097960

Smits, I. A. M., Dolan, C. V, Vorst, H., Wicherts, J. M., & Timmerman, M. E. (2011). Cohort
differences in Big Five personality factors over a period of 25 years. *Journal of
Personality and Social Psychology*, 100, 1124–1138. doi: 10.1037/a0022874

Sterne, J. A. C., & Egger, M. (2005). Regression Methods to Detect Publication and Other
Bias in Meta-Analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.),
Publication bias in meta-analysis. Chichester: John Wiley & Sons.

*Van Beest, I., Carter-Sowell, A. R., van Dijk, E., & Williams, K. D. (2012). Groups being
ostracized by groups: Is the pain shared, is recovery quicker, and are groups more likely
to be aggressive? *Group Dynamics: Theory, Research, and Practice*, 16, 241–254. doi:
10.1037/a0030104

*Van Beest, I., & Williams, K. D. (2006). When inclusion costs and ostracism pays, ostracism
still hurts. *Journal of Personality and Social Psychology*, 91, 918–928. doi:
10.1037/0022-3514.91.5.918

Viechtbauer, W. (2005). Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model. *Journal of Educational and Behavioral Statistics*, 30, 261–293. doi: 10.3102/10769986030003261

Viechtbauer, Wolfgang. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36, 1–48.

Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *The American Psychologist*, 61, 726–8. doi: 10.1037/0003-066X.61.7.726

*Wirth, J. H., Lynam, D. R., & Williams, K. D. (2010). When social pain is not automatic: Personality disorder traits buffer ostracism's immediate negative impact. *Journal of Research in Personality*, 44, 397–401. doi: 10.1016/j.jrp.2010.03.001

*Williams, K D, Cheung, C. K., & Choi, W. (2000). Cyberostracism: effects of being ignored over the Internet. *Journal of Personality and Social Psychology*, 79, 748–62. doi: 10.1037/0022-3514.79.5.748

Williams, K.D. (2009). Ostracism: a temporal need-threat model. *Advances in Experimental Social Psychology*, 41, 275–314. doi: 10.1016/S0065-2601(08)00406-1

Williams, Kipling D., & Jarvis, B. (2006). Cyberball: A program for use in research on interpersonal ostracism and acceptance. *Behavior Research Methods*, 38, 174–180. doi: 10.3758/BF03192765

*Wirth, J. H., & Williams, K. D. (2009). 'They Don't Like Our Kind': Consequences of Being Ostracized While Possessing a Group Membership. *Group Processes & Intergroup Relations*, 12, 111–127. doi: 10.1177/1368430208098780

*Zadro, L., Boland, C., & Richardson, R. (2006). How long does it last? The persistence of the effects of ostracism in the socially anxious. *Journal of Experimental Social Psychology, 42*, 692–697. doi: 10.1016/j.jesp.2005.10.007

*Zadro, L., Williams, K. D., & Richardson, R. (2004). How low can you go? Ostracism by a computer is sufficient to lower self-reported levels of belonging, control, self-esteem, and meaningful existence. *Journal of Experimental Social Psychology, 40*, 560–567. doi: 10.1016/j.jesp.2003.11.006

Footnotes

1. Note that exclusion encompasses both social rejection and social ostracism.
2. The direct link: <https://osf.io/ht25n/>
3. It has been updated since, but the list that was used can be found on the Open Science Framework, see Footnote 2.
4. Oaten, Williams, Jones and Zadro (2008) was applicable, but was excluded due to being an outlier with respect to effect size ($ds > 15$). See also Gerber and Wheeler (2009; p. 473): “*One study (Oaten, Williams, Jones, & Zadro, 2007) had need effect sizes that were clear outliers (effect sizes were 5–7 standard deviations above the means) [...and...] were excluded from the analyses.*”
5. The reported fundamental need selection can be specified even further to only include studies that explicitly focus on composite need satisfaction as typically defined by Kip Williams. Such a selection again provides support for the hypothesis that immediate fundamental need satisfaction is less moderated, $\Delta d = -0.18$, 95% CI [-0.47, -0.11] ($k = 15$), than delayed need satisfaction, $\Delta d = -0.93$, 95% CI [-1.67, -0.19] ($k = 3$). Note, however, that such a selection is based on 3 studies for delayed measures.
6. We used G*Power 3.1.7 to calculate this between-subjects interaction effect (F -test, fixed effects, .8 power); with $k = 4$ and the smaller interaction (last time point; numerator $df = k - 1$). The effect size Δd was transformed in to f by means of $\sqrt{[d^2/(2k)]}$, resulting in $f = .0707$.
7. DeWall et al. (2007) was not included in the meta-analysis, because we were not able to retrieve all information.
8. Note that out of the 72 data requests, we received timely replies of 52 (i.e., ~72%). However, these requests were only for specific information and not for raw datasets, as was the case in Wicherts et al. (2006).

Table 1

(a) Negative moderator, negative measure					(b) Positive moderator, negative measure						
		Moderated	Not-moderated/control	Raw	Correct			Moderated	Not-moderated/control	Raw	Correct
Ostracism factor	Ostracism	13	11	2	2	Ostracism factor	Ostracism	9	11	-2	2
	Inclusion	8	8	0	0		Inclusion	8	8	0	0
	Raw	5	3				Raw	1	3		
	Correct	-5	-3				Correct	-1	-3		
(c) Negative moderator, positive measure					(d) Positive moderator, positive measure						
		Moderated	Not-moderated/control	Raw	Correct			Moderated	Not-moderated/control	Raw	Correct
Ostracism factor	Ostracism	3	5	-2	2	Ostracism factor	Ostracism	7	5	2	2
	Inclusion	8	8	0	0		Inclusion	8	8	0	0
	Raw	-5	-3				Raw	-1	-3		
	Correct	-5	-3				Correct	-1	-3		

Hypothetical data example of coding correction

Note. Raw denotes the simple effect in the hypothetical data before correction whereas correct denotes the simple effect after correction. Column wise effects are multiplied by the type of measure only, whereas column wise effects are multiplied by both the type of moderator and type of measure.

Table 2

Effect sizes per study for the confirmatory hypotheses

First author	Year	<i>N</i>	<i>d</i> T1	(<i>SE</i>)	<i>d</i> T2	(<i>SE</i>)	Δd T1	(<i>SE</i>)	Δd T2	(<i>SE</i>)
Alvares	2010	74	-1.21	0.12	-0.10	0.10	-0.15	0.24	1.12	0.23
Ambrosini	2013	40	-1.69	0.13	-0.97	0.11	-	-	-	-
Anonymous	-	49	-3.97	0.24	-1.32	0.10	-	-	-	-
Anonymous	-	91	-3.17	0.20	-0.48	0.09	0.75	0.56	0.53	0.18
Aydin	2012	68	-0.95	0.13	-0.40	0.12	-1.19	0.24	0.72	0.23
Banki	2012	89	-1.87	0.07	-0.35	0.05	-	-	-	-
Bastian	2010	72	-2.75	0.11	-1.42	0.07	-	-	-	-
Bernstein	2012	24	-0.41	0.16	-	-	-	-	-	-
Bernstein	2012	25.50	-1.04	0.17	-	-	-	-	-	-
Bernstein	2010	73	-1.63	0.16	-1.63	0.16	-0.86	0.37	-1.11	0.40
Bernstein	2010	138	-2.67	0.10	-1.96	0.08	-0.53	0.22	-0.51	0.17
Bernstein	2012	67	-2.00	0.17	-0.99	0.13	-1.07	0.45	-0.80	0.30
Bernstein	2012	27	-1.39	0.17	-	-	-	-	-	-
Boyes	2009	89	-0.43	0.05	-0.80	0.05	-	-	-	-
Boyes	2009	87	-0.20	0.05	-0.84	0.05	-	-	-	-
Brochu	-	35	-2.51	0.20	-0.48	0.11	-	-	-	-
Brown	2009	52	-0.64	0.08	-	-	-	-	-	-
Carter	2008	143	-0.28	0.06	0.20	0.06	0.34	0.11	0.17	0.11
Carter-Sowell	2008	65	-2.86	0.12	-1.48	0.08	-	-	-	-
Carter-Sowell	2010	74	-1.60	0.14	-1.49	0.13	-1.23	0.33	-1.15	0.34
Carter-Sowell	2010	70.67	-2.09	0.17	-0.56	0.11	-0.65	0.39	-0.63	0.24
Chen	2012	60	-1.04	0.14	-	-	-1.35	0.27	-	-
Chen	2012	83	-1.32	0.11	-	-	-1.32	0.21	-	-
Chernyak	2010	76	-1.52	0.10	0.15	0.08	-	-	-	-
Chow	2008	75	-1.20	0.06	-1.31	0.06	-	-	-	-
Chrisp	2012	77	-0.70	0.06	-0.15	0.05	-	-	-	-

First author	Year	<i>N</i>	<i>d</i> T1	(SE)	<i>d</i> T2	(SE)	Δd T1	(SE)	Δd T2	(SE)
Coyne	2011	40	-0.56	0.10	-	-	-	-	-	-
De Waal-Andrews	2012	136	-3.55	0.16	-2.55	0.11	-1.29	0.24	-0.87	0.18
De Waal-Andrews	2012	112	-4.21	0.22	-2.17	0.11	-1.56	0.31	-1.20	0.18
DeBono	-	57	-1.07	0.15	-0.05	0.13	-1.55	0.29	-0.48	0.27
DeBono	-	81	-1.07	0.11	-0.10	0.09	-0.33	0.21	0.24	0.19
DeBono	-	83	-0.13	0.09	-	-	-0.75	0.19	-	-
Dietrich	2010	75	1.43	0.07	-	-	-	-	-	-
Duclos	2012	59	-0.63	0.07	-	-	-	-	-	-
Eisenberger	2006	48	-0.15	0.08	-1.24	0.10	-	-	-	-
Fayant	-	60	-2.04	0.20	-1.12	0.15	0.22	0.38	-0.44	0.28
Floor	2007	88	-1.92	0.13	-0.73	0.09	-0.21	0.28	-0.59	0.19
Gallardo-Pujol	2012	57	-1.18	0.16	-0.52	0.15	-1.17	0.31	0.11	0.29
Gan	2012	72	-0.54	0.03	-0.07	0.03	-0.62	0.06	0.02	0.06
Garczynski	2013	83	-1.51	0.19	0.39	0.15	-1.29	0.33	-0.01	0.29
Geniole	2011	74	0.19	0.06	-0.11	0.06	-	-	-	-
Gerber	-	38	-2.09	0.16	-	-	-	-	-	-
Gerber	-	89	-3.38	0.21	-	-	-	-	-	-
Gonsalkorale	2007	97	-1.31	0.14	0.26	0.12	0.49	0.30	1.31	0.25
Goodwin	2010	300	-1.81	0.04	-0.94	0.03	0.20	0.08	-0.43	0.07
Goodwin	2010	314	0.13	0.02	-0.09	0.02	0.35	0.06	-0.10	0.06
Greitemeyer	2012	56	-0.48	0.07	-0.23	0.07	-	-	-	-
Gruijters	-	113	-0.26	0.06	-1.07	0.07	-	-	-	-
Hackenbracht	2013	51	-1.92	0.11	-0.18	0.08	-	-	-	-
Hawes	2012	55	-2.16	0.23	0.69	0.15	0.00	0.38	-1.05	0.28
Hellmann	-	76	-1.21	0.12	0.19	0.10	-1.40	0.22	0.74	0.21
Hess	2010	162	-2.34	0.04	-0.87	0.03	-	-	-	-
Hess	2011	38	-0.64	0.11	-	-	-	-	-	-
Horn	-	68	-0.77	0.12	-0.99	0.13	-0.99	0.23	1.49	0.24
IJzerman	2012	86	-1.67	0.12	-	-	-1.07	0.22	-	-

First author	Year	<i>N</i>	<i>d</i> T1	(SE)	<i>d</i> T2	(SE)	Δd T1	(SE)	Δd T2	(SE)
Jamieson	2010	33	-1.56	0.15	-1.06	0.13	-	-	-	-
Jamieson	2010	68	-1.94	0.09	-1.47	0.07	-	-	-	-
Johnson	2010	104	-0.73	0.04	-0.79	0.04	-	-	-	-
Kassner	-	85	-1.72	0.13	-1.02	0.11	-0.87	0.31	-0.30	0.21
Kassner	2012	49	-2.11	0.12	-1.78	0.11	-	-	-	-
Kerr	2008	250	-1.66	0.02	-0.05	0.02	-	-	-	-
Kesting	2013	76	-0.28	0.05	-0.79	0.06	-	-	-	-
Knowles	2010	62	-0.38	0.12	-	-	-0.99	0.25	-	-
Knowles	2012	60	-0.60	0.07	-	-	-	-	-	-
Krijnen	2008	144	-4.74	0.11	-0.18	0.03	-	-	-	-
Krill	2008	119	-2.11	0.05	-0.57	0.03	-	-	-	-
Lakin	2008	36	-1.53	0.14	-0.51	0.11	-	-	-	-
Lau	2009	56	-2.50	0.23	-1.09	0.15	-0.06	0.58	1.36	0.46
Lustenberger	2010	71	-0.83	0.06	0.04	0.06	-	-	-	-
Lustenberger	2010	156	-0.70	0.03	-	-	-	-	-	-
MacDonald	2008	63	-0.15	0.06	-	-	-	-	-	-
McDonald	2012	270	-0.06	0.02	-2.40	0.03	-	-	-	-
Nordgren	2011	71	-0.74	0.06	-	-	-	-	-	-
Nordgren	2011	74	-0.80	0.06	-	-	-	-	-	-
Nordgren	2011	46	-2.24	0.14	-	-	-	-	-	-
Nordgren	2011	44.67	-0.55	0.09	-0.75	0.09	-	-	-	-
Nordgren	2011	58.67	-0.65	0.07	-	-	-	-	-	-
Oberleitner	2012	88	-2.36	0.08	0.42	0.05	-	-	-	-
O'Brien	2012	125	-0.58	0.03	-0.69	0.03	-	-	-	-
Peterson	2011	40	-0.89	0.11	-0.91	0.11	-	-	-	-
Pharo	2011	74	-1.33	0.13	-0.58	0.11	-1.01	0.30	-0.84	0.23
Plaisier	2012	149	-0.36	0.05	0.23	0.05	-0.40	0.11	-0.56	0.11
Ramirez	2009	121	-2.26	0.05	-1.02	0.04	-	-	-	-
Ren	2012	53	-2.18	0.12	-0.17	0.07	-	-	-	-

First author	Year	<i>N</i>	<i>d</i> T1	(SE)	<i>d</i> T2	(SE)	Δd T1	(SE)	Δd T2	(SE)
Renneberg	2011	60	-1.46	0.16	-1.30	0.15	0.47	0.29	0.51	0.29
Riva	2011	100	-2.10	0.13	-1.09	0.09	-	-	-	-
Ruggieri	-	91	-0.39	0.04	-0.57	0.05	-	-	-	-
Ruggieri	-	74	-0.06	0.13	-0.23	0.13	-0.31	0.24	-0.68	0.23
Sacco	2011	51	-2.40	0.13	-1.45	0.10	-	-	-	-
Sacco	2011	21	-2.28	0.29	-1.46	0.22	-	-	-	-
Sacco	2011	38	-1.74	0.14	-1.04	0.11	-	-	-	-
Salvy	2010	59	-1.45	0.08	-1.43	0.08	-	-	-	-
Salvy	2009	103	-1.48	0.05	-1.31	0.05	-	-	-	-
Schaafsma	2012	720	-1.42	0.02	-0.49	0.02	0.09	0.03	0.33	0.03
Segovia	2012	56	0.14	0.13	-	-	-1.89	0.32	-	-
Staebler	2011	68	-0.79	0.12	-0.05	0.12	0.50	0.23	0.42	0.23
Stillman	2009	121	-0.74	0.15	-1.13	0.16	0.57	0.22	-1.19	0.24
Stock	2011	155	-2.00	0.04	-0.13	0.03	-	-	-	-
Van Beest	2011	87	-0.94	0.10	-0.58	0.09	-0.40	0.24	-0.44	0.19
Van Beest	2011	183	-2.64	0.13	-0.50	0.07	-0.76	0.22	-0.11	0.13
Van Beest	2006	135	-1.29	0.07	-0.65	0.06	-0.10	0.14	-0.13	0.12
Van Beest	2006	111.33	-2.11	0.11	0.09	0.07	-0.09	0.22	-0.19	0.14
Van Beest	2012	125	-2.68	0.11	-1.24	0.07	0.06	0.35	-0.23	0.15
Van Beest	2012	85	-3.10	0.20	0.05	0.09	-0.28	0.44	0.07	0.18
Van Dijk	-	51	-1.50	0.10	-0.04	0.08	-	-	-	-
Webb	-	170	-0.91	0.05	-0.38	0.05	0.03	0.10	0.04	0.09
Weik	2010	65	0.16	0.12	-0.22	0.12	-0.43	0.24	0.66	0.24
Wesselmann	2009	82	-0.71	0.10	-2.03	0.14	-1.30	0.24	-0.20	0.28
Wesselmann	2012	91	-1.46	0.06	-	-	-	-	-	-
Williams	2002	390	-0.39	0.01	-2.35	0.02	-	-	-	-
Williams	2000	732	-0.79	0.01	-1.44	0.01	-	-	-	-
Williams	2000	111	-0.26	0.06	-1.01	0.07	-0.20	0.15	-0.98	0.15
Wirth	2009	159.33	-2.29	0.08	-0.76	0.05	0.05	0.17	0.46	0.11

First author	Year	<i>N</i>	<i>d</i> T1	(SE)	<i>d</i> T2	(SE)	Δd T1	(SE)	Δd T2	(SE)
Wirth	2010	76	-0.96	0.06	-1.64	0.07	-	-	-	-
Zadro	2004	62	-1.63	0.16	-0.19	0.12	-0.11	0.32	-1.12	0.28
Zadro	2004	77	-1.75	0.14	-0.33	0.10	-0.29	0.28	-0.70	0.21
Zadro	2006	56	-3.70	0.19	-0.87	0.08	-	-	-	-
Zhong	2008	52	-0.72	0.15	-	-	-	-	-	-
Zoller	2010	57	-0.24	0.07	-0.09	0.07	-	-	-	-
Zwolinski	2012	56	-2.01	0.11	-0.28	0.07	-	-	-	-

Note. *d* T1 refers to ostracism effect on first measure; *d* T2 refers to ostracism effect on last measure; Δd represent interactions. Non-integer *Ns* arise from division of full sample *N* for included conditions, appropriate due to random assignment.

Table 3

Interaction effect per subset

							95% CI	95% CI
		k	Estimate	(SE)	Z-value	p -value	Lowerbound	Upperbound
Overall	T1	52	-0.46	0.09	-5.08	< .001	-0.64	-0.28
	T2	46	-0.19	0.11	-1.82	.069	-0.40	0.02
Fundamental	T1	30	-0.39	0.12	-3.42	< .001	-0.62	-0.17
	T2	17	-0.77	0.25	-3.05	.002	-1.27	-0.28
Intrapersonal	T1	42	-0.31	0.09	-3.38	< .001	-0.49	-0.13
	T2	39	-0.21	0.11	-1.87	.062	-0.44	0.01
Interpersonal	T1	10	-1.03	0.18	-5.69	<.0001	-1.38	-0.67
	T1 _{listwise}	6	-0.36	0.22	-1.63	.104	-0.79	0.07
	T2	6	0.63	0.62	1.02	.309	-0.58	1.84
Model	T1	36	-0.29	0.10	-2.99	.003	-0.48	-0.10
	T2	23	0.01	0.17	0.08	.938	-0.31	0.34

Note. Overall estimates are based on all data, where the rest form subsets. Model indicates that the first measure was indeed reflexive and the last measure reflective. Listwise deletion for equal *ks* across time points within a subset yielded highly similar results, except for interpersonal measures, which is depicted in the row labeled T1_{listwise}.

Table 4

Meta regression coefficients for composition effects (first measure; k = 45)

	Estimate	(SE)	Z-value	p-value	95% CI Lowerbound	95% CI Upperbound
Intercept	-1.49	3.27	-0.46	0.649	-7.90	4.92
Country = US	-	-	-	-	-	-
Country = Western	-0.42	0.36	-1.15	0.249	-1.13	0.29
Country = Asian	-0.30	1.13	-0.26	0.793	-2.51	1.92
Proportion male	1.54	1.09	1.42	0.156	-0.59	3.68
Mean age	-0.05	0.05	-0.97	0.332	-0.16	0.05
Nr. of players	-0.22	1.05	-0.21	0.837	-2.28	1.85
Ostracism <5 min	-	-	-	-	-	-
Ostracism 5-10 min	0.75	0.81	0.92	0.358	-0.84	2.34
Need scale = Williams (2000)	-	-	-	-	-	-
Need scale = Zadro et al. (2004)	-0.36	0.41	-0.88	0.381	-1.16	0.45
Need scale = Van Beest & Williams (2006)	0.07	0.54	0.13	0.894	-0.98	1.12
Need scale = Williams Zadro	-0.03	0.62	-0.04	0.965	-1.25	1.19
Need scale = Gonsalkorale & Williams (2007)	0.68	0.83	0.82	0.414	-0.94	2.30
Nr. of throws	0.03	0.02	1.49	0.137	-0.01	0.07

Note. This can be interpreted as a standard regression formula. Empty rows represent reference categories.

Table 5

Meta-regression coefficients for composition effects (last measure; k = 41)

	Estimate	(SE)	Z-value	p-value	95% CI	
					Lowerbound	Upperbound
Intercept	-5.76	2.42	-2.38	0.017	-10.51	-1.01
Country = US	-	-	-	-	-	-
Country = Western	0.26	0.30	0.87	0.387	-0.33	0.86
Country = Asian	0.85	0.84	1.01	0.313	-0.80	2.49
Proportion male	0.29	0.83	0.35	0.730	-1.34	1.91
Mean age	-0.01	0.04	-0.25	0.806	-0.10	0.08
Nr. of players	1.55	0.78	1.98	0.047	0.02	3.07
Ostracism <5 min	-	-	-	-	-	-
Ostracism 5-10 min	0.38	0.62	0.61	0.539	-0.83	1.59
Need scale = Williams (2000)	-	-	-	-	-	-
Need scale = Zadro et al. (2004)	-0.14	0.32	-0.44	0.658	-0.77	0.49
Need scale = Van Beest & Williams (2006)	-0.21	0.41	-0.51	0.613	-1.02	0.60
Need scale = Williams Zadro	-0.12	0.53	-0.22	0.826	-1.16	0.92
Need scale = Gonsalkorale & Williams (2007)	-0.07	0.65	-0.10	0.916	-1.33	1.20
Nr. of throws	0.01	0.02	0.59	0.556	-0.02	0.04

Note. This can be interpreted as a standard regression formula. Empty rows represent reference categories.

Figure 1. A screenshot of the Cyberball game.



Figure 2. Dotplots of the average estimated simple effects with 95% confidence intervals, where T1 represents first measure, and T2 represents last measure. Traditional ostracism effect refers to the between-subjects effect of being ostracized with *no* moderator present, whereas moderated ostracism effect refers to being ostracized *with* a moderator present. Vice versa, moderator effect within ostracism/inclusion level refers to the between-subjects effect of the moderator factor, within the ostracized/inclusion conditions. All = all measures; Fundamental = only fundamental need measures; Intrapersonal = all intrapersonal measures; interpersonal = all interpersonal measures; model = first is immediate and last is delayed. For lists of studies in each subset, see Supplementary Materials.

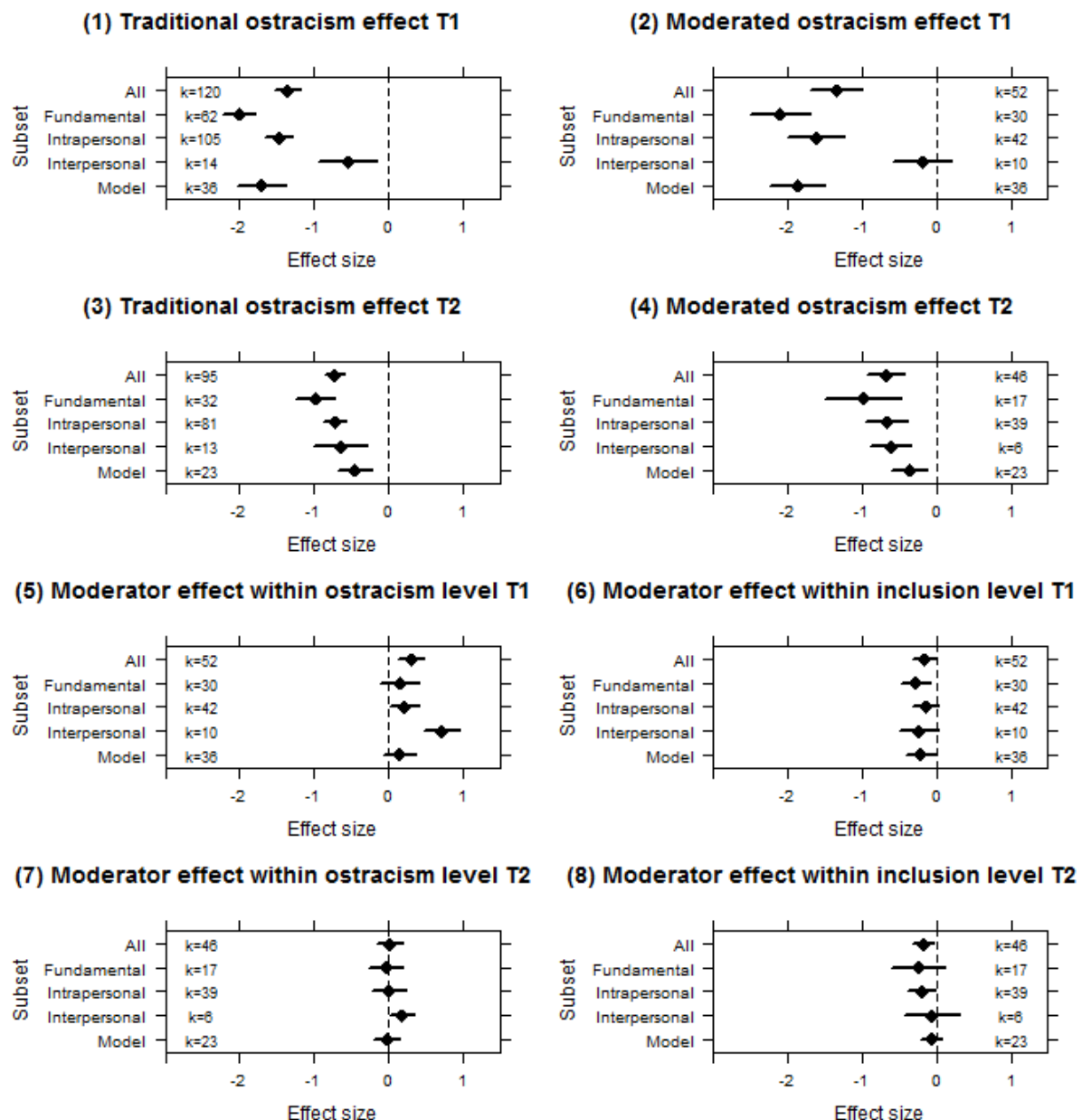
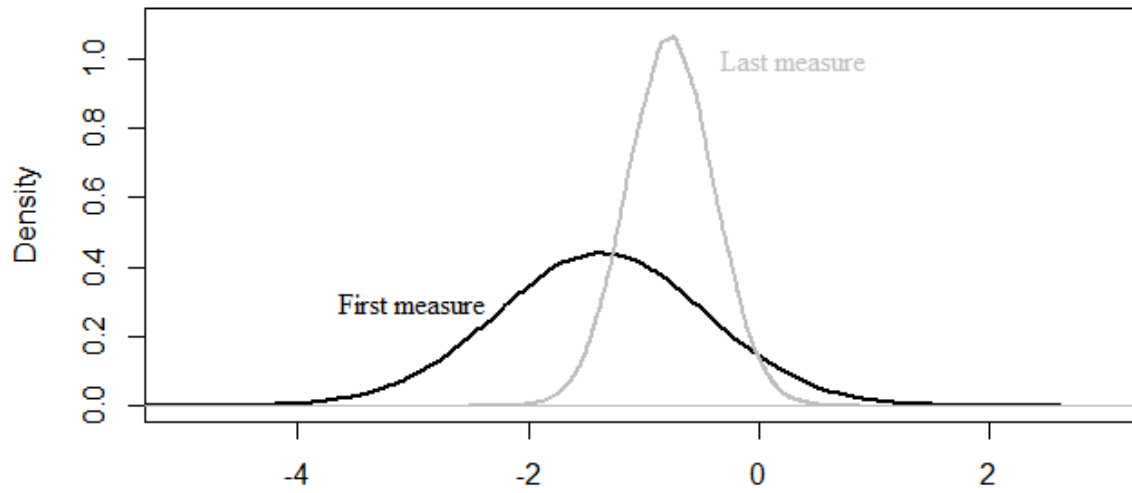


Figure 3. Simulated effects under the model estimates for the standard ostracism effect, showing higher estimated heterogeneity on the first measure than last measure.



Appendix

All formulae reported below originate from the chapter by Michael Borenstein (2009).

Hedges' g was calculated as

$$g = d \left(1 - \frac{3}{4(df_w) - 9} \right)$$

where d is the standardized main effect. For the standardized interaction effect d was calculated as

$$\Delta d = \frac{(\bar{X}_{11} - \bar{X}_{12}) - (\bar{X}_{21} - \bar{X}_{22})}{s_p}$$

where the first term in the nominator is the ostracism effect and the second term is the ostracism effect in the moderator conditions. This Δd corresponds to the partial eta-squared of the interaction. Sampling variance of g was calculated by multiplying the sampling variance of d by the squared correction factor, that is

$$s_g = \left(1 - \frac{3}{4(df_w) - 9} \right)^2 s_d$$

where the sampling variance of the interaction was calculated as the sum of the sampling variances of both the simple main effects.

Reference list of studies included in meta-analysis

[Click here to download Supplemental Material - Integral: Reference list of included studies.docx](#)

Comprehensive list of studies selected into subsets

[Click here to download Supplemental Material - Additional: Panels subsets list.xlsx](#)