

Many Labs 5: Replication of van Dijk, van Kleef, Steinel, and van Beest (2008)



Lauren Skorb¹, Balazs Aczel², Bence E. Bakos², Lily Feinberg¹, Ewa Hałasa³, Mathias Kauff⁴, Marton Kovacs², Karolina Krasuska³, Katarzyna Kuchno³, Dylan Manfredi⁵, Andres Montealegre⁶, Emilian Pękala³, Damian Pieńkosz³, Jonathan Ravid¹, Katrin Rentzsch⁷, Barnabas Szaszi², Stefan Schulz-Hardt^{8,9}, Barbara Sioma³, Peter Szecsi², Attila Szuts², Orsolya Szöke², Oliver Christ¹⁰, Anna Fedor¹¹, William Jiménez-Leal¹², Rafał Muda³, Gideon Nave⁵ , Janos Salamon^{2,13}, Thomas Schultze^{8,9} , and Joshua K. Hartshorne¹ 

¹Psychology and Neuroscience Department, Boston College; ²Institute of Psychology, ELTE Eötvös Loránd University; ³Faculty of Economics, Maria Curie-Skłodowska University; ⁴Department of Psychology, Medical School Hamburg; ⁵Marketing Department, The Wharton School of the University of Pennsylvania; ⁶Department of Psychology, Cornell University; ⁷Psychologische Hochschule Berlin; ⁸Institute of Psychology, University of Göttingen; ⁹Leibniz Science Campus Primate Cognition, Göttingen, Germany; ¹⁰Faculty of Psychology, FernUniversität in Hagen; ¹¹Institute of Evolution, Centre for Ecological Research, Budapest, Hungary; ¹²Department of Psychology, Universidad de los Andes; and ¹³Doctoral School of Psychology, ELTE Eötvös Loránd University

Abstract

As part of the Many Labs 5 project, we ran a replication of van Dijk, van Kleef, Steinel, and van Beest's (2008) study examining the effect of emotions in negotiations. They reported that when the consequences of rejection were low, subjects offered fewer chips to angry bargaining partners than to happy partners. We ran this replication under three protocols: the protocol used in the Reproducibility Project: Psychology, a revised protocol, and an online protocol. The effect averaged one ninth the size of the originally reported effect and was significant only for the revised protocol. However, the difference between the original and revised protocols was not significant.

Keywords

replication, Many Labs 5, ultimatum, open data, open materials, preregistered

Received 4/24/18; Revision accepted 3/27/20

Negotiation is a crucial aspect of many major decisions, in areas spanning policy, medicine, education, and business. Because of the integral role that negotiation plays in large-scale, as well as small-scale, decisions, a wealth of research has explored the aspects of negotiation that might influence outcomes. The previously studied predictors of negotiation outcomes include positive affect (Carnevale & Isen, 1986), gender (Mazei et al., 2015), time pressure (Stuhlmacher & Champagne, 2000), and physical appearance (Solnick & Schweitzer, 1999), to name a few.

Emotion has long been studied as an important predictor of negotiation outcomes (Olekalns & Druckman, 2014), and a large body of research has found that negotiating partners who express anger are more successful than those who express happiness (van Kleef, De Dreu, & Manstead, 2004). In "A Social Functional

Corresponding Author:

Joshua K. Hartshorne, McGuinn Hall 522, 140 Commonwealth Ave., Chestnut Hill, MA 02467
E-mail: joshua.hartshorne@bc.edu

Advances in Methods and
Practices in Psychological Science
1–11

© The Author(s) 2020

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/2515245920927643

www.psychologicalscience.org/AMPPS



Approach to Emotions in Bargaining: When Communicating Anger Pays and When It Backfires,” van Dijk, van Kleef, Steinel, and van Beest’s (2008) reported a study in which they examined this effect of expressed emotion in negotiation further by manipulating the structure of the negotiation itself. In their Experiment 3, when the consequences of rejection were low, subjects offered fewer chips to angry bargaining partners than to happy partners. (The interactions were actually simulations with a computer.)

This study by van Dijk et al. (2008) was included in the Reproducibility Project: Psychology (RP:P), a large-scale reproducibility study by the Center for Open Science (Open Science Collaboration, 2015). The goal of this project was to assess the reliability of the statistical findings reported in high-profile psychological publications. A single key statistical result was selected from each report as the target for replication. Note that it is rare for the theoretical claims advanced in psychological work to depend solely on a single statistical result. Thus, this project was not designed to provide a holistic assessment of the robustness of the key theoretical claims in each of the original reports. Rather, it was designed to provide a holistic assessment of the robustness of key statistical findings across a representative sample of high-profile studies.

In the case of van Dijk et al.’s (2008) article, the replication team selected the just-mentioned finding that when the consequences of rejection were low, subjects offered fewer chips to angry bargainers than to happy bargainers, $F(1, 99) = 16.62, p < .0001$ (Slowik & Voracek, 2015). This statistical finding was clearly identified by van Dijk et al. as a key result, and it was highlighted in the article’s abstract. However, the attempt to replicate it failed: The authors found no significant difference between the offers in the happy and angry conditions, and the reported effect was in the direction opposite the direction of the effects van Dijk et al. observed

There were several methodological differences between the RP:P replication and the original study. The sample in van Dijk et al.’s (2008) study was limited to students with no psychology or economics background, no prior participation in behavioral experiments, and no knowledge of the other subjects. The RP:P replication did not enforce these restrictions. Subjects completed the original study individually in cubicles, where they could not hear or see one another. In the RP:P replication, subjects were situated in the same room, with partial partitioning, so that they could hear, but not see, each other. The differences in the exclusion criteria and testing environments might explain the divergent result, as these factors are potentially theoretically relevant moderators of the observed effect. In

addition, one could argue that the exclusion criteria and testing environment in the RP:P replication might have threatened the credibility of the cover story and the simulated interaction. Both the theoretical and the methodological considerations speak to the necessity of an additional replication study.

In the current study, we aimed to investigate the possibility that these methodological differences account for the differences in results by running three different protocols across eight different sites: (a) a replication of the replication protocol used by the RP:P (Open Science Collaboration, 2015), (b) a revised, peer-reviewed protocol intended to most closely match the original study by van Dijk et al. (2008), and (c) an online version of the revised protocol.

Disclosures

Preregistration

Our methodology and analyses were preregistered on the Open Science Framework (<https://osf.io/b2sgw/>). Our reported analyses depart from the preregistration in two ways. First, reviewers of the results-blind protocol requested some additional analyses involving different sets of exclusion criteria.

Second, some of the analyses described in the results-blind manuscript were either underspecified or misspecified. Specifically, the approved results-blind text indicated that we would report regression coefficients and t statistics for fixed effects in the regression models. However, this is not possible for multilevel categorical factors, for which regression coefficients and t statistics are not well defined. Moreover, there are a number of methods for assessing the significance of mixed-effects models (Bolker, 2020), and we neglected to state which one we would use.

We resolved these issues as follows. When t statistics were well defined, we calculated them and assessed significance using the Wald test, which treats the t value as a z value. This is the standard choice in the lead team’s (L. Skorb and J. K. Hartshorne’s) lab, and likely what we had in mind. For clarity, we follow convention in labeling these t statistics *Wald’s zs*. For multilevel factors, for which regression coefficients and t statistics cannot be calculated, we assessed significance using model comparison with Akaike information coefficients (AICs; Akaike, 1974), with one exception noted later. Despite the fact that these analytic decisions were not preregistered, they were decided on prior to analysis, and no other methods were considered or used.

Finally, writing a results-blind manuscript¹ for a contingent analysis plan raises some difficulties: One must write text describing analyses without knowing what

analyses will actually be conducted. In some places, our results-blind manuscript did not fully capture the branching possibilities. All preregistered analyses included in our results-blind manuscript are reported here, and additional analyses are explicitly marked as “exploratory.”

Data, materials, and online resources

All materials, data, and code are available on the Open Science Framework (<https://osf.io/f6n27/>). Note that because of edits requested after acceptance, in some parts these materials diverge significantly from this article, though the overall import remains the same. The Supplemental Material (<http://journals.sagepub.com/doi/suppl/10.1177/2515245920927643>) includes analyses of the effects of exclusions (which were minimal) and protocol differences that were hypothesized to modulate the effect (but did not). It also includes several analyses that were not relevant to the current investigation but were reported in the original article by van Dijk et al. (2008) and thus may be of independent scientific interest.

Reporting

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

Ethical approval

Testing was conducted with the approval of the institutional review boards of the data-collection sites. In addition, the research was conducted in accordance with the Declaration of Helsinki, which requires pre-registration before data collection begins.

Method

Table 1 provides a summary of the methods for all three protocols, as well as the original study and RP:P replication.

Demographics of the sample

A total of 1,018 people initially participated across our three replication protocols. No subjects in the replication of the RP:P protocol were excluded from analyses, but on the basis of the preregistered exclusion criteria, we excluded from analyses a total of 157 (57%) subjects in the revised protocol and a total of 193 (64%) subjects in the online protocol. Across those two protocols, 15 people failed the manipulation checks, 337 expressed doubt that their partner was real, 13 knew at least 2

other subjects well, and 20 expressed knowledge about the ultimatum game. Note that we did not explicitly exclude subjects who had previous coursework in psychology or economics or had participated in behavioral experiments in the past because our recruitment strategy was designed to screen out these subjects in the beginning. However, we still considered the effect of all exclusion criteria in our analyses.

We excluded 5 subjects who enrolled but who could not be tested in accordance with the protocol: Two were excluded because no other subjects showed up for the testing time, and another 3 were excluded because the software crashed.

After application of these preregistered exclusion criteria, 663 subjects remained for inclusion in the analyses (mean age = 22.26 years, $SD = 5.93$): 436 (mean age = 20.54 years, $SD = 2.84$) under the original replication protocol, 119 (mean age = 20.88 years, $SD = 2.68$) under the revised protocol, and 108 (mean age = 30.70 years, $SD = 9.50$) under the online protocol.

Note that we present the results of our main analyses without any exclusions whatsoever (including for failed manipulation checks) in the Supplemental Material. These results were consistent with the primary confirmatory analyses reported here.

Materials and procedure

The materials were identical at all testing sites and in all protocols (see Table 1 for which protocols were run at each site). In order to ensure this consistency, we created a software package that could run on all modern Web browsers, so that all the sites would be able to use it. The stimuli were hosted and the data were stored on the gameswithwords.org Web servers.

The original materials from van Dijk et al. (2008) were translated into English, Spanish, German, Hungarian, and Polish (so that the local language could be used at each testing site) and were programmed into an online format through gameswithwords.org. Each site and protocol used the same materials and the same online format; only the language varied.

At the start of the instructions, the subjects were informed that they would participate in a study on bargaining and that they would be paired with one of the other subjects, who would be referred to as “Y” (the subjects were always assigned to be “X”). They then saw a message that they were being connected to another subject. After this, subjects were informed that before the bargaining would start, they would answer some questions about how they viewed bargaining. Six general statements (e.g., “Strategy is important in negotiations,” “In negotiations I quickly try to reach agreement”) were presented, and subjects had to indicate on

Table 1. Comparison of the Methodology in the Five Protocols

Feature	Current study			
	Original protocol (van Dijk, van Kleef, Steinel, & van Beest, 2008)	Original RP:P protocol	Replication of the RP:P protocol	Revised protocol
Recruitment of subjects	The method of recruiting subjects was unreported.	Subjects were from a university subject pool but were not advanced students of psychology or economics.	Subjects were from a university subject pool, and specifically were freshmen, who by definition are less likely than other students to have taken psychology and economics.	At FernUniversität in Hagen, subjects were recruited through a general online forum for psychology students; the forum of the introductory class; and the “virtual lab,” an official university website for students on which all university studies are posted. On CrowdFlower, subjects were recruited by posting the study as a job on CrowdFlower.
Exclusion criteria	The authors tried to avoid participation of subjects who had prior experience with psychology or economics, who had participated in previous experiments, or who knew other subjects. It is not clear how these criteria were operationalized.	The authors reported that they tried to recruit presumably hypothesis-naïve participants and aimed to exclude advanced psychology or economics students. It is unclear how these exclusion criteria were operationalized.	No additional exclusion criteria were used.	The exclusion criteria included knowledge of the ultimatum game, expression of doubt that the partner was real, familiarity with other subjects, and answering a manipulation check incorrectly.
Sample	103 participants (22 males, 81 females; mean age = 20.8 years)	83 participants (27 males, 56 females; mean age = 24.5 years)	436 participants (177 males, 258 females, 1 other; mean age = 21 years)	301 participants (165 males, 135 females, 1 other; mean age = 32 years); after exclusion: 108 participants (61 males, 47 females; mean age = 31 years)
Size of testing group	Groups of 8	Groups of 4	Groups of 4, 6, or 8 (confederates were available to substitute for missing subjects if necessary)	Individual

(continued)

Table 1. (*continued*)

Feature	Current study		
	Original RP:P protocol	Replication of the RP:P protocol	Revised protocol
Original protocol (van Dijk, van Kleef, Steinel, & van Beest, 2008)			
Testing environment	A room with 4 cubicles that were open in the back	A room with 4, 6, or 8 cubicles or small rooms that were open in the back	A room with 4, 6, or 8 enclosed cubicles or smaller rooms
Presentation of offer	Subjects wrote their offer by hand on a sheet of paper, allegedly to be given to their partner, before also typing it on the keyboard.	The offer was made only on a computer.	The offer was made only on a computer.
Design	2 (recipient's emotion: angry vs. happy) × 2 (consequences of rejection: low vs. high) between-subjects factorial ANOVA design Leiden University	2 (recipient's emotion: angry vs. happy) × 1 (consequences of rejection: low) between-subjects factorial ANOVA design Boston College, ELTE Eötvös Loránd University, Maria Curie-Skłodowska University, Universidad de los Andes, University of Göttingen School	2 (recipient's emotion: angry vs. happy) × 1 (consequences of rejection: low) between-subjects factorial ANOVA design Boston College, Maria Curie-Skłodowska University, Universidad de los Andes, University of Göttingen
Software	Unreported	OpenSesame 2.9 (Mathôt Schreijf, & Theeuwes, 2012)	Web-based software shared across the sites
Site			Web-based software shared across the sites

Note: Because of an institution change early in the project, the Hungarian Academy of Sciences is referred to as the Milestone Institute in the online archives. RP:P = Reproducibility Project: Psychology; ANOVA = analysis of variance.

a 5-point scale the extent to which they agreed or disagreed with each. Only after answering these questions did subjects learn that their answers would be sent to their bargaining partner. The reason given for this was that in reality, people often have some information about the other party in a negotiation.

After the bargaining situation was explained, the manipulation of the bargaining partner's emotion was implemented. Subjects learned that their partners had typed a reaction after reading their six answers to the bargaining statements. It was emphasized that when the partners typed their reactions, they did not think that the reactions would be sent back to the subjects (cf. van Kleef et al., 2004). This was done to ensure that the subjects trusted that the typed emotion was genuine (rather than a strategic message). In the angry condition, subjects read, "I've read what X typed and it makes me angry!!" In the happy condition, subjects read, "I've read what X typed and it makes me happy!!" Note that because of concerns with the translations of the Dutch to American English (e.g., the original translation of the first statement was "I've read what X typed and I must say I am angry about it!"), we edited the translations to make them more natural.

After this manipulation, all subjects learned that they would bargain with their partner over the distribution of 100 chips; 1 chip was said to be worth \$0.10 (or the equivalent in local currency). In both conditions, subjects learned that their partner would be the recipient of an offer they would make regarding how they wanted the 100 chips to be allocated. If the recipient agreed to the division, the chips would be distributed accordingly, and if the recipient turned down the division, the division would be reduced by 10% (thus, low consequences of rejection were established). (Note that all subjects received instructions indicating low consequences of rejection; we did not include high consequences of rejection in this replication.) Subjects then made their offer to the recipient by typing on the keyboard.

Prior to making the offer, subjects responded to some manipulation checks. They were asked to rate the extent to which the recipient was angry and to which the recipient was happy (Likert scales from 1 to 7). They then rated how much they and their partner could affect the eventual distribution of chips (Likert scale from 1, *X could have more impact*, to 7, *Y could have more impact*). Subjects were also asked what would happen if the recipient rejected the offer (the offer would be reduced by 10%).

Next, subjects were asked to rate their own anger and happiness (7-point Likert scales), how likely they thought it was that the recipient would accept the offer (Likert scale from 1, *very unlikely*, to 7, *very likely*), and the extent to which the recipient having the possibility to reject their offer affected their decision (Likert scale

from 1, *absolutely not*, to 7, *absolutely*). Subjects were also asked, "How many chips do you think Y should be offered to have him/her accept the offer?" Finally, they answered some questions about the identity of their bargaining partner (male, female, or computer) and their previous psychology and economics courses, participation in prior behavioral experiments, familiarity with other subjects, and demographics. Data for questions not discussed in this article are discussed in the Supplemental Material or accessible from the open data set.

Differences among the protocols

As already noted, the protocols differed in the exclusion criteria applied. They also differed in the testing environment.

In the replication of the RP:P protocol, 4 or more people participated at the same time. In the original RP:P study, the testing environment may have allowed subjects to hear each other, but they were not able to see one another. Thus, adjacent rooms with the door open or cubicles that were in the same room and not fully enclosed were appropriate for our replication. Participants were led to a computer and completed the study individually in a Web browser.

In the revised protocol, 4, 6, or 8 people participated at the same time. We waited for at least 2 subjects to arrive and immediately led them to separate cubicles or rooms, instructing them to wait until all the subjects had arrived, and then shutting the door. It should not have been possible to see or hear other subjects from inside. The doors to the cubicles or rooms were visible from the arrival area, providing additional evidence to the subjects that there were indeed other subjects. Note that because subjects were seated asynchronously, it was not necessary to have confederates replace no-shows. Subjects were informed that there were 3, 5, or 7 other subjects at the same time.

In the online protocol at FernUniversität in Hagen and on CrowdFlower, subjects completed the study online. They received a link to the study's website, where they followed the instructions on their own. At FernUniversität in Hagen, subjects could start the study only at certain times during the day, so that they could not access the website at a point in the night when they would not believe that somebody was up and participating at the same time. Except for this time restriction, the environment was relatively uncontrolled. On CrowdFlower, subjects could start the study at any point after receiving a link to it, and therefore the study environment was uncontrolled.

The original authors stressed that subjects should not all be seated at the same time, and also that they should not be seated strictly one at a time. No further clarification was available. In consultation with the journal's

Table 2. Effect of Emotion Condition on the Initial Offer in the Original Study, the RP:P Study, and the Current Project After Exclusions

Protocol	Happy condition		Angry condition		Difference	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	Cohen's <i>d</i>
Original (van Dijk, van Kleef, Steinel, & van Beest, 2008)	45.21	9.03	32.56	16.31	-12.65	-0.96
Original RP:P	36.20	17.45	37.60	16.62	1.40	0.08
Current project						
RP:P replication	41.93	20.29	38.73	20.68	-3.20	-0.16
Revised	46.40	11.67	38.98	21.23	-7.42	-0.48
Online	44.60	21.07	42.07	28.43	-2.53	-0.10
Protocols combined	43.32	18.90	39.35	22.26	-3.97	-0.19

Note: RP:P = Reproducibility Project: Psychology.

Editor, we decided that for the RP:P and revised protocols, the experimenters would begin seating subjects only after at least 2 had arrived. This consideration did not apply to the online protocols.

Results

Preliminary analyses

First, we used a binomial regression to analyze the effect of previous psychology courses, previous economics courses, participation in behavioral experiments, protocol, and their interactions on being excluded. We implemented these analyses in R using the *lme4* package, Version 1.1-12 (Bates, Mächler, Bolker, & Walker, 2015). These analyses included all 1,013 subjects. We expected to see a main effect of protocol. Specifically, we expected that subjects in the revised and online protocols would be significantly more likely to be excluded than subjects in the replication of the RP:P protocol, given that the latter protocol did not have any exclusion criteria. We observed a pattern that was consistent with that prediction (AICs = 817 and 1,092 for models with and without the effect of protocol, respectively). Model comparison provided very slight evidence for interactions of protocol with the number of experiments subjects had participated in previously (AICs = 816.5 vs. 816.9) and with the number of psychology courses they had taken (also AICs = 816.5 vs. 816.9). Generally, such a small difference in AIC is not considered meaningful. No other effects or interactions were significant.

Next, we conducted the same between-subjects manipulation checks as in the original study by van Dijk et al. (2008). These analyses included only subjects who passed the exclusion criteria. An independent-samples *t* test compared the perceived anger in the angry and happy conditions. We expected that people in the angry condition would perceive significantly greater anger in the bargaining partner, and this what we found (angry condition: $M = 5.26$, $SD = 1.93$; happy

condition: $M = 1.50$, $SD = 1.27$), $t(522.66) = 29.23$, $p < 1 \times 10^{-5}$). This pattern of results confirms that the manipulation of anger was successful. An independent-samples *t* test was also conducted to compare the perceived happiness in the angry and happy conditions. We expected that people in the angry condition would perceive significantly lower less happiness in the bargaining partner, and this is what we found (angry condition: $M = 2.06$, $SD = 1.67$; happy condition: $M = 6.47$, $SD = 1.11$), $t(528.56) = -39.48$, $p < 1 \times 10^{-5}$. This pattern of results confirms that the manipulation of happiness was successful.

Main analyses

The main effect of interest was the effect of expressed emotion (angry vs. happy) on the number of coins that participants initially offered to their bargaining partner. In the original study, van Dijk et al. (2008) found that emotion was a significant predictor of subjects' initial offer to the other player when there were low consequences of rejection. Subjects offered less to the angry recipient ($M = 32.56$, $SD = 16.31$) than to the happy recipient ($M = 45.21$, $SD = 9.03$). Thus, perceived emotion did affect subjects' offers to their negotiation partners.

Table 2 shows the means and standard deviations for the number of coins offered in van Dijk et al.'s (2008) original study, the original RP:P replication, and our three protocols (RP:P replication, revised, online), as well as effect sizes for the difference between conditions.

We used mixed-effects regression for our primary analyses, rather than analysis of variance (used in the original study), because mixed-effects regression is robust to unbalanced designs and allows for more appropriate random-effects structures (Baayen, 2008). In order to determine the best regression model to test the effect of interest in our data, we first ran a stepwise linear regression, starting with emotion condition, protocol, and testing site (university) as predictors. On the

Table 3. Interaction Effect of Emotion Condition and Protocol on the Initial Offer in the Current Project After Exclusions

Predictor	Estimate	SE	Wald's z	p
Intercept	38.73	1.40	27.57	$< 1 \times 10^{-5}$
Happy condition	3.20	1.97	1.62	.11
Revised protocol	0.24	3.50	0.07	.94
Online protocol	3.34	3.11	1.07	.28
Happy Condition \times Revised Protocol	4.23	4.42	0.95	.34
Happy Condition \times Online Protocol	-0.66	4.42	-0.15	.88

Note: In this analysis, the central question could not be addressed using Akaike's information criterion, as planned, so we dummy-coded the three-level protocol factor with the replication of the Reproducibility Project: Psychology protocol as the baseline.

basis of the results of this stepwise regression, we used a mixed-effects regression model that considered the main effects of emotion condition and protocol and estimated the random intercept of testing site. A successful replication would be indicated by a main effect of emotion condition in which expressed happiness predicted a higher initial offer than expressed anger. We replicated the significant effect of emotion: Expressed happiness predicted a higher initial offer than expressed anger ($b = 3.78$, $SE = 1.61$, Wald's $z = 2.35$, $p = .019$).

We also determined that the main effect of protocol was significant (AICs = 6,078 and 6,084 for models with and without the main effect of protocol, respectively). Initial offers were highest for the revised protocol (44.55, $SD = 15.00$), slightly lower for the online protocol (43.31, $SD = 25.01$), and lowest for the replication of the RP:P protocol (40.00, $SD = 20.70$).

The stepwise regression did not license testing an interaction between protocol and emotion condition. However, this interaction was of core interest in our project. Thus, we ran an additional (exploratory) model with this interaction term added. There was a significant

interaction effect (AICs = 6,072 and 6,078 for models with and without the interaction, respectively). The online protocol showed the smallest effect of emotion condition, and the revised protocol showed the largest (Table 2). The critical effect of emotion condition, however, was not significantly different between our RP:P replication protocol and either the revised protocol or the online protocol (see Table 3).

We additionally conducted mixed-effects regressions with emotion condition as the only predictor and with the random intercept of testing site, separately for each of the protocols (see Table 4). The only protocol with a significant effect of emotion was the revised protocol.

Follow-up analyses

An anonymous results-blind reviewer requested that we analyze the results without excluding subjects who indicated that they thought their partner was a computer. This question on the identity of the bargaining partner—which was not included in the original study—may have retroactively caused subjects to doubt their

Table 4. Main Effect of Emotion Condition and Random Intercept of Testing Site in the Current Project After Exclusions

Predictor and statistic	Replication of the RP:P protocol	Revised protocol	Online protocol	Protocols combined
Intercept				
Estimate	38.68	38.98	42.07	39.31
SE	1.48	2.44	3.38	1.21
Wald's z	26.11	15.98	12.44	32.52
p	$< 1 \times 10^{-5}$	$< 1 \times 10^{-5}$	$< 1 \times 10^{-5}$	$< 1 \times 10^{-5}$
Happy condition				
Estimate	3.18	7.42	2.53	3.97
SE	1.96	3.01	4.83	1.6
Wald's z	1.62	2.46	0.52	2.48
p	.11	.014	.60	.013

Note: RP:P = Reproducibility Project: Psychology.

Table 5. Follow-Up Analysis: Effect of Emotion Condition on the Initial Offer in the Original Study, the RP:P Study, and the Current Project (Including Subjects Who Believed Their Partner Was a Computer)

Protocol	Happy condition		Angry condition		Difference	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	Cohen's <i>d</i>
Original (van Dijk, van Kleef, Steinel, & van Beest, 2008)	45.21	9.03	32.56	16.31	-12.65	-0.96
Original RP:P	36.20	17.45	37.60	16.62	1.40	0.08
Current project						
RP:P replication	41.93	20.29	38.73	20.68	-3.20	-0.16
Revised	43.10	15.49	38.39	20.42	-4.71	-0.26
Online	42.39	20.03	42.94	23.79	0.55	0.02
Protocols combined	42.38	19.00	39.87	21.61	-2.51	-0.12

Note: RP:P = Reproducibility Project: Psychology.

partner's validity. Alternatively, subjects who were stingy in bargaining may have been motivated to claim that they thought their partner was a computer.

In response to this reviewer's request, we again ran mixed-effects models with the main effect of emotion and random intercept of testing site, this time including subjects who said they thought their partner was a computer (Tables 5 and 6). (Note that although this analysis was not included in the preregistration, it was included in the approved results-blind manuscript.) The results of these models suggest that excluding these subjects had little effect on the overall results, despite the original authors' suspicion that the RP:P replication failed because of such subjects (author correspondence is available at <https://osf.io/a6pz3>). In the Supplemental Material, we show that removing all exclusions similarly had little effect on the pattern of results.

Discussion

In this article, we have assessed the replicability of a single key statistical result from van Dijk et al. (2008):

that when the consequences of rejection were low, expressing anger, rather than happiness, resulted in a lower offer from a negotiation partner. The current study follows on a prior failed replication (Open Science Collaboration, 2015). We conducted three replications: one using (roughly) the same protocol as that used in the RP:P, a protocol that was revised to better reflect van Dijk et al.'s original protocol and that went through peer review prior to its use, and a variant of the revised protocol that was modified to be suitable for running over the Internet. In all three protocols, the effect was in the original direction, though it was significant only for the revised protocol (see Table 4).

What does this mean for the original effect of interest? Expressing positive emotion resulted in a more generous negotiation offer in four out of five experiments conducted by three independent research teams—and this effect was significant in two cases. These results are consistent with there being at least a small effect in the predicted direction in at least some contexts. We found little evidence for a large effect that is robust to context. Indeed, our data suggest that the

Table 6. Follow-Up Analysis: Main Effect of Emotion Condition and Random Intercept of Testing Site in the Current Project (Including Subjects Who Believed Their Partner Was a Computer)

Predictor and statistic	Replication of the RP:P protocol	Revised protocol	Online protocol	Protocols combined
Intercept				
Estimate	38.68	38.02	42.71	39.45
<i>SE</i>	1.48	1.99	2.35	1.15
Wald's <i>z</i>	26.11	19.07	18.2	34.44
<i>p</i>	$< 1 \times 10^{-5}$	$< 1 \times 10^{-5}$	$< 1 \times 10^{-5}$	$< 1 \times 10^{-5}$
Happy condition				
Estimate	3.18	4.72	-0.44	2.54
<i>SE</i>	1.96	2.21	2.63	1.3
Wald's <i>z</i>	1.62	2.13	-0.17	1.96
<i>p</i>	.11	.033	.87	.05

Note: RP:P = Reproducibility Project: Psychology.

effect of expressing anger during negotiation is far more modest than originally reported. When the data were collapsed across our three experiments, initial offers were only 10% higher for partners who had expressed happiness than for those who had expressed anger (see Table 2). This increase is only a quarter the size of what van Dijk et al. (2008) reported (39%), despite the fact that we used a number of quality-control measures not employed by the original authors, which resulted in large numbers of exclusions. Without those exclusions, our measured effect was only 5%—or about one ninth of what van Dijk and colleagues observed (see Table S1 in the Supplemental Material). The effect was slightly stronger for the revised protocol, in which the increase was 19% with exclusions and 9% without. We emphasize effect size because this effect is one for which effect size has direct practical implications. We leave it to other researchers—and negotiators—to determine whether a 5% difference in an initial offer is important.

The overarching question addressed by the Many Labs 5 project is whether peer-reviewed replication protocols are more likely to result in successful replication than are protocols that were used in a previous replication project and mostly failed to replicate prior findings. Our study provides at best limited evidence for the claim that they are. Although only the revised protocol provided statistically significant support for the original effect, all three protocols yielded an effect in the same direction as the original. Critically, the effect was not statistically larger for the revised protocol than for the protocol that followed the RP:P replication.

Interestingly, most of the differences introduced into the revised protocol involved procedures intended to ensure subjects' naivete and willingness to believe that they were interacting with a live partner. However, these procedures seem to have had little, if any, effect. For example, as we have shown, the pattern of the results was the same whether subjects who saw through the deception were included or not. We present a more detailed analysis of specific suspected moderators of the effect (e.g., prior exposure to psychology research) in the Supplemental Material. Although certainly we cannot say that peer review would never identify flaws in a replication protocol, in this case peer review focused on features that ultimately made little difference.

However, this lack of an effect of peer review should not be overinterpreted. We tested the efficacy of one set of peer reviewers at one journal in improving the replication protocol designed by one set of researchers for one experiment. That is, the apparent fact that our peer reviewers did not identify a significant flaw in our

protocol does not mean that other peer reviewers would not have detected such a flaw, or that flaws would not be detected in other protocols by ourselves or other researchers. Moreover, our total number of subjects may be large relative to the standards of the field, but it is modest relative to the demands of statistics. The fact that we did not find a large effect of peer review does not mean that a larger study would not find a small effect.

An important limitation on our conclusions about the effect of expressing anger during negotiations is that across the five studies, we tested only one method of expressing anger and one method of expressing happiness. Other methods—particularly more dramatic expressions of affect—might have larger or different effects.

Note that because our study was designed to test replicability, it necessarily leaves other questions open as well. In particular, we are unable to fully evaluate some of the design decisions made by the original authors. For instance, in keeping with the instructions of the original authors, although we informed all subjects about the number of simultaneously tested subjects, they personally observed variable numbers of other subjects (depending on exactly when they and the other subjects arrived). Unfortunately, it is not known what the distribution of subject-subject exposure was in the original study, and thus it is likely that the distribution in ours was different. This could matter in a number of ways. First, it could have affected how willing subjects were to believe that they were negotiating with a live individual. Second, it meant that some of the subjects knew who some of their potential partners were, which could have affected how they responded to the negotiation (note that by design, subjects always had some uncertainty as to who their partner was). The fact that the results were (nonsignificantly) weakest for the online protocol could support the claim that seeing other subjects is important for realism. Conversely, we obtained mixed results as to whether familiarity with the other subjects interacted with the effect of interest (for the revised protocol, greater familiarity was associated with a smaller effect, but for the replication of the RP:P protocol, the interaction was not significant; see the Supplemental Material). Thus, although the impact of physical copresence of the subjects does not appear to have been sufficient to qualitatively affect our results, it is possible that smaller quantitative effects were present.

Finally, we emphasize that we did not test all the statistical findings in van Dijk et al. (2008). Therefore, we cannot provide a holistic assessment of the broader theoretical claims made in the original article.

Transparency

Action Editor: Daniel J. Simons

Editor: Daniel J. Simons

Author Contributions

L. Skorb and J. K. Hartshorne conducted the analyses and drafted the manuscript. All the authors contributed to collecting data and revising the manuscript. All the authors approved the manuscript submitted for publication.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

This project was supported by a grant from the Association for Psychological Science and from Arnold Ventures.

Open Practices

Open Data: <https://osf.io/f6n27/>

Open Materials: <https://osf.io/f6n27/>

Preregistration: <https://osf.io/b2sgw/>

All data and materials have been made publicly available via the Open Science Framework and can be accessed at <https://osf.io/f6n27/>. The design and analysis plans were preregistered at the Open Science Framework and can be accessed at <https://osf.io/b2sgw/>. The complete Open Practices Disclosure for this article can be found at <http://journals.sagepub.com/doi/suppl/10.1177/2515245920927643>. This article has received badges for Open Data, Open Materials, and Preregistration. More information about the Open Practices badges can be found at <http://www.psychologicalscience.org/publications/badges>.



ORCID iDs

Gideon Nave  <https://orcid.org/0000-0001-6251-5630>

Thomas Schultze  <https://orcid.org/0000-0002-7900-0809>

Joshua K. Hartshorne  <https://orcid.org/0000-0003-1240-3598>

Supplemental Material

Additional supporting information can be found at <http://journals.sagepub.com/doi/suppl/10.1177/2515245920927643>

Note

1. Because the editorial decision to use the results-blind review process was made after data collection was completed, L. Skorb and J. K. Hartshorne were not entirely blind to the data when they drafted the manuscript.

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, England: Cambridge University Press.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). doi:10.18637/jss.v067.i01
- Bolker, B. (with others). (2020). *GLMM FAQ*. Retrieved from <http://bbolker.github.io/mixedmodels-misc/glmmFAQ.html>
- Carnevale, P. J., & Isen, A. M. (1986). The influence of positive affect and visual access on the discovery of integrative solutions in bilateral negotiation. *Organizational Behavior and Human Decision Processes*, *37*, 1–13.
- Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*, 314–324.
- Mazei, J., Hüffmeier, J., Freund, P. A., Stuhlmacher, A. F., Bilke, L., & Hertel, G. (2015). A meta-analysis on gender differences in negotiation outcomes and their moderators. *Psychological Bulletin*, *141*, 85–104.
- Olekalns, M., & Druckman, D. (2014). With feeling: How emotions shape negotiation. *Negotiation Journal*, *30*, 455–478.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*, Article aac4716. doi:10.1126/science.aac4716
- Slowik, A., & Voracek, M. (2015). *Replication of “A social functional approach to emotions in bargaining: When communicating anger pays and when it backfires” by E van Dijk, G A van Kleef, W Steinel, I van Beest (2008, Journal of Personality and Social Psychology)*. Retrieved from <https://osf.io/2idfu/>
- Solnick, S. J., & Schweitzer, M. E. (1999). The influence of physical attractiveness and gender on ultimatum game decisions. *Organizational Behavior and Human Decision Processes*, *79*, 199–215.
- Stuhlmacher, A. F., & Champagne, M. V. (2000). The impact of time pressure and information on negotiation process and decisions. *Group Decision and Negotiation*, *9*, 471–491.
- van Dijk, E., van Kleef, G. A., Steinel, W., & van Beest, I. (2008). A social functional approach to emotions in bargaining: When communicating anger pays and when it backfires. *Journal of Personality and Social Psychology*, *94*, 600–614.
- van Kleef, G. A., De Dreu, C. K. W., & Manstead, A. S. R. (2004). The interpersonal effects of anger and happiness in negotiations. *Journal of Personality and Social Psychology*, *86*, 57–76.