Stake size effects in Ultimatum Game and Dictator Game offers: A meta-analysis

Andrea Larney, Amanda Rotella and Pat Barclay*
Department of Psychology, University of Guelph

Acknowledgements
We thank the following people for useful discussions: Maxwell Burton-Chellew, Stuart West, Miguel Dos Santos, Jillian O'Connor, Megan Kelly, and Savreet Chuckal.
We thank the following researchers for providing data or discussing their papers: Felix Bader, Michael Carr, Todd Cherry, Andreas Diekmann, Anthony Gabay, Glenn Harrison, Freya Harrison, Shachar Kariv, Anand Krishna, John List, Ananta Neelim, Nichola Raihani, David Reinstein, Bradley Ruffle, Laura Schons, and Israel Waichman. This work was supported by the Social Sciences and Humanities Research Council of Canada [grant number 430287].

Abstract

Are people more generous when less money is at stake? The Ultimatum Game (UG) and Dictator Game (DG) are often used as models of bargaining and charitable giving, respectively. Previous studies have produced conflicting results on whether UG and DG offers are lower when the stakes are high, and many previous studies had insufficient statistical power to detect significant effects of stake size. To resolve this, we conducted a meta-analysis of 31 existing studies that manipulated the size of participants' endowments in the UG and DG (3233 total participants). We hypothesized that: 1) proposer offers would be lower with larger stakes in both games, owing to an increased cost of giving; and 2) offers would decrease more with stake size in the DG than the UG because proposers would not want to risk their offer being rejected in the UG. Our results found almost zero effect of stake size on UG offers (d=0.02), and a small but significant effect of stake size on DG offers (d=0.14). Furthermore, larger differences in stakes had little impact on the effect sizes in the UG, but had a medium-large impact on the effect sizes in the DG. These results show that higher stakes reduce donations in the DG, albeit not by much, and have little to no effect in the UG.

Keywords: bargaining, altruism, donations, endowment, stake size, meta-analysis
JEL codes: C78 (bargaining theory), C91 (Laboratory, individual behavior), D90 (Micro-based behavioral economics, general),

* Corresponding Author
Pat Barclay
Department of Psychology
University of Guelph
50 Stone Rd. E.
Guelph, ON, Canada, N1G 5L5
Phone: 1-519-824-4120 ext. 58247
Fax: 1-519-837-8629
ORCID ID: 0000-0002-7905-9069
barclayp@uoguelph.ca
www.patbarclay.com

Introduction

Bargaining Games

Researchers have used bargaining game experiments in the fields of psychology and economics for many years with the hope of extrapolating their results to real-world bargaining situations. Güth et al. (1968) introduced one such type of game, termed the Ultimatum Game (UG), to model bargaining situations such as contract or other business negotiations. The UG is a two-person game where one person (the proposer) makes an offer to the second person (the responder) about how they would suggest dividing a given stake. The responder can either accept or reject this offer; acceptance would split the stake as proposed, but rejection (of an unfair offer) would result in neither party getting any of the stake. Güth et al. (1968) note the need for strategy formation in the UG; each person is essentially acting independently from the other in their own self-interest. The UG examines how strategy formation (i.e. finding the lowest offer the responder will accept) interacts with self-serving behaviors (trying to keep as much of the stake as possible) (Güth & Kocher, 2013) under varying circumstances by measuring either how much the proposer will offer, or how little the responder will accept.

Kahneman, Knetsch & Thaler (1986) introduced a second type of bargaining game, the Dictator Game (DG), to model charitable giving and examine other-regarding behavior. Forsythe et al. (1994) developed a simplified version of the DG that is commonly used in experiments (Engel, 2011). The DG is also a two-person game in which the proposer offers a portion of their stake to a responder who does not have the choice to reject. The DG measures how much proposers give under experimentally manipulated conditions, to mimic giving in charitable settings.

General trends in the DG and UG. Researchers typically compare results from the DG and UG to economic theoretical predictions to examine how human characteristics contribute to proposer and responder behavior. They generally find that human behavior deviates from these predictions. In the DG, theory predicts that to maximize their payoff, proposers should offer zero to the responder, as there is no risk of punishment for being selfish or benefit for being generous. However, Engel's (2011) meta-analysis of the DG found that DG proposers only offer zero 36% of the time, on average. Instead, Engel (2011) reports DG offers tend to be greater than zero – averaging about 28.3% of the given stake – and deviations from this average tend to be smaller than 28.3%. This deviation from the theoretical prediction can be explained by the presence of various human characteristics that standard economic theory does not account for. These characteristics largely include other-regarding behaviors such as adhering to cultural fairness norms (Rabin, 1993), generosity, and approval-seeking. Individual differences in these human characteristics can also contribute to explaining between-subject behavior differences in the DG.

In the UG, theory predicts that to maximize their payoff, proposers should offer the smallest possible non-zero amount that the responder will accept (Webster, 2013). The responder, to maximize their own payoff, should be willing to accept any offer above zero, regardless of the proportion. Rejections would only occur, then, in response to zero offers as it has no cost to the responder to do so (Webster, 2013). Early UG studies, however, report that UG proposers typically reject offers they perceive to be unfair to punish the proposer for being selfish (Cameron, 1999). Offers perceived to be unfair are, on average, those less than 20-30% of the stake (Camerer & Thaler, 1995). Because of the tendency of the responders to reject unfair offers and the ability of the proposer to predict such rejections, offers in the UG average 40-50% of the stake in industrialized societies (Oosterbeek, Sloof & Van De Kuilen, 2004). Although

small-scale societies sometimes have lower offers than in industrialized societies, no societies conform to the standard economic prediction of near-zero offers and universal rejections (Henrich et al., 2010). In addition to human characteristics that influence behavior in the DG, UG behavior is also influenced by risk aversion on the part of the proposer; since proposers can predict the human tendency of the responder to reject unfair offers, proposers tend to offer higher amounts than the theoretically predicted, smallest non-zero, offer to reduce the risk of their offer being rejected (Holt & Laury, 2002). Like the other human characteristics that affect these games, participants can also have individual differences in risk aversion that may account for between-subject behavior differences.

Since the establishment of these trends, a main focus of research on these games has increasingly been put on finding explanatory factors for deviations from these averages to uncover confounds and make the games more representative of real-world bargaining situations.

Effect of Stake Size

One of the factors researchers commonly examine in these games is how characteristics of the stake itself affect proposer behavior. Authors of early DG and UG studies (e.g. Hoffman, McCabe & Smith, 1996; Cameron,1999; Güth et al., 1968) mention that the size of the stake may act as a confound, as lab studies typically use relatively small stake sizes due to budgetary constraints. However, many of the real-world bargaining or charitable situations to which these games hope to generalize involve much larger stakes. To reconcile this potential conflict, researchers studying stake size effects ask whether the average deviations from theory would be similar enough in higher stake conditions to justify generalizations of low stake, budget-friendly experiments to higher stake conditions. There have been conflicting findings, however, about the effect of stake size in bargaining games, which has led researchers using the UG or DG to either

control for stake size or not, depending on the state of research at the time of their study. It is important to know if there is an effect or not so researchers can have an idea of whether stake size is an important confound to control for.

Measuring stake size effects. Past studies that have examined stake size effects have done so by creating separate low, high, and possibly intermediate stake size conditions. Due to the limited budget a lab may have to conduct their study, the high-stake condition may still be a relatively low amount. To make high stakes conditions more practical, some researchers have used hypothetical stake conditions, allowing them to set the stake size to any desired amount (Xu et al., 2016; Amir et al., 2012; Ben-Ner et al., 2008). Hypothetical stake conditions assume that people will accurately estimate how they would behave in a real stakes condition with the same amount of money; however, some researchers question this assumption (e.g. Xu et al., 2016). Other researchers have overcome budgetary constraints by conducting their studies in developing countries where lower currency values relative to USD$ lowers the cost of creating high stakes conditions (e.g. Slonim & Roth, 1998; Andersen et al., 2011).

State of the Research

Stake effects in non-bargaining games. In a review of incentive effects on participant performance in experiments, Camerer & Hogarth (1999) found little effect of stakes on behavior, but found that variance in performance may decrease as stakes increase. Additionally, Kocher, Martinsson & Visser (2008) found no effect of stakes on cooperation or punishment in the Public Goods Game (PGG). Johanssen-Stenman, Mahmud & Martinsson (2005), however, found that offers in trust games decreased significantly as stake size increased.

Stake effects in the DG and UG. Among the studies examining the effects of stake size on proposer behavior, Hoffman et al. (1996), Slonim & Roth, (1998), Andersen (2011), Forsythe

et al. (1994), and Engel (2011) are some of the most frequently cited in studies examining stake size effects. Hoffman et al. (1996) compared the effect of $10 and $100 stakes on proposer offers in the UG, but found no significant differences in offers between conditions. Slonim & Roth (1998) conducted a UG in the Slovak Republic, varying the stakes from 60, 300, and 1500 Slovak Crowns (USD$1.90, $9.70, $48.40), but found no effect on proposer behavior. Similarly, Forsythe et al., (1994) found no effect of comparing $5 and $10 stakes.

In contrast, Andersen et al. (2011) compared stakes of 20, 200, 2000 and 20 000 rupees (a substantially large stake) (USD$0.41, $4.10, $41.00, $410.00) in an UG in Northeast India, and found that proposers offered significantly lower proportions in the higher stake condition. Similarly, a frequently cited meta study by Engel (2011) examined the effects of several factors on the DG, and found a small but significant effect of stake size where proposers offered lower proportions in the higher stake conditions.

Of the studies that have found an effect of stakes, many are newer studies that have increased the differences between stake conditions compared to older studies (e.g. Leibbrandt et al., 2015 (TK100 (USD$1.22) vs. TK10000 (USD$122.00); Andersen et al., 2011) indicating that the high stakes conditions in the studies that did not have an effect may have had too small of an amount as the high stake condition to be able to see this difference (e.g. Forsythe et al.'s (1994) stakes of $5 and $10). Additionally, any studies that have found an effect of stake size in the DG or UG have found that offers decrease with increasing stakes.

Large-scale studies. Other large scale studies attempting to consolidate stake effects in the DG and UG included an array of studies with potential confounds of the effect of stake size on proposer behavior (Engel, 2011: small effect of stakes in the DG; Karagözoğlu & Urhan, 2016: inconclusive). These confounds include: stake origin (earned versus windfall), where

earned stakes reduce willingness to give in the DG (Bediou et al., 2012; Cherry, 2001); whether the responder knows the stake size or not, where offers decrease when the stake is unknown to the responder (Rapoport & Sundali, 1996; Rapoport, Sundali, & Seale, 1996; Straub & Murnighan, 1995); varying levels of inequality when the responder also has a starting stake, where offers decrease when responders start with a higher stake (Korenok, Millner, & Razzolini, 2012); and hypothetical versus real stakes, where results are varied (Amir et al., 2012; Ben-Ner et al., 2008). The current study will select studies that do not involve these confounding factors to isolate effects of the size of the stake only.

The Current Study

We use a meta-analytic approach consolidate the data on the effects of stake size on proposer behavior in the UG and DG that have been observed thus far to determine if there is a significant effect of stake size on offers. This will update past studies with newer research while controlling for confounds not previously accounted for. This information should provide insight into whether stake size is a variable that researchers should consider as a confound in UG and DG studies as it relates to proposer behavior, and whether low stakes games are generalizable to high stake conditions.

Hypotheses. We predict that: 1) in the UG and DG, there will be an effect of stake size on proposer behavior, causing proposer offers to decrease as the stake size increases; and 2) there will be a larger effect of stake size in the DG in this way than in the UG because of the added influence of risk in the UG.

***Rationale: Theoretical basis of stake size effects*.** When studies do report an effect of stakes, they tend to find that offers decrease as stake size increases. Higher stakes increase the cost (i.e. the total amount of money that would have to be given) of making an offer from what

the same proportion would have cost in a lower stakes condition (e.g. parting with $500 of a $1000 stake (50%), compared to parting with only $5 of a $10 stake). This higher cost of giving may make the proposer less willing to part with the same proportion of the stake they might have in lower stake conditions, where the cost of giving is much lower (Andersen et al., 2011; Slonim & Roth, 1998; Hoffman et al., 1996). Additionally, Fu, Kong & Yang (2007) suggest that as the cost of giving and stake size increases this way, the money becomes more salient than any social concerns that influence the offer size at lower stakes. Similarly, Bethwaite & Tompkinson (1993) suggest that other-regarding behaviors and fairness norms become less important as stakes increase. This model would predict that proposers would decrease their offers as stakes increase.

Compared to the DG, the UG may be less impacted by stake size because of the added presence of risk in the UG. As stakes increase, risk aversion also increases in the UG as there is more to lose if the offer is rejected (Karagözoğlu & Urhan, 2016), causing proposers to increase offers to prevent rejection from the proposer (Holt & Laury, 2002). This factor works to minimize any increased saliency of the money due to the higher stake, resulting in the UG showing less of an effect of stake size than the DG.

Method. We will use a meta-analysis to answer this question. A meta-analysis can work to increase the confidence of a result from single experiments because of the large cumulative sample size that results from pooling data across samples in multiple studies. If many studies report a similar effect, consolidating these studies using a meta-analysis can provide support of the existence of this effect with greater confidence than one study could have on its own. Furthermore, while any single study is subject to sampling error and may over- or underestimate the true effect size, a meta-analysis can average out this sampling error by combining multiple studies. Our meta-analysis will consolidate UG and DG studies that have manipulated stake size

by searching the PSYCInfo, Web of Science, Google Scholar, and Econlit databases using specific search, inclusion, and exclusion criteria, and by sending messages to relevant listservs to find unpublished studies. We calculated the standardized mean differences of offers between conditions to analyze the average effect size over all studies.

Method

Locating Studies

We conducted searches of the PSYCInfo, Google Scholar, Web of Science, and Econlit databases using various combinations of the following search terms: dictator game or games, ultimatum game or games, bargaining game or games, stake or stake size, endowment or endowment size, pie or pie size; we note that Google Scholar searches many working papers and unpublished articles. We also searched reference lists of the articles located by the database searches for additional relevant studies. Additionally, we sent e-mails asking for published or unpublished studies to the e-mail lists of the Economic Science Association, Society for Personality and Social Psychology, and Human Behavior and Evolution Society. Finally, we put out a call for studies on Twitter which resulted in 21,015 impressions and 127 total engagements. Whenever a study had insufficient statistical detail to calculate effect sizes for stakes (e.g., no standard deviations or inferential statistics on stakes), we attempted to contact the authors for clarification or raw data.

Study Selection Criteria

Inclusion criteria. We selected two-player UG and DG games, as studies using more than two-players introduce additional layers of complexity that may confound the findings on stake size effects. We only included games using adult players; although studies have been done with children (Blake & Rand, 2010; Posid et al., 2015), other studies examining child behavior in

these games have found children's perceptions of fairness and giving tendencies to fluctuate with age (Kogut, 2012). This finding gives us reason to include only adults since the results of UGs and DGs are typically extrapolated to adult bargaining or giving situations. We also included studies that use modified versions of the UG and DG in which the researchers present the proposer with a limited set of offers to choose from, as this is seen to assess offer behavior the same way a free-choice experiment would (Bolton, Katok & Zwick, 1998; Engel, 2011).

Exclusion criteria. Table 1 summarizes the exclusion criteria along with the rationale for exclusion and any exceptions for a study's inclusion despite matching exclusion criteria. The Appendix includes a list of excluded studies and the rationales for their exclusion.

Table 1: Exclusion criteria, rationale, and exceptions

Exclusion Criteria	Rationale for Exclusion	Exception for Inclusion
Earned endowments	Confounds the effects of stakes with the effects of entitlement	Data from unearned or windfall conditions of the same paper can be included
Alternate framings without control	Does not assess offering behavior in the same way (e.g. framed as proposer taking from the responder's stake instead of giving part of their own)	Data from normal framing (i.e. give) conditions can be included if they involve varying stakes. Data from alternate framings is included if both high and low stakes have same framing
Priming or other experimental manipulation before the main task	May introduce a confound on proposer behavior	Data from control (i.e. unprimed) conditions can be included
Conditions where responder does not know stake size	May introduce a confound on proposer behavior; e.g. Straub & Murnighan (1995) show that when responders do not know the stake size, they tend to accept lower offers, and proposers tend to propose lower offers, independent of stake size effects	Data from control (i.e. known stakes) conditions can be included

Cross-cultural studies where stake sizes are equivalent in purchasing power	Stake sizes appear to vary, but are only varied as a method to keep stake size *equivalent* across cultures by accounting for variation in purchasing power (Henrich et al., 2010)	N/A
Hypothetical stakes	May introduce a confound on proposer behavior; may or may not be seen to increase offer size as risk decreases from lack of real consequences of losing money (e.g. Amir et al., 2012; Keushnigg, Bader & Bracher, 2016)	N/A
Paper does not contain new data on stakes in Ultimatum or Dictator Games (e.g., theoretical model, review, different game)	Meta-analysis combines empirical studies, not theoretical results	N/A
Insufficient statistical detail to calculate effect size (e.g., no SD or SEM, imprecise statistics like "n.s.") *plus* authors could not provide raw data for us to calculate it ourselves	If there is insufficient detail to calculate an effect size for a study, then there is no number to include in our quantitative meta-analysis	N/A

Coding Process

We organized the included studies using the following categories: type of game (UG or DG/ classic or modified), sizes of stakes between conditions, and reported effect of stake size including statistical data.

Data Analysis

Conditions analyzed. The analyses include the effect of stake size on proposer behavior separately for the DG and the UG, together for both games, and for the difference in effect size between the DG and UG, resulting in three overall measures.

Measure of effect. We extracted data necessary from each study to calculate standardized mean differences (Cohen's *d*) and confidence intervals. The type of data collected depended on what had been reported by the authors which included sample sizes, t values of differences between

offer size in each stake condition, means and standard deviations of offers in each stake condition, or the raw data reporting individual offers in each stake condition. Once converted to Cohen's *d,* we used the metaphor (Viechtbauer, 2010) package in the R statistical software (R Core Team, 2013) to conduct both a random and fixed-effects meta-analysis to yield average effect sizes and confidence intervals for the UG and DG studies separately, and together. We calculated unbiased Cohen's *d* values, which uses δ to weight each study's SD by the sample size, in order to correct for any bias due to small samples common among the included studies (Cumming, 2013). Studies were weighted by inverse variance. We used the Q test for heterogeneity to determine whether a random-effects model – used in cases of high heterogeneity between studies – was justified over a fixed-effects model. Study data and analysis scripts are available at [https://osf.io/hc3py/?view_only=8a78540a4d1546a4be97628fd67a8016. Note: this is a blind link for reviewers, and will be replaced by our non-blinded link upon publication].

Because of the wide variety of stake conditions the authors used, we also calculated correlation coefficients (*r*) to examine any potential relationships between the calculated effect sizes and the differences in stake sizes used.

Results

Included and Excluded Studies

The meta-analysis included a total of 21 papers using the UG, the DG, or both (UG only, N=7; DG only, N=10; both, N=4*),* resulting in 31 effect sizes (UG, N_d=13*;* DG, N_d=18) from 3233 total participants, as summarized in Table 2. The weights are based on the inverse variance of the estimate, which is strongly affected by sample size – larger studies are weighted more. There was a wide range of stakes: the median high stake condition was $100 in the UG (range:

$10 to >1 year income (Andersen et al., 2011) and $20 in the DG (range: $10 to 100 days salary (Leibbrandt et al., 2015)). In the median study, the high stakes were worth 10 times more than the low stakes in the UG (range: 2 to 10,000) and 4.5 times more in the DG (range: 1.75 to 100).

We excluded a total of 80 papers that went against our exclusion criteria. The most common reasons for exclusion were: they were not empirical studies of a UG or DG (14 papers), had hypothetical stakes instead of varying real stakes (11 papers), had earned endowments without an unearned control (7 studies), UG responders did not know the stakes (6 studies), or did not provide sufficient statistical detail (e.g., standard deviations, inferential statistics) to calculate effect sizes and we were unable to contact the authors for data (11 studies). Many of the studies were not principally designed to test the effects of stakes, which is perhaps why they did not present sufficient statistical detail on this question, and why they met our exclusion criteria. The Appendix (Table 3) contains the full list of excluded studies and rationales.

There is no evidence for publication bias. The funnel plot (Figure 1) is symmetrical, and a regression test for funnel plot asymmetry was non-significant, z=1.25, p=.21. Anecdotally, the earliest studies on stake size tended to find no effect of stakes (e.g., Cameron, 1999; Forsythe et al., 1994; Hoffman et al., 1996; Ruffle, 1998), and this finding was seen as reassuring because it suggested that these games were invariant to stakes. Some studies even touted no effect of stakes, despite finding a small-medium effect that was non-significant due to low statistical power (e.g., Carpenter et al., 2005). As such, the usual reasons for publication bias (i.e., non-significant results not getting published) do not appear to exist for this research question, and the symmetrical funnel plot supports this contention.

Table 2: Ultimatum Games (UG) and Dictator Games (DG) included in the meta-analysis.

Authors	UG or DG	Repeated Measures	N	Stake size[a]	Diff. in Stakes (x)	Weights (%)	Effect Size (Cohen's d)[b]	95% CI
Andersen et al., 2011	UG	No	126	Rs20/Rs200/ Rs2000/Rs20000	1000	4.0%	0.27	[-0.09, 0.62]
Cameron 1999, UG2	UG	Yes	37	Rp5000/Rp40000	8	4.8%	-0.16	[-0.47, 0.14]
Cameron 1999, UG1	UG	Yes	35	Rp5000/Rp200000	40	6.3%	-0.09	[-0.33, 0.15]
Carpenter et al., 2005 UG	UG	No	39	$10/$100	10	1.6%	0.36	[-0.27, 1.00]
Carpenter et al., 2005 DG	DG	No	40	$10/$100	10	1.6%	0.33	[-0.29, 0.96]
Carr & Mellizo, 2017	UG	No	84	$10/$20/$40	4	3.0%	-0.14	[-0.57, 0.29]
Cherry, 2001	DG	No	50	$16-$28	1.75	1.9%	-0.00	[-0.56, 0.56]
Cherry & Shogren, 2008	DG	No	68	$10/$20	2	2.6%	-0.22	[-0.69, 0.26]
Cherry et al., 2002	DG	No	52	$10/$40	4	2.1%	0.11	[-0.43, 0.66]
Forsythe et al., 1994 UG	UG	No	67	$5/$10	2	2.4%	-0.12	[-0.62, 0.38]
Forsythe et al., 1994 DG	DG	No	69	$5/$10	2	2.4%	-0.06	[-0.55, 0.44]
Fu et al., 2007	UG	No	397	NT$200/NT$1000	5	5.4%	0.25	[-0.03, 0.53]
Gabay et al., 2018	UG	Yes	20	£6-£53	20	0.8%	0.62	[-0.32, 1.57]
Harrison & El Mouden 2011	DG	No	30	£2/£4/£6/£8/£10	5	1.2%	0.22	[-0.51, 0.95]
Heinz et al., 2012	DG	No	83	€5/€10	2	3.0%	0.25	[-0.19, 0.68]
Hoffman et al., 1996	UG	No	51	$10/$100	10	2.0%	-0.07	[-0.62, 0.48]
Kettner & Waichman, 2016 Take	DG	No	43	$5/$20	4	1.7%	0.14	[-0.46, 0.74]
Kettner & Waichman, 2016 Give	DG	No	44	$5/$20	4	1.7%	0.39	[-0.22, 1.00]
Keuschnigg et al., 2016 UG USA	UG	No	186	$1/$4/$10	10	5.1%	-0.07	[-0.36, 0.22]
Keuschnigg et al., 2016 UG India	UG	No	186	$0.40/$1.60/$0.40	10	5.1%	-0.07	[-0.36, 0.22]
Keuschnigg et al., 2016 DG USA	DG	No	190	$1/$4/$10	10	5.2%	-0.08	[-0.36, 0.21]
Keuschnigg et al., 2016 DG India	DG	No	190	$0.40/$1.60/$0.40	10	5.2%	0.17	[-0.12, 0.45]
Leibbrandt et al., 2015 Take	DG	No	45	100Tk/10000Tk	100	1.8%	-0.10	[-0.69, 0.42]
Leibbrandt et al., 2015 Give	DG	No	45	100Tk/10000Tk	100	1.6%	1.10	[0.47, 1.72]
Raihani et al., 2013 India	DG	No	282	$1/$5/$10	10	6.4%	0.32	[0.08, 0.56]
Raihani et al., 2013 USA	DG	No	292	$1/$5/$10	10	6.5%	-0.00	[-0.23, 0.23]
Reinstein et al., 2012	DG	No	102	€5/€7.5/€10	2	3.3%	0.54	[0.14, 0.94]
Ruffle, 1998 UG	UG	No	44	$4/$10	2.5	1.7%	0.55	[-0.05, 1.15]
Ruffle, 1998 DG	DG	No	52	$4/$10	2.5	2.1%	-0.11	[-0.66, 0.43]
Schier et al., 2016	DG	No	202	Tickets for $10/$500	50	5.4%	0.23	[-0.05, 0.51]
Slonim & Roth, 1998	UG	No	82	Sk60/Sk300/Sk1500	25	2.2%	0.05	[-0.47, 0.58]

[a] Rs=Indian rupees (day's wages ~100 RS). Rp=Indonesian Rupiah (3 months wages ~200000 Rp). NT$=Taiwan New Dollar (hourly wage ~NT$100). Tk=Bangladeshi Taka (daily wage ~100Tk). Sk=Slovak crowns (hourly wage ~60 Sk)

[b] This represents unbiased d-values. Positive effect sizes mean that offers are higher when stakes are small.

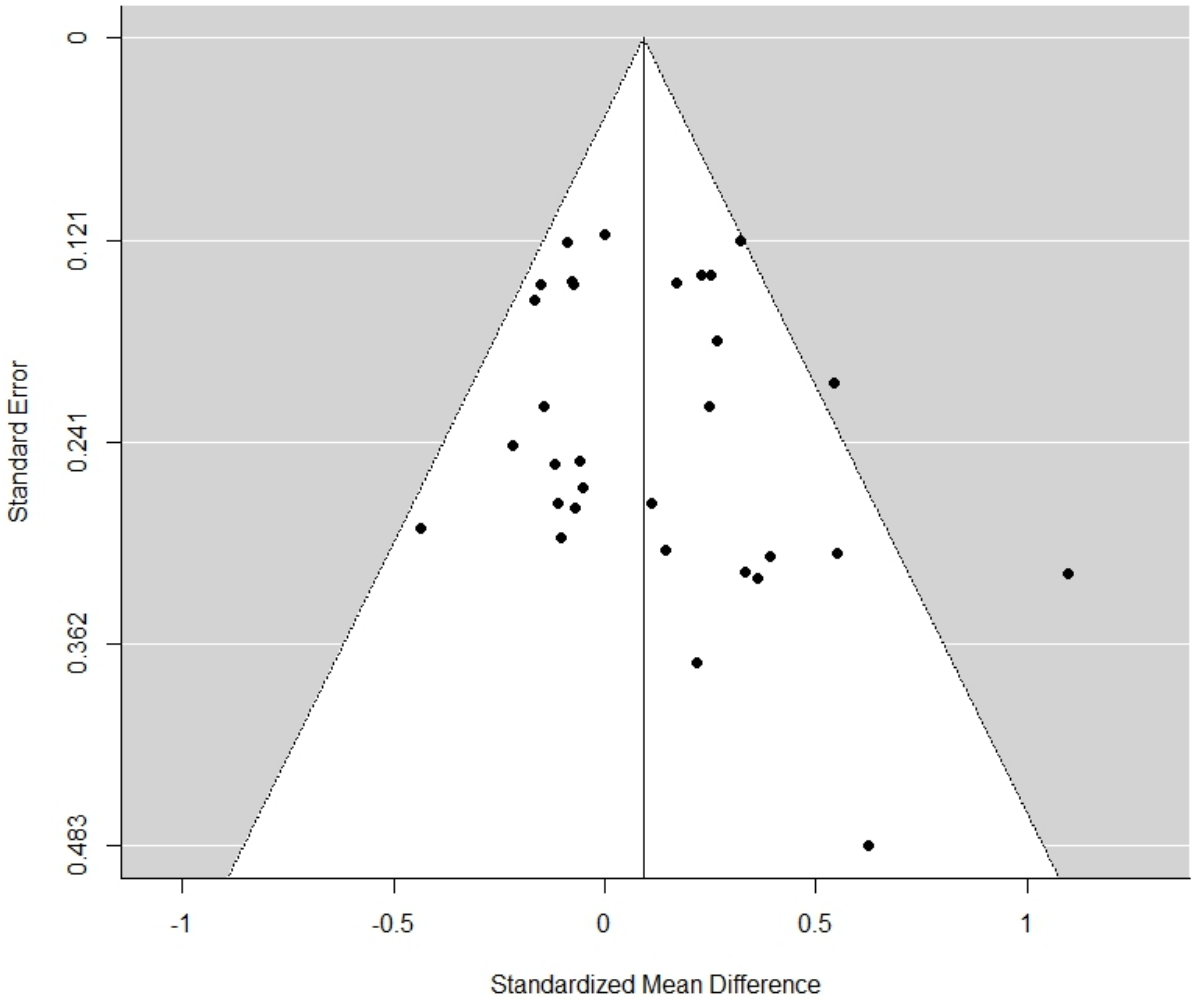

Figure 1: Funnel plot of effect sizes (Cohen's d) and their standard error for each study (which is a function of sample size). Each dot represents one study. The funnel plot shows little to no asymmetry, and a regression test for asymmetry was not significant (z=1.25, p=.21). This suggests that publication bias is not a problem.

Measures of Effect Size

Figure 2 is a forest plot of the effect sizes and confidence intervals of each study, as well as for each game type. Positive effect sizes indicate that offers were higher with lower stakes (i.e., in accordance with our predictions). Negative effect sizes indicate that offers were lower with lower stakes (i.e., contrary to predictions). A random-effects model was justified because the measure of heterogeneity, Q=47.27, d.f. = 30, p=0.0234, indicated significant heterogeneity.

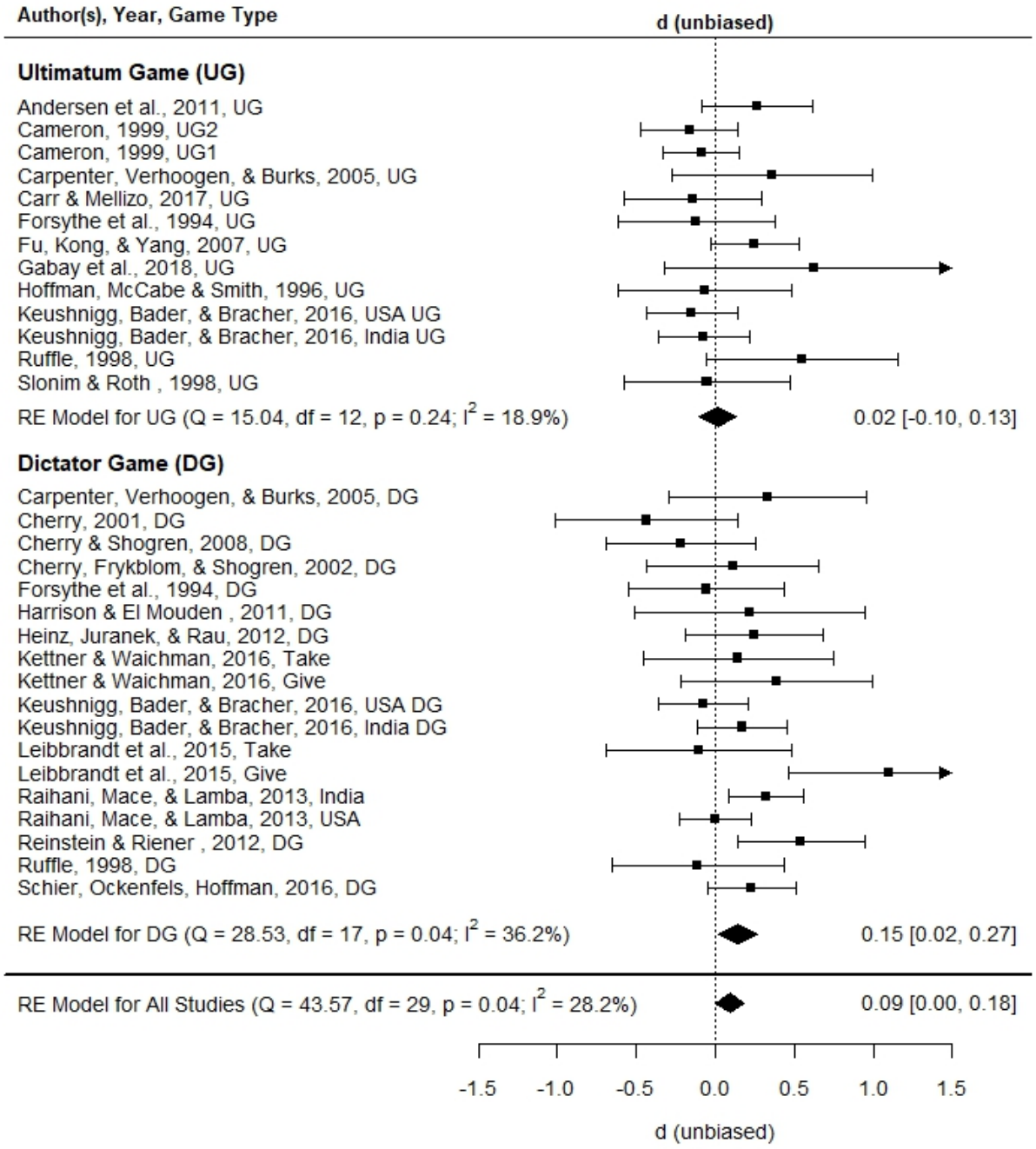

Figure 2: Forest plot of the random-effects meta-analysis of effect sizes (Cohen's *d*) with 95% confidence intervals for Ultimatum Games (top) and Dictator Games (bottom). Each dot represents the effect size of one study and the confidence intervals of that effect size; an arrow means that the confidence interval extends beyond the range of the graph. Diamonds represent the average effect size (top of the diamond) and 95% confidence interval of that average effect size (width of the diamond) within a category (i.e., within all UG, all DG, or overall). The vertical dashed line represents zero effect. Positive effect sizes indicate that offers are higher with low stakes.

Average Effect of Stake Size in UG and DG

The average effect size of stake size over all UG and DG studies (N_d=31*)* was *d*=0.091, 95% CI [0.002, 0.180], p=0.045. The positive value indicates that offers were significantly higher at lower stakes, albeit the effect size was small. Few studies reached statistical significance of α=.05 on their own, but when combined, the overall effect was significant at the α=.05 level. However, there was high heterogeneity among studies: τ^2=0.0195 (total heterogeneity), I^2=33.76% (total heterogeneity/total variability), H^2=1.51 (total variability/sampling variability). This heterogeneity is highly significant: Q=47.27, d.f.=30, p=.0234. To attempt to resolve this heterogeneity, we first analyzed each game separately, and then conducted a moderator analysis with game type (UG vs. DG) as a predictor.

Comparing stake size effects between UG and DG. Stakes had almost no effect in the thirteen UG studies: *d*= 0.017, 95% CI [-0.101, 0.135]. There was some heterogeneity among UG studies: τ^2=0.0086 (total heterogeneity), I^2=18.90% (total heterogeneity/total variability), H^2=1.23 (total variability/sampling variability), but this heterogeneity was not significant Q=15.04, d.f.=12, p=.239.

Stakes had a small but significant effect in the eighteen DG studies: *d*=0.145, 95% CI [0.022, 0.269]. DG studies were heterogeneous: τ^2=0.0231 (total heterogeneity), I^2=36.17% (total heterogeneity/total variability), H^2=1.57 (total variability/sampling variability), and this heterogeneity was significant Q=28.53, d.f.=17, p=.039.

We conducted a moderator analysis of all 31 studies with UG as the baseline and DG as the moderator. The intercept (UG) was not significant, with an estimated effect size of d=.021, s.e.=0.066, z=0.327, p=.744, 95% CI [-0.107, 0.150]. Dictator Games have an effect size greater than zero (d=0.145, s.e.=0.059, z=2.475, p=.013, 95% CI [0.030, 0.259]). The moderator (DG)

accounts for 21.19% of the heterogeneity, but the difference between Dictator Games and Ultimatum Games does not reach traditional significance: effect of moderator d=0.123, s.e.=0.088, z=1.404, p=.160, 95% CI [-0.0488, 0.296]. The 83% confidence interval for the moderator excludes zero [0.003, 0.244], suggesting that there may be an effect of the moderator, but we should place less confidence in this result than if it had reached the traditional $p<.05$. There was high residual heterogeneity among studies: $\tau^2=0.0154$ (total heterogeneity), $I^2=28.24\%$ (residual heterogeneity/total variability), $H^2=1.39$ (unaccounted variability/sampling variability), and this residual heterogeneity was significant Q=43.57, d.f.=29, p=.040. Thus, there may be other factors that causes differences among effect sizes (especially within DG studies).

Relationship of Effect Size to Difference in Stake Conditions Used

Does the effect of stakes depend on how different the stakes were? Studies varied widely in how different the stakes were, from a 1.75 times difference between lowest and highest stakes (Cherry, 2001) to a 100 times difference (Leibbrandt et al., 2015) or even a 1000 times difference (Andersen et al., 2011). These differences in stake sizes are on different orders of magnitude, so we addressed this non-linearity by taking the base 10 logarithm of the stake size differences. We then correlated the effect sizes of the various studies with the logs of stake differences (Figure 3). We note that the results are qualitatively similar if we use raw differences in stakes instead of log differences in stakes.

Across all studies, there was a small-medium correlation between effect size and log differences in stakes ($r_{29}=.201$), which was not significant due to sample size (p=.278). However, UG and DG differed in how the differences in stakes correlated with effect size. In the UG, there was a very weak and non-significant correlation between effect size and log difference in stakes ($r_{11}=.066$, p=.830). By contrast, in the DG there was a medium-large correlation between effect

size and log difference in stakes (r_{16}=.411, p=.090); this is significant with a directional one-tailed test (p=.045), which is justified because the correlation cannot meaningfully be negative (i.e., it would make no sense for stakes to have more effect in studies when high and low stakes are the same than in studies where high and low stakes are very different). Combined with the results of the meta-analysis, our results suggest that stake size affects DG offers, and the effect gets bigger as the stakes get more different. By contrast, the evidence suggests that stake size does not affect UG offers no matter how different the stakes are.

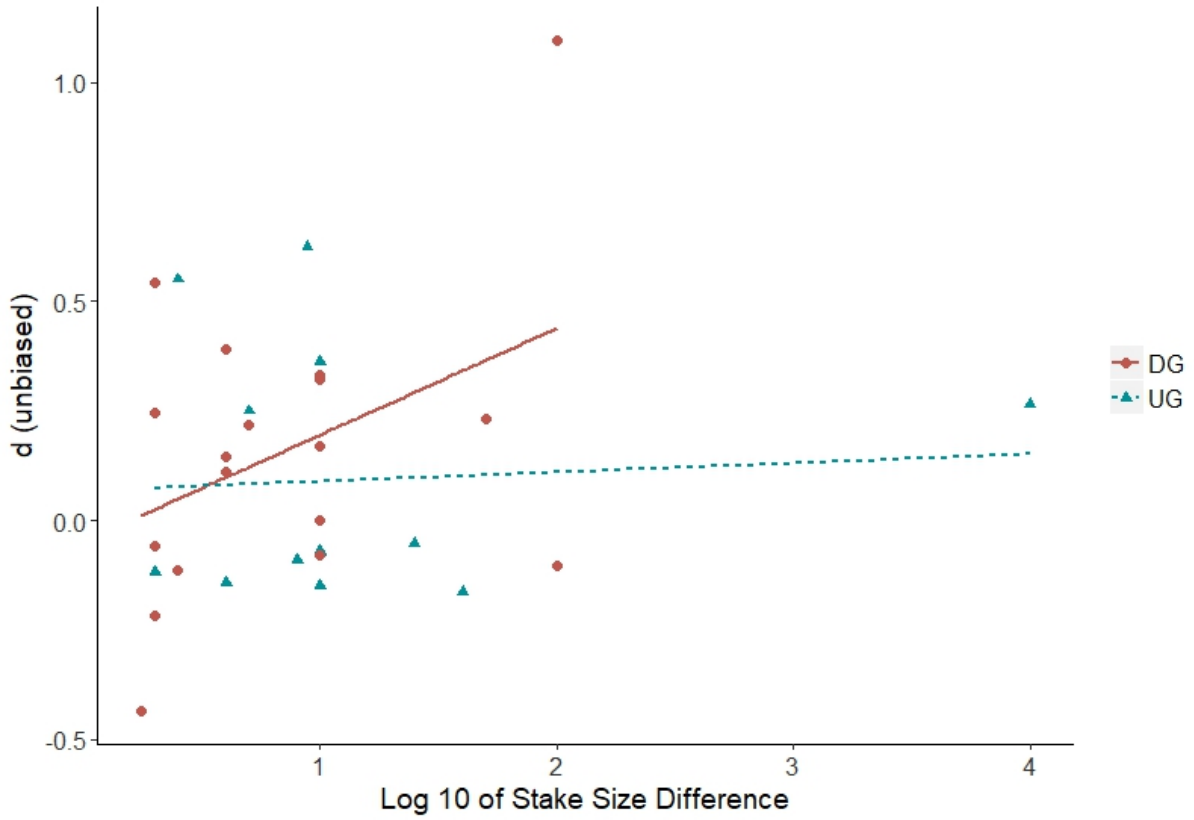

Figure 3: Scatterplot of the regression between the difference in stake size of a study and the effect size for that study (unbiased Cohen's d). Each red circle represents one Dictator Game (DG) study, and each blue triangle represents one Ultimatum Game (UG) study. The solid red line is the regression line for DG studies (estimated d = -0.048 + 0.243*(log10 of stake difference)), and the dashed blue line is the regression line for UG studies (estimated d = 0.068 + 0.021*(log10 of stake differences)). If we remove the UG outlier (Andersen et al., 2011), then the regression slope for the UG becomes negative.

Discussion

The purpose of this study was to consolidate data on stake size effects from a number of UG and DG studies, in an attempt to summarize the current state of knowledge about the effect of stake size on proposer behavior. This would be the first meta-study on this topic, to the best of our knowledge, that included only games with real stakes and controlled for additional confounds as outlined in our exclusion criteria (Table 1), to solely measure effects on one-shot, 2-player UG and DG studies. The hypotheses included that 1) as stake size increases, offer size will decrease in both the DG and UG owing to a higher cost of giving, and 2) DG offers will decrease more than UG offers because of the UG's added dimension of risk.

DG offers do decrease slightly with increasing stake size. The effect was small (d=0.14), but seems to get larger as the difference in stakes increases (r=.41). This pattern fits with Engel's (2011) meta-regression finding that DG studies with large stake size had slightly smaller offers than those with small stake size. These two meta-analyses differ in their goals and their methodologies: Engel (2011) sought to understand what factors predict DG offers and used all previous DG studies in a regression (i.e., not just studies on stake size), whereas the current meta-analysis specifically examined studies that experimentally manipulated stake size. Together, their different methodologies provide convergent evidence that stake size does affect DG offers, albeit weakly. Previous studies have claimed to find no effect of stake size, but may have had insufficient power to detect small effects (e.g., total sample size of 40 in Carpenter et al., 2005).

By contrast, there is no evidence that people offer less money in high-stakes UG. The effect of stakes is almost zero in UG, and there is little evidence that stakes have more effect when the difference in stakes is larger. This finding refutes studies that find an effect of stakes in

the UG, but supports studies that propose that the presence of risk aversion will prevent a substantial change in offers with increasing stakes (e.g. Holt & Laury, 2002). However, we must note that our results say nothing about responder behaviour in the UG – high stakes might reduce people's willingness to reject unfair UG offers, but this is beyond the scope of our study.

As for Hypothesis 2, stake size did have a bigger effect on DG offers than UG offers (difference in d=0.12), but the 95% confidence intervals for this effect included zero. The 83% confidence intervals excluded zero, which is suggestive, but far from conclusive. As such, we cannot be confident about this effect and will need more studies on stake size to determine if it exists. This difference might have been significant if there had been more high-stakes DGs, because studies with large differences in stakes found stronger effects of stakes in DGs. If future studies confirm this effect, it would support the idea that risk aversion prevents people from changing their offers at higher UG stakes, when compared to the DG (e.g. Holt & Laury, 2002).

Comparison to Other Stake Size Studies

Our study adds to the current literature investigating stake size effects in other bargaining and trust games (e.g., stake size effects in trust games, Johanssen-Stenman et al., 2005; no stake size effects in public goods games, Kocher et al., 2008; inconclusive stake size effects, Karagözoğlu & Urhan, 2016). Karagözoğlu and Urhan's (2016) survey of bargaining games included the DG and UG, and could not come to a conclusion about stake effects because of the wide variation in findings. They included studies of games other than the DG and UG, as well as those with confounds we were able to exclude. We found somewhat less variation in the current study, with the majority of effect sizes around the zero mark with some wide outliers (e.g., Leibbrandt et al., 2015, Give condition) but these studies did not have much weight on the

overall measure because of their small sample sizes. These consistencies allow for the relatively small confidence intervals around the average effect sizes we found.

Limitations

Sample sizes. Some studies had small sample sizes, which leads us to question the reliability of their findings. For example, Leibbrandt and colleagues (2015) had only 21 participants in the high stakes conditions in each of the DG and UG. This is understandable for budgetary reasons, but results in substantial sampling variation of what the "true" effect size is. For this reason, the meta-analytic software weighed each study's contribution to the overall effect size based on inverse variance, so that studies with larger samples were weighed more heavily (see weightings in Table 1). This means that those outliers with large effect sizes but small samples did not contribute as heavily to the overall effect size. This correction should have reduced these significant findings' potential to skew the results if, in fact, there is only a small effect size. Alternatively, this weighting and use of small sample size could have caused us to miss a larger effect of stakes. That these effects were on either side of zero, however, would have balanced out their strong effects to contribute to our finding of a near zero average for the UG.

Stake sizes. We only included real-stake experiments and excluded those with hypothetical stakes, because participants will take the former more seriously, whereas in the latter there are no consequences for losing money, and it is unclear whether people treat different hypothetical stakes differently (see also Hertwig & Ortmann, 2001). This resulted in some studies with relatively low stake sizes for the high-stake condition. The median high stakes condition was $20 in the DG and was $100 in the UG, which is somewhat high but not exorbitant. Only a few studies from Western societies used stakes more than US$20 (Carpenter et al., 2005; Carr & Mellizo, 2017; Cherry, 2001; Cherry et al., 2002; Gabay et al., 2018;

Hoffman et al., 1996), but studies in non-Western societies often had larger stakes in the local currency (Fu et al., 2007; Slonim & Roth, 1998), including some with more than three months' salary (Andersen et al., 2011; Cameron, 1999; Leibbrandt et al., 2015). Nevertheless, because there were not many studies with stake sizes of more than $100, we may have been limited in our ability to see an effect that may be present with much higher stakes.

In the DG, we found a medium-large correlation between the stake size difference in each study and the effects of stake size, such that studies with larger differences in stakes found larger effects of stakes. This correlation was significant at $p<.05$ using a directional one-tailed test, despite being underpowered (N=18). Perhaps we only found a small effect of stakes in the DG because most studies used only small differences in stakes. That being said, the regression equation suggests that even if stakes differ by two orders of magnitude (i.e., 100 times), the effect of stakes would only be d=0.44 (d = -0.048 + 0.243*2 = 0.438), which is a medium effect size. The regression equation suggests that stakes will only have a large effect (d=0.8) on DG offers when stakes differ by more than three orders of magnitude – we leave it to readers to decide how important that is in practice. In contrast to the results in DGs, in UGs there was only a weak and non-significant correlation between effect sizes and stake size differences, suggesting that using larger stake sizes would not change the effect sizes in the UG very much, if at all.

Practical Implications

Overall, from the 18 DG studies analyzed, these findings suggest that people give less money in DG as the stakes increase, but not in the UG. This finding means that researchers could anticipate seeing slightly lower offers from high-stakes DG dictators than they expect from general averages found at low stake conditions, because of an increase of the cost of giving that results in increased selfishness. Thus, depending on the aim of the study, researchers should take

caution when determining which stake size to use. Lower, budget-friendly stakes may not be as representative of dictator giving as higher stakes would be, but provide a reasonable approximation (given that stakes had only a small effect on dictator giving). By contrast, the near-zero effect of stakes in the UG means that stake size may be less important to consider in these games, perhaps because of the UG's added complexity (e.g., risk of rejection, anticipation of responder behavior).

In addition, the use of effect sizes allows us to see the size of any potential effects, however small, of stake size on offer behavior. When multiple studies find the same small effect size, we can be more confident that the effect is real, even if no single study reached statistical significance. In this way, the use of effect sizes may refute studies that have reported no effect of stakes based on a lack of statistical significance. For example, Carpenter et al., (2005) report a non-significant effect of stakes in the UG and DG, but a reinterpretation of their results in terms of Cohen's *d* reveals low to medium effects of increasing stakes on decreasing offers (e.g., d=0.33 in the DG). This indicates that individual studies reporting no effect of stakes sizes may (or may not) have found an effect that supports the current finding of a small effect size.

We should note that even if different stakes do change the absolute amounts given in Dictator Games, this should not be problematic for most experiments. If there are two experimental conditions, each with the same high or low stakes, then those stakes should have the same effect on the experimental and the control condition. Any differences between an experimental and a control condition are due to the experimental manipulation, not the stakes (which are constant in both conditions). Thus, while stake size affects *absolute* amounts in Dictator Games, there is no reason for it to affect *relative* amounts between two experimental conditions, and the latter is what matters in most experiments.

Future Directions

Future studies on stake size effects should attempt to ensure a large sample size to get a representative effect, as many of the currently available studies use relatively small samples. Representative sampling may be easier for contemporary studies with the use of sampling technology such as MTurk (e.g. Amir et al., 2012). The other option, to use hypothetical stakes, could be analyzed for stake size effects if the effect or lack of effect of hypothetical stakes becomes more clear in future studies. If hypothetical stakes were shown not to be a confound, their use to analyze stake size effects at much larger stakes may be justified.

Acknowledgments

We thank [removed for blinded review].

References

Amir, O., Rand, D. G., & Gal, Y. K. (2012). Economic games on the internet: The effect of $1 stakes. *PLoS ONE*, *7*(2). https://doi.org/10.1371/journal.pone.0031461

Andersen, S., Ertaç, S., Gneezy, U., Hoffman, M., & List, J. A. (2011). Stakes matter in ultimatum games. *American Economic Review*, *101*(7), 3427–3439. https://doi.org/10.1257/aer.101.7.3427

Andreoni, J., & Miller, J. (2002). Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism. *Econometrica*, *70*(2), 737–753. https://doi.org/10.1111/1468-0262.00302

Bardsley, N. (2008). Dictator game giving: Altruism or artefact? *Experimental Economics*, *11*(2), 122–133. https://doi.org/10.1007/s10683-007-9172-2

Barr, A., Burns, J., Miller, L., & Shaw, I. (2015). Economic status and acknowledgement of earned entitlement. *Journal of Economic Behavior & Organization, 118,* 55-68. http://dx.doi.org/10.1016/j.jebo.2015.02.012

Batista, C., Silverman, D., & Yang, D. (2015). Directed giving: Evidence from an inter-household transfer experiment. *Journal of Economic Behavior & Organization*, *118*, 2–21. https://doi.org/10.1016/j.jebo.2015.03.008

Bechler, C. (2013). *Choices in Two-Person Interactions: The Effect of Amount and Social Distance on Offers in the Dictator and Ultimatum Games* (Unpublished thesis). Washington University. Retrieved from https://economics.wustl.edu/files/economics/imce/christopher_bechler_2013.pdf

Bechler, C., Green, L., & Myerson, J. (2015). Proportion offered in the Dictator and Ultimatum Games decreases with amount and social distance. *Behavioural Processes*, *115*, 149–155. https://doi.org/10.1016/j.beproc.2015.04.003

Bediou, B., Sacharin, V., Hill, C., Sander, D., & Scherer, K. R. (2012). Sharing the fruit of labor: Flexible application of justice principles in an ultimatum game with joint-production. *Social Justice Research, 25,* 25-40. DOI 10.1007/s11211-012-0151-1

Bekkers, R. (2007). Measuring altruistic behavior in surveys: The all-or-nothing dictator game. *Survey Research Methods*, *1*(3), 139–144. https://doi.org/10.18148/srm/2007.v1i3.54

Ben-Ner, A., Kramer, A., & Levy, O. (2008). Economic and hypothetical dictator game experiments: Incentive effects at the individual level. *Journal of Socio-Economics*, *37*(5), 1775–1784. https://doi.org/10.1016/j.socec.2007.11.004

Bethwaite, J. and Tompkinson, P. (1993). The ultimatum game-Understanding the taste for fairness. *Economic Notes, 22*(1), 37-48.

Bhogal, M. S., Galbraith, N., & Manktelow, K. (2016). Physical attractiveness and altruism in two modified dictator games. *Basic and Applied Social Psychology, 38*(4), 212-222. http://dx.doi.org.subzero.lib.uoguelph.ca/10.1080/01973533.2016.1199382

Blake, P. R., & Rand, D. G. (2010). Currency value moderates equity preference among young children. *Evolution and Human Behavior*, *31*(3), 210–218. https://doi.org/10.1016/j.evolhumbehav.2009.06.012

Bolton, G. E., Katok, E., & Zwick, R. (1998). Dictator game giving: Rules of fairness versus acts of kindness. *International Journal of Game Theory*, *27*(2), 269–299. https://doi.org/10.1007/s001820050072

Bühren, C., & Kundt, T. C. (2015). Imagine being a nice guy: A note on hypothetical vs. incentivized social preferences. *Judgment and Decision Making*, *10*(2), 185–190. Retrieved from papers3://publication/uuid/1723F6F0-9A06-4B61-9307-6B885E01CF84

Busch, J., & Krishna, A. (unpublished). Shifting utility focus – how emphasizing comparative or categorical utility influences the impact of social norms in Ultimatum Games. Unpublished BSc thesis, 2015, Julius Maximilians Universität Würzburg.

Cameron, L. a. (1999). Raising the stakes in the ultimatum game: Experimental evidence from Indonesia. *Economic Inquiry*, *37*(1), 47–59. https://doi.org/10.1111/j.1465-7295.1999.tb01415.x

Carr, M. D., & Mellizo, P. (2017). The effect of endogenous endowments: evidence from a mini-ultimatum game. *Economics Bulletin*, *37*, 2552-2560.

Carpenter, J., Verhoogen, E., & Burks, S. (2005). The effect of stakes in distribution experiments. *Economics Letters*, *86*(3), 393–398. https://doi.org/10.1016/j.econlet.2004.08.007

Chang, S-C., Lin, L-Y., Horng, R-Y., & Wang, Y-D. (2014). The effect of amount and tangibility of endowment and certainty of recipients on selfishness in a modified dictator game. *Psychological Reports, 114*(3), 720-739. https://doi.org/10.2466/31.01.PR0.114k24w8

Charness, G., & Rabin, M. (2002). Understanding Social Preferences with Simple Tests. *The Quarterly Journal of Economics*, *117*(3), 817–869. https://doi.org/10.1162/003355302760193904

Chen, C-C., Chiu, I-M., Smith, J., & Yamada, T. (2013). Too smart to be selfish? Measures of cognitive ability, social preferences, and consistency. *Journal of Economic Behavior & Organization*, *90*, 112–122. https://doi.org/10.1016/j.jebo.2013.03.032

Cherry, T. L. (2001). Mental accounting and other-regarding behavior: Evidence from the lab. *Journal of Economic Psychology*, *22*(5), 605–615. https://doi.org/10.1016/S0167-4870(01)00058-7

Cherry, T. L., Frykblom, P., & Shogren, J. F. (2002). Hardnose the dictator. *American Economic Review*, *92*(4), 1218–1221. https://doi.org/10.1257/00028280260344740

Cherry, T. L., & Shogren, J. F. (2008). Self-interest, sympathy and the origin of endowments. *Economics Letters*, *101*(1), 69–72. https://doi.org/10.1016/j.econlet.2008.04.007

Crockett, M. J., Clark, L., Tabibnia, G., Lieberman, M. D., & Robbins, T. W. (2008). Serotonin modulates behavioral reactions to unfairness. *Science*, *320*(5884), 1739. https://doi.org/10.1126/science.1155577

Cumming, G. (2012). *Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis* (Multivariate Applications Series). London: Taylor and Francis.

Dalbert, C., & Umlauft, S. (2009). The role of the justice motive in economic decision making. *Journal of Economic Psychology*, *30*(2), 172–180. https://doi.org/10.1016/j.joep.2008.07.006

De Bruyn, a., & Bolton, G. E. (2008). Estimating the Influence of Fairness on Bargaining Behavior. *Management Science*, *54*(10), 1774–1791. https://doi.org/10.1287/mnsc.1080.0887

Dickinson, D. L. (2000). Ultimatum decision-making: A test of reciprocal kindness. *Theory and Decision*, *48*(2), 151–177. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-0007180856&partnerID=tZOtx3y1

Diekmann, A. (2004). The Power of Reciprocity: Fairness, Reciprocity, and Stakes in Variants of the Dictator Game. *The Journal of Conflict Resolution*, *48*(4), 487–505. https://doi.org/10.1177/0022002704265948

El Harbi, S., Bekir, I., Grolleau, G., & Sutan, A. (2015). Efficiency, equality, positionality: What do people maximize? Experimental vs. hypothetical evidence from Tunisia. *Journal of Economic Psychology*, *47*, 77–84. https://doi.org/10.1016/j.joep.2015.01.007

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, *14*(4), 583–610. https://doi.org/10.1007/s10683-011-9283-7

Fehr, E., Tougareva, E., & Fischbacher, U. (2014). Do high stakes and competition undermine fair behaviour? Evidence from Russia. *Journal of Economic Behavior & Organization*, *108*, 354-363.

Fiala, J., Starý, O., Fialová, M., Holasová, A., Mejzlíková, T., & Bems, J. (2016). Value perception in the ultimatum game: A blinded randomized trial. *Ekonomicky Casopis, 64*(6), 519-538. Retrieved from http://sfx.scholarsportal.info/guelph/docview/1864074495?accountid=11233

Fisman, R., Kariv, S., & Markovits, D. (2007). Individual preferences for giving. *American Economic Review*, *97*(5), 1858–1876. https://doi.org/10.1257/aer.97.5.1858

Freiburg, L., & Krishna, A. (unpublished). Fairness-sensitivity in ultimatum games – how bargaining over losses and regulatory focus interact. Unpublished BSc thesis, 2016, Julius Maximilians Universität Würzburg.

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in Simple Bargaining Experiments. *Games and Economic Behavior*. https://doi.org/10.1006/game.1994.1021

Fu, T.-T., Kong, W.-H. & Yang, C. (2007). Monetary stakes and socioeconomic characteristics in ultimatum games: An experiment with nation-wide representative subjects, Working paper.

Gabay, A.S., Carhart-Harris, R.L., Mazibuko, N., Kempton, M.J., Morrison, P.D., Nutt, D.J., & Mehta, M.A. (2018). Psilocybin and MDMA reduce costly punishment in the Ultimatum Game. *Scientific Reports*, *8*, 8236.

Greitemann, J., & Krishna, A. (unpublished). Und tschüss rationalität: der einfluss impulsiver verarbeitung im Ultimatumspiel. Unpublished BSc thesis, 2015, Julius Maximilians Universität Würzburg.

Grossman, P. J., & Eckel, C. C. (2015). Giving versus taking for a cause. *Economic Letters, 123,* 28-30. http://dx.doi.org.subzero.lib.uoguelph.ca/10.1016/j.econlet.2015.04.002

Güth, W. (2010). The Generosity Game and calibration of inequity aversion. *The Journal of Socio-Economics*, *39*(2), 155–157. https://doi.org/10.1016/j.socec.2009.10.012

Güth, W., & Kirchkamp, O. (2012). Will you accept without knowing what? The Yes-No game in the newspaper and in the lab. *Experimental Economics*, *15*(4), 656–666. https://doi.org/10.1007/s10683-012-9319-7

Güth, W., & Kocher, M. G. (2013). More than thirty years of ultimatum bargaining experiments: Motives, variations, and a survey of the recent literature. *Jena Economic Research Papers, 035,* 1-36. http://hdl.handle.net/10419/85032

Güth, W., Levati, M., & Ploner, M. (2012). An experimental study of the generosity game. *Theory and Decision, 72*(1), 51-63. doi:10.1007/s11238-011-9253-8

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization*, *3*(4), 367–388. https://doi.org/10.1016/0167-2681(82)90011-7

Güth W., Tietz R. (1988) Ultimatum Bargaining for a Shrinking Cake — An Experimental Analysis. *Bounded Rational Behavior in Experimental Games and Markets, 314,* 111-128. doi: 10.1007/978-3-642-48356-1_9

Haas, K. (2009). *Changing endowment sizes and prices of giving in ultimatum games* (Unpublished thesis). University of California, Berkeley. Retrieved from http://www.econ.berkeley.edu/sites/default/files/kevin_haas_thesis.pdf

Halali, E., Bereby-Meyer, Y., & Ockenfels, A. (2013). Is it all about the self? The effect of self-control depletion on ultimatum game proposers. *Frontiers in Human Neuroscience*, *7*, 240. http://doi.org/10.3389/fnhum.2013.00240

Harrison, G. W., & Rutström, E. (2002). Do Higher Stakes Change Behavior in Ultimatum Games? *Unpublished manuscript*, Moore School of Business, University of South Carolina.

Harrison, F., & El Mouden, C. (2011). Exploring the effects of working for endowments on behaviour in standard economic games. *PLoS ONE*, *6*(11), e27623 -e27623. https://doi.org/10.1371/journal.pone.0027623

Heinz, M., Juranek, S., & Rau, H.A. (2012). Do women behave more reciprocally than men? Gender differences in real effort dictator games. *Journal of Economic Behavior and Organization*, *83*, 105-110.

Henrich, J., Ensminger, J., Mcelreath, R., Barr, A., Barrett, C., Bolyanatz, A., Camilo Cardenas, J., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., & Ziker, J. (2010). Markets, religion, community size, and the evolution of fairness and punishment. *Science*, *327*(*1480*), 1-94. DOI: 10.1126/science.1182238

Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: a methodological challenge for psychologists? *Behavioral and Brain Sciences*, *24*, 383-451.

Hoffman, E., McCabe, K. a., & Smith, V. L. (1996). On expectations and the monetary stakes in ultimatum games. *International Journal of Game Theory*, *25*(3), 289–301. https://doi.org/10.1007/BF02425259

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, *92*(5), 1644–1655. https://doi.org/10.1257/000282802762024700

Hou, Y., Zhao, L., Yao, Q., & Ding, L. (2016). Altered economic decision-making in abstinent heroin addicts: Evidence from the ultimatum game. *Neuroscience Letters*, *627*, 148–154. https://doi.org/10.1016/j.neulet.2016.06.002

John, K., & Thomsen, S. L. (2015). School-track environment or endowment: What determines different other-regarding behavior across peer groups? *Games and Economic Behavior*, *94*, 122–141. https://doi.org/10.1016/j.geb.2015.10.007

Jordan, J., McAuliffe, K., & Rand, D. (2015). The effects of endowment size and strategy method on third party punishment. *Experimental Economics*, *19*(4), 741–763. https://doi.org/10.1007/s10683-015-9466-8

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986). Fairness and the Assumptions of Economics. *The Journal of Business*, *59*(S4), S285. https://doi.org/10.1086/296367

Karagözoğlu, E., & Urhan, Ü B. (2016). The Effect of Stake Size in Experimental Bargaining and Distribution Games: A Survey. *Group Decision and Negotiation,26*(2), 285-325. doi:10.1007/s10726-016-9490-x

Kench, B. T., & Niman, N. B. Of altruists and thieves. *Eastern Economic Journal, 36*(3), 317-343. doi:10.1057/eej.2009.10

673 Kettner, S. E., & Waichman, I. (2016). Old age and prosocial behavior: Social preferences or
674 experimental confounds? *Journal of Economic Psychology*, *53*, 118–130.
675 https://doi.org/10.1016/j.joep.2016.01.003
676 Keuschnigg, M., Bader, F., & Bracher, J. (2016). Using crowdsourced online experiments to
677 study context-dependency of behavior. *Social Science Research*, *59*, 68–82.
678 https://doi.org/10.1016/j.ssresearch.2016.04.014
679 Klaffehn, A.L., & Krishna, A. (unpublished). It's not fair, but do I care? The influence of
680 impulsivity and attentional focus on Ultimatum Game decisions. Unpublished BSc thesis,
681 2016, Julius Maximilians Universität Würzburg.
682 Kocher, M.G., Martinsson, P., & Visser, M. (2008). Does stake size matter for cooperation and
683 punishment? *Economics Letters*, *99*, 508-511.
684 Kogut, T. (2012). Knowing what I should, doing what I want: From selfishness to inequity
685 aversion in young children's sharing behavior. *Journal of Economic Psychology*, *33*(1),
686 226–236. https://doi.org/10.1016/j.joep.2011.10.003
687 Korenok, O., Millner, E. L., & Razzolini, L. (2012). Are dictators averse to inequality? *Journal*
688 *of Economic Behavior and Organization*, *82*(2–3), 543–547.
689 https://doi.org/10.1016/j.jebo.2012.03.009
690 Korenok, O., Millner, E. L., & Razzolini, L. (2013). Impure altruism in dictators' giving. *Journal*
691 *of Public Economics*, *97*(1), 1–8. https://doi.org/10.1016/j.jpubeco.2012.08.006
692 Kriss, P. H., Nagel, R., & Weber, R. A. (2013). Implicit vs. Explicit deception in ultimatum
693 games with incomplete information. *Journal of Economic Behavior and Organization*, *93*,
694 337–346. https://doi.org/10.1016/j.jebo.2013.03.024
695 Lee, C. C., & Lau, W. K. (2013). Information in repeated ultimatum game with unknown pie
696 size. *Economics Research International*. Retrieved from
697 http://dx.doi.org.subzero.lib.uoguelph.ca/10.1155/2013/470412
698 Leibbrandt, A., Maitra, P., & Neelim, A. (2015). On the redistribution of wealth in a developing
699 country: experiment evidence on stake and framing effects. *Journal of Economic Behavior*
700 *and Organization*, *118*, 360-371.
701 Limback, E. *Influences on preschoolers' altruism* (Unpublished doctoral dissertation). University
702 of Nottingham. Retrieved from http://eprints.nottingham.ac.uk/27679/1/576485.pdf
703 List, J. A. (2007). On the Interpretation of Giving in Dictator Games. *Journal of Political*
704 *Economy, 115*(3). 482-493. Retrieved from
705 http://home.uchicago.edu/jlist/papers/dictator_game.pdf
706 List, J. A., & Cherry, T. L. (2008). Examining the role of fairness in high stakes allocation
707 decisions. *Journal of Economic Behavior and Organization*, *65*(1), 1–8.
708 https://doi.org/10.1016/j.jebo.2003.09.021
709 Marwell, G., & Ames, R.E. (1980). Experiments on the provision of public goods. Provision
710 points, stakes, experience, and the free-rider problem. *American Journal of Sociology, 85*,
711 926–937. Retrieved from
712 https://www.jstor.org/stable/2778712?seq=6#page_scan_tab_contents
713 Mitzkewitz, M., & Nagel, R. (1993). Experimental results on ultimatum games with incomplete
714 information. *International Journal of Game Theory*, *22*(2), 171–198.
715 https://doi.org/10.1007/BF01243649
716 Munier, B., & Zaharia, C. (2002). High stakes and acceptance behavior in ultimatum bargaining:
717 A contribution from an international experiment. *Theory and Decision*, *53*(3), 187–207.
718 https://doi.org/10.1023/A:1022815832351

719 Neelin, J., Sonnenschein, H., & Spiegel, M. (1988). A further test of noncooperative bargaining
720 theory: Comment. *American Economic Review*, *78*(4), 824.
721 Neilson, W. S. (2009). A theory of kindness, reluctance, and shame for social preferences.
722 *Games and Economic Behavior*, *66*(1), 394–403. https://doi.org/10.1016/j.geb.2008.04.004
723 Novakova, J., & Flegr, J. (2013). How much is our fairness worth? The effect of raising stakes
724 on offers by proposers and minimum acceptable offers in dictator and ultimatum games.
725 *PLoS ONE*, *8*(4). https://doi.org/10.1371/journal.pone.0060966
726 Oberholzer-Gee, F., Waldfogel, J., & White, M.W. (2010). Friend or foe? Cooperation and
727 learning in high-stakes games. *Review of Economics and Statistics*, *92*, 179-187.
728 Ockenfels, A., & Werner, P. (2012). "Hiding behind a small cake" in a newspaper dictator game.
729 *Journal of Economic Behavior and Organization*, *82*(1), 82–85.
730 https://doi.org/10.1016/j.jebo.2011.12.008
731 Ploner, M., & Regner, T. (2013). Self-image and moral balancing: An experimental analysis.
732 *Journal of Economic Behavior and Organization*, *93*, 374–383.
733 https://doi.org/10.1016/j.jebo.2013.03.030
734 Posid, T., Fazio, A., & Cordes, S. (2015). Being sticker rich: Numerical context influences
735 children's sharing behavior. *PLoS ONE*, *10*(11).
736 https://doi.org/10.1371/journal.pone.0138928
737 R Core Team (2013). R: A language and environment for statistical computing. R Foundation for
738 Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.
739 Rabin, M. (1993). Incorporating Fairness into Game Theory and Economics. *The American*
740 *Economic Review*. https://doi.org/10.1257/aer.91.4.1180
741 Raihani, N. J., Mace, R., & Lamba, S. (2013). The Effect of $1, $5 and $10 Stakes in an Online
742 Dictator Game. *PLoS ONE, 8*(8), E73131. doi:10.1371/journal.pone.0073131
743 Rapoport, A., & Sundali, J. A. (1996). Ultimatums in two-person bargaining with one-sided
744 uncertainty: Offer games. *International Journal of Game Theory*, *25*(4), 475-494. doi:
745 10.1007/BF01803952
746 Rapoport, A., Sundali, J. A., & Seale, D. A. (1996). Ultimatums in two-person bargaining with
747 one-sided uncertainty: Demand games. *Journal of Economic Behavior and Organization*,
748 *30*(2), 173–196. https://doi.org/10.1016/S0167-2681(96)00856-6
749 Reinstein, D., & Reimer, G. (2012). Decomposing desert and tangibility effects in a charitable
750 giving experiment. *Experimental Economics*, *15*, 229-240.
751 Rese, M. & Schons, L. M. (2013) Norm Enforcement in high-cost-situations: An experimental
752 investigation of the effect of an economic constraint on responder behavior in the
753 ultimatum game. *Zeitschrift fur Soziologie*, *42*(6), 463-482.
754 Roth, A., Prasnikar, V., Okuno-Fujiwara, M., & Zamir, S. (1991). Bargaining and Market
755 Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study.
756 *American Economic Review*, *81*(5), 1068–95. https://doi.org/10.1126/science.151.3712.867-
757 a
758 Ruffle, B.J. (1998). More is better, but fair is fair: tipping in Dictator and Ultimatum Games.
759 *Games and Economic Behavior*, *23*, 247-265.
760 Schier, U. K., Ockenfels, A., & Hofmann, W. (2016). Moral values and increasing stakes in a
761 dictator game. *Journal of Economic Psychology*, *56*, 107–115.
762 https://doi.org/10.1016/j.joep.2016.06.004

Schulz, J. F., Fischbacher, U., Thöni, C., & Utikal, V. (2011). Affect and fairness: Dictator games under cognitive load. *Journal of Economic Psychology*, *68*, 1-26. Retrieved from http://kops.uni-konstanz.de/handle/123456789/16205

Slonim, R., & Roth, A. E. (1998). Learning in high stakes ultimatum games: an experiment in the Slovak Republic. *Econometrica*, *66*(3), 569–596. https://doi.org/10.2307/2998575

Straub, P. G., & Murnighan, J. K. (1995). An experimental investigation of ultimatum games: information, fairness, expectations, and lowest acceptable offers. *Journal of Economic Behavior and Organization*, *27*(3), 345–364. https://doi.org/10.1016/0167-2681(94)00072-M

Sundelin, T., & Axelsson, J. (unpublished data). [No title]

Tompkinson, P., & Bethwaite, J. (1995). The ultimatum game: raising the stakes. *Journal of Economic Behavior and Organization*, *27*(3), 439–451. https://doi.org/10.1016/0167-2681(94)00035-D

Tonin, M., & Vlassopoulos, M. (2017). Sharing one's fortune? An experimental study on earned income and giving. *Journal of Economic Behavior & Organization*, *66*, 112-118.

Van Donge, K. (2015). *Changing endowment sizes and prices of giving in ultimatum games* (Unpublished thesis). University of California, Berkeley. Retrieved from https://www.econ.berkeley.edu/sites/default/files/Kent%20Van%20Donge%20-%20Thesis.pdf

Van Koten, S., Ortmann, A., & Babicky, V. (2013). Fairness in Risky Environments: Theory and Evidence. *Games*, *4*(2), 208–242. https://doi.org/10.3390/g4020208

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1-48. URL: http://www.jstatsoft.org/v36/i03/

Wang, X., Chen, X., & Wang, L. (2014). Random allocation of pies promotes the evolution of fairness in the Ultimatum Game. *Scientific Reports*, *4*(4534). https://doi.org/10.1038/srep04534

Yamagishi, T., Li, Y., Matsumoto, Y., & Kiyonari, T. (2016). Moral Bargain Hunters Purchase Moral Righteousness When it is Cheap: Within-Individual Effect of Stake Size in Economic Games. *Scientific Reports*, *6*(February), 27824. https://doi.org/10.1038/srep27824

Zhou, Y., Wang, Y., Rao, L-L., Yang, L-Q., & Li S. (2014). Money talks: neural substrate of modulation of fairness by monetary incentives. *Frontiers in Behavioral Neuroscience, 8*:150. doi: 10.3389/fnbeh.2014.00150

801

Appendix

Table 3

Summary of Excluded Studies and Reason for Exclusion

Authors	Type of Game	Reason for Exclusion
Amir, Rand & Kobi Gal (2012)	DG/UG	Hypothetical versus real stakes instead of varying the real stake level
Andersen et al. (2011), Wealth condition	UG	Earned stakes (data from No Wealth condition were included)
Andreoni & Miller (2002)	DG variant	Non-equivalence of tokens across conditions because worth different amounts to proposer & responder
Bardsley (2008)	DG	Give vs. take frame; give frame had responders start with an endowment to vary inequality
Barr et al. (2015)	DG	4-player DG; earned vs. unearned stakes
Batista, Silverman & Yang (2015)	DG	Data unavailable and unable to contact
Bechler (2013)	DG/UG	Hypothetical stakes
Bechler, Green & Myerson (2015)	DG/UG	Hypothetical stakes
Bediou et al. (2012)	DG	Earned stakes, no control condition
Bekkers (2007)	DG variant	Earned endowments
Ben-Ner, Kramer & Levy (2008)	DG	Hypothetical vs. real stakes instead of varying the real stake level
Bethwaite & Thompkinson (1993)	UG	Data unavailable and unable to contact any authors
Bhogal, Galbraith & Manktelow (2016)	DG	Earned monetary vs. non-monetary stakes
Blake & Rand (2010)	DG	Participants were children
Bolton, Katok & Zwick (1998)	DG	Stakes did not vary across conditions; repeated interactions
Bühren & Kundt (2015)	DG	Hypothetical vs. real stakes instead of varying the real stake level
Busch & Krishna (unpublished)	UG	Participants primed before decisions
Carr & Mellizo (2017), Responder Produces and Responder Gambles conditions	UG	Responder earns the endowments (data from Exogenous condition were included)
Chang et al. (2014)	DG	Hypothetical stakes
Charness & Rabin (2002)	DG	Two options given for a single offer/round, but each option totaled to a different stake size
Chen et al. (2013)	DG/SVO	Stakes only differ between the DG and SVO games

Cherry (2001), Earned Money condition	DG	Earned stakes (data from Allocated Money condition were included)
Cherry, Frykblon & Shorgen (2002), Blind and Double Blind with Earnings conditions	DG	Earned endowments (data from Baseline were included)
Cherry & Shogren (2008), Earned Endowment condition	DG	Earned stakes (data from Windfall condition were included)
Crockett et al. (2008)	UG	Data unavailable at time of submission
Dalbert & Umlauft (2009)	DG	Hypothetical vs. real stakes instead of varying the real stake level
De Bruyn & Bolton (2008)	-	Non-experimental model of giving in bargaining games
Dickinson (2000)	UG	Non-experimental model based on previous experimental results
Diekmann (2004)	DG variant	Priming before task: proposers given offers from computerized proposer to examine reciprocity in future games
El Harbi et al. (2015)	DG	Participants are given 3 alternative ratios to split the stake (5/7, 4/1, 3/3) that do not clearly reflect stake size effects
Engel (2011)	DG	Non-experimental/meta-analysis, measure of effect included studies that warrant exclusion here
Fehr, Tougareva, & Fischbacher (2014)	-	Game is neither a DG nor UG
Fiala et al. (2016)	UG	Hypothetical vs. real stakes instead of varying the real stake level
Fisman, Kariv & Markovitz (2007)	DG	Data unavailable and unable to contact
Freiburg & Krishna (unpublished)	UG	Participants primed before decisions
Greitemann & Krishna (unpublished)	UG	No data on proposers, only responders
Grossman & Eckel (2015)	DG	Give vs. take frame, no control or stake size variation
Güth (2010)	-	Neither DG/UG; non-experimental model
Güth & Kirchkamp (2012)	UG variant	Yes-No game: responder did not know stake size
Güth, Levati & Ploner (2012)	DG/UG	Proposer chooses stake size in all conditions
Güth & Tietz (1988)	UG	Responder could make counter-offer, which changes proposer's strategies
Haas (2009)	UG	Data unavailable and could not contact author
Halali, Bereby-Meyer & Ockenfels (2013)	DG/UG	Data unavailable and unable to contact
Harrison & El Mouden (2011), T1 and T2 conditions	DG	Earned stakes (data from M1 condition were included)
Harrison & Rutström (2002)	UG	Unpublished manuscript, not accessible online and authors do not have it

Heinz Juranek Rau (2012), Real Effort condition	DG	Earned stakes (data from Windfall condition were included)
Henrich et al. (2010)	DG	Cross-cultural study where all stake sizes are equivalent in purchasing power
Holt & Laury (2002)	-	Neither a DG or UG
Hou et al. (2016)	UG	Data unavailable and could not contact any authors
John & Thomsen (2015)	UG	Participants were children
Jordan, McAuliffe & Rand (2015)	DG	Included 3rd party punishment
Karagözoğlu & Urhan (2016)	DG/UG	Non-experimental/meta-analysis, measure of effect included studies that warrant exclusion here
Kench & Niman (2010)	DG	Earned stakes, no control condition
Klaffehn & Krishna (unpublished)	UG	Stakes so small (<€0.01) as to be hypothetical
Kocher, Martinsson & Visser (2008)	-	Game is neither a DG nor UG
Korenok, Millner & Razzolini (2012)	DG	Responders started with an endowment to vary inequality; control conditions do not vary in stake size
Korenok, Millner & Razzolini (2013)	DG	Responders started with an endowment to vary inequality; control conditions do not vary in stake size
Kriss, Nagel & Weber (2013)	UG	Responders did not know stake size, no control conditions present
Lee & Lau (2013)	UG	Responders did not know stake size, no control conditions present
Limback (2012)	DG	Participants were children
List (2007)	DG	Responders started with an endowment to vary inequality
List & Cherry (2008)	DG	Data not provided and authors did not have them
Marwell & Ames (1980)	-	Neither DG/UG
Mitzkewitz & Nagel (1993)	UG	Responder did not know stake size, no control conditions
Munier & Zaharia (2003)	UG	Stake size was confounded with order effects (stake size varied within participants, but order was not counterbalanced)
Neelin (1988)	-	Neither a DG or UG
Neilson (2009)	DG	Non-experimental model
Novakova & Flegr (2013)	DG/UG	Hypothetical stakes only

Oberholzer-Gee, Waldfogel & White (2010)	-	Game is neither a DG nor UG
Ockenfels & Werner (2012)	DG	Repeated interactions: participants made offers for each stake condition at the same time
Ploner & Regner (2013)	DG	Stakes earned by rolling a die
Posid, Fazio & Cordes (2015)	DG	Participants were children
Rapoport & Sundali (1996)	UG	Responder did not know stake size/no control conditions
Rapoport, Sundali & Seale (1996)	UG	Responder did not know stake size/no control conditions
Reinstein & Reiner (2012), Performance condition	DG	Earned stakes (data from Random condition were included)
Rese & Schons (2013)	UG	Repeated interactions; Computer proposer
Roth et al. (1991)	UG	Cross-cultural study where all stake sizes are equivalent in purchasing power
Ruffle (1998), Hypothetical and Skill conditions	DG/UG	Either not real stakes or earned stakes (data from Real and Coin conditions were included)
Schulz et al. (2011)	DG	Stake value did not vary across conditions; repeated interactions
Straub & Murnighan (1995)	UG	Data unavailable and unable to contact any authors
Sundelin & Axelsson (unpublished)	UG	No data for proposers, only responders
Tompkinson & Bethwaite (1995)	UG	Data unavailable and unable to contact any authors
Tonin & Vlassopoulos (2017)	-	Game is neither a DG nor UG (giving to charity); also earned endowments
Van Donge (2015)	-	Non-experimental prediction model
VanKoten, Ortmann & Babicky (2013)	DG/UG	Stake consisted of lottery tickets with different levels of risk, cannot define stake sizes independent of risk manipulation
Wang, Chen & Wang (2014)	UG	Non-experimental prediction model
Yamagashi et al. (2016)	PDG	Neither DG/UG
Zhou et al. (2014)	UG	No data for proposers, only responders; does not distinguish between computer and human proposers

802

803