Rewarding more is better for soliciting help, yet more so for cash than for goods: Revisiting and reframing the Tale of Two Markets with replications and extensions of Heyman and Ariely (2004)

*Hirotaka Imada
School of Psychology, University of Kent
hi67@kent.ac.uk
ORCID: 0000-0003-3604-4155

*Wan Fei Chan
Department of Psychology, University of Hong Kong, Hong Kong SAR
fei666@connect.hku.hk / zeong1998@gmail.com

*Yuk Ki Ng
Department of Psychology, University of Hong Kong, Hong Kong SAR
ngyukki@connect.hku.hk / ngyukki0711@gmail.com

*Lee Hing Man
Department of Psychology, University of Hong Kong, Hong Kong SAR
u35538333@connect.hku.hk / claricemanleehing@gmail.com

*Mei Sze Wong
Department of Psychology, University of Hong Kong, Hong Kong SAR
u3554152@connect.hku.hk / meisze80@gmail.com

Bo Ley Cheng
Department of Psychology, University of Hong Kong, Hong Kong SAR
boleyc@hku.hk / boleystudies@gmail.com

^Gilad Feldman
Department of Psychology, University of Hong Kong, Hong Kong SAR
gfeldman@hku.hk / giladfel@gmail.com
ORCID: 0000-0003-2812-6599

In press at Collabra:Psychology
Accepted for publication on February 9, 2022

*Contributed equally, joint first author
^Corresponding author
Word count: 4579 words

Author bios:

Hirotaka Imada is a PhD student at University of Kent. His research interest is formed around prosocial behavior and reputation.

Wan Fei Chan, Lee Hing Man, Yuk Ki Ng, and Mei Sze Wong were students with the psychology department at the University of Hong Kong during the academic year 2019.

Boley Cheng was a teaching assistant at the University of Hong Kong psychology department during the academic year 2019.

Gilad Feldman is an assistant professor with the University of Hong Kong psychology department. His research focuses on judgment and decision-making.

Declaration of Conflict of Interest:

The authors declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

Financial disclosure/funding:

The authors received no financial support for the research and/or authorship of this article.

Authorship declaration:

Wan Fei Chan, Lee Hing Man , Yuk Ki Ng, and Mei Sze Wong conducted the replication, including original article analysis, power analyses, designed the experiments, constructed the surveys, wrote the pre-registrations, and conducted the data analysis and initial write-ups.

Boley Cheng guided and assisted the replication effort .

Hirotaka Imada verified and extended analyses, integrated the studies, and wrote the final manuscript for submission.

Gilad Feldman led the replication effort, supervised each step in the project, conducted the pre-registrations, and ran data collection. Gilad edited the final draft for submission.

Corresponding author

Gilad Feldman, Department of Psychology, University of Hong Kong, Hong Kong SAR; gfeldman@hku.hk

Acknowledgements

We would especially like to thank Max Korbmacher for serving as our red team for this project, pushing us to do better and to Ashleigh Haydock-Symonds for careful copyediting and suggestions for improvement. We would also like to thank Farid Anvari and Jerome Olsen for their feedback and involvement in the initial stage of this project. Finally, we would like to thank Raluca Diana Szekely, Raj Aiyer, Qinyu Xiao, and Nicole Russell for their feedback on earlier drafts of this manuscript.

Contributor Roles Taxonomy

In the table below, employ CRediT (Contributor Roles Taxonomy) to identify the contribution and roles played by the contributors in the current replication effort. Please refer to the url (https://www.casrai.org/credit.html) for details and definitions of each of the roles listed below.

Role	**Hirotaka Imada**	**Gilad Feldman**	**Wan Fei Chan, Yuk Ki Ng, Lee Hing Man, Mei Sze Wong**	**Boley Cheng**
Conceptualization		X		
Pre-registration		X	X	
Data curation		X	X	
Formal analysis	X		X	
Funding acquisition		X		
Investigation		X	X	X
Pre-registration peer review / verification	X	x		X
Data analysis peer review / verification			X	X
Methodology		X	X	
Project administration		X		
Resources		X		
Software				
Supervision		X		X
Validation	X		X	
Visualization	X		X	
Writing-original draft	X	X	X	
Writing-review and editing	X	X		

Abstract

Heyman and Ariely (2004) demonstrated that the expected effectiveness of soliciting help varied depending on the "market", a money market represented by cash rewards versus a social market represented by goods as rewards. They showed that, as cash rewards increase, individuals expected others to be more willing to help, yet, when offering social goods as rewards such as candy, expected willingness to help was insensitive to rewards' monetary worth. We conducted two pre-registered replication studies (total: *N* = 3302, MTurk/Prolific) of Study 1 in Heyman and Ariely (2004) and found support for one of their main claims that people are more sensitive to worth when the reward is cash than goods. However, the rewards' monetary worth impacted expected willingness to help even in social markets, deviating from the original findings. Extensions further compared between-subject and within-subject designs, examined perceived affect (joy and regret), and added a new control condition. We concluded that higher compensation is generally perceived as better when soliciting help, yet more so for the money market cash rewards than for the social market goods rewards. All materials, data, and code are provided on https://osf.io/y9p7u/

Keywords: social utility; helping; judgment and decision making; money market; social market; compensation; replication

Rewarding more is better for soliciting help, yet more so for cash than for goods: Revisiting and reframing the Tale of Two Markets with replications and extensions of Heyman and Ariely (2004)

Individuals often face situations where they need help from others. Nevertheless, helping behavior requires people to incur costs (e.g., time and effort), and it is not always easy to solicit desired help. Individuals, thus, utilize rewards to incentivize others to lend them a hand. Since such incentives can vary in types (e.g., money or goods) and size, it is of vital importance to understand how we can best elicit quality effort; what and how much should we provide to maximize the level of effort in solicited help?

Addressing this question, Heyman and Ariely (2004) have argued that the effectiveness of the level of incentives to motivate others depends on the perceived exchange relationship: money market or social market relationships. Drawing upon Fiske’s (1992) relational theory, they defined the money market relationship as an exchange where effort level is determined in accordance with reciprocity and, thus, the level of compensation directly shapes behavior. Accordingly, they hypothesized that the larger the amount of reward was, the more willing individuals were to help others, and the more effort was exerted in the money market relationship. By contrast, the social market relationship refers to an exchange where effort level is most influenced by altruistic motivations and remains high irrespective of the amount of reward. This led them to hypothesize that in the social market relationship, the amount of rewards would not affect the willingness to offer help and the effort invested in helping. In sum, they predicted that the influence of the reward level would be conditional on the exchange relationship.

Heyman and Ariely (2004) used different types of rewards, cash and candies, as a means to induce the money- and social market relationships, respectively. They conducted a set of studies, and these yielded empirical support for their hypotheses, revealing that increases in the amount of cash reward, but not non-cash reward, led to increased willingness to help others and enhanced effort invested in helping. In addition, they found that when cues of both relationships are presented (i.e., when candies with a price tag were offered), this primed the money market relationship.

Their findings have become a theoretical cornerstone of a wide range of subsequent research in various disciplines, and there are 1216 Google Scholar citations for Heyman and Ariely (2004) as of February 2022 (Bowles & Polanía-Reyes, 2012; Gneezy et al., 2011; Lacetera & Macis, 2010; Newman & Shen, 2011; Shampanier et al., 2007; Yam et al., 2012). The asymmetry in the effectiveness of big compensations compared to small ones between cash and non-monetary goods has guided people working in diverse fields, such as marketing (Shampanier et al., 2007) and conservation (Cifor, 2005; Wunder, 2007).

The present research sought to replicate and extend findings in Heyman and Ariely (2004) for three reasons: its substantial impact, the lack of direct replications, existing contradictory findings, and erroneous reporting of results. First, despite the substantial impact on a broad audience, to our knowledge, there have not been any direct replications of Heyman and Ariely (2004).

Second, a previous study with a similar experimental paradigm found contradictory results. Following Heyman and Ariely (2004), Liu et al. (2012) investigated the relative effectiveness of three payment forms (cash, soap, and soap with a price tag) with two different payment levels (low vs. medium) in encouraging individuals to participate in and take time to respond to a short survey. They failed to find support for effort level change depending on the level of the cash payment. Moreover, the effort level rather decreased when the non-monetary payment level increased. Regarding the willingness to help, they replicated the original finding that in the monetary payment condition, there was a positive relationship between the willingness to help and the payment level. However, while Heyman and Ariely (2004) argued that the payment level would have the same effect when the payment form was monetized goods (i.e., soap with a price tag), Liu et al. (2012) failed to find support for payment level affecting monetized goods. Though their research design was different from the original studies, their findings raised doubts regarding the robustness and generalizability of Heyman and Ariely (2004). These mixed findings raise the need for well-powered pre-registered direct close replications of this work.

Last, we found several erroneous and ambiguous reports of statistical analyses in the original article; there were inconsistencies between the reported sample size and degrees of freedom for F statistics. We also examined reported *p* values and found that nine out of 10 values appeared inconsistent with reported F statistics (see S1 in Supplementary for details). These have led to a recent expression of concern by one of the original authors and the journal, which recognized the issues (Bauer & Ariely, 2021). Follow-up research, therefore, cannot rely on the original article's reporting for estimating effect sizes. These issues raise the need for a careful reproduction of the materials and analyses to reassess these effects, amendment of the historical record, and replication work to verify the findings and obtain accurate estimates to allow future research.

Method Overview (Study 1 and 2)

We conducted two parallel well-powered pre-registered replications of Study 1 in Heyman and Ariely (2004), with extensions. According to LeBel et al.'s (2018) criteria, our studies were classified as very close replications. We pre-registered hypotheses and analytic plans of Study 1 (https://osf.io/h9wus/) and Study 2 (https://osf.io/5j7fg/), and we provided data, study materials, and results with analysis code at https://osf.io/y9p7u/ .

Participants

Due to erroneous statistics reported in the original article (Bauer & Ariely, 2021), we could not rely on the original effects for our power analyses. We, therefore, sought to recruit the maximum number of participants that was possible with the budgetary constraints of our project, which would far exceed even the most conservative effect estimates. For Study 1, we recruited a total of 2203 American participants from the United States via Amazon Mechanical Turk. According to Simonsohn's (2015) suggestion for simpler designs, our sample size was well beyond 2.5 times larger than the original sample size. In Study 2, we employed an adjusted within-subject design that is better powered and recruited a total of 999 British participants from Prolific Academic. We summarize the key demographic information of the samples in Table 1.

Table 1

Demographic information and study features in the original and replication studies.

	Heyman and Ariely (2004)	Study 1	Study 2
Participants	University Students	Amazon MTurk	Prolific Academic
Design	Between-subject	Between-subject	Within-subject
Sample size	614	2203	999
Geographic origin	United States American	United States American	British
Gender	NA	1058 males 1132 females	388 males 608 females
Median age	NA	37	38
Mean age	NA	39.70	39.70
Medium	Survey	Online Survey	Online Survey
Compensation	NA	Nominal payment	Nominal payment
Year	NA	2019	2019

Note. NA = not available/unknown. See “Original versus Replication” section in the supplementary material for a summary on differences in experimental procedures between our replication studies and the original study.

Study Designs

Our studies followed a 3 (payment form: cash vs. candy vs. monetized candy) x 2 (payment level: small vs. medium) design. In Study 1, we employed a between-subject design, similar to the original study. In Study 2, we made adjustments to a within-subject design. In addition to the six conditions, we had two control conditions, therefore, eight experimental conditions in total.

In the original study, however, they employed the same 3 x 2 between-subject design with one control condition where no payment was introduced. In addition to the original seven conditions, we introduced a new control condition as an extension; in the original study, there was a control condition where a helper would not be paid at all. Nevertheless, the authors did not clearly report whether they explicitly instructed participants that no payment would be given for helping. We sought to address the ambiguity in the original study and, thus, had two control conditions: nonpayment-without-mention and nonpayment-with-mention conditions. In the former condition, we did not mention payment at all. By contrast, in the latter, we told participants that no payment would be provided. The explicit mention of the absence of monetary compensations would prime the money market relationship. In the original study, it was hypothesized that their control condition would induce the social market relationship and that the expected willingness to help would be similar in the control and candy conditions (social market relationship). Thus, it seems that they did not explicitly mention nonpayment, and we decided to use the nonpayment-without-mention condition as a reference group for hypothesis testing. We summarized the operationalized hypotheses in Table 2.

Experimental Vignettes and Hypotheses

Following Heyman and Ariely (2004), we constructed eight scenarios in which a person was seeking someone to help load a sofa into a van. In the scenarios, we manipulated the payment form and level (see Table 3), and we used them for both studies.

Table 2

Replications and extensions: Summary of hypotheses

Replication (Studies 1 and 2)	
Hypothesis 1	The relationship between the payment level and the expected willingness to help is different in social vs. money market relationships.
Hypothesis 1a	In the cash condition, the expected willingness to help increases with the payment level
Hypothesis 1b	In the candy condition, the expected willingness to help is unaffected by the payment level and remains high. [null hypothesis]
Hypothesis 1c	The expected willingness to help in the nonpayment condition is higher than in the low monetary payment condition.
Hypothesis 2	Monetized candy is processed as a money market mindset, thereby resulting in the same pattern as predicted by the money market hypothesis (H1a).
Extension (Study 1)	
Hypothesis 3	There is an interaction between the form and level of payment on the expected joy.
Hypothesis 3a	In the cash condition, expected joy is higher when the payment level is medium compared to when it is low.
Hypothesis 3b	In the candy condition, the payment level does not affect the expected joy. [null hypothesis]
Hypothesis 4	The expected joy is higher in a social market relationship (i.e., the candy condition) than in a money market relationship (i.e., the cash and monetized candy conditions).

Table 3

Experimental conditions: Summary

Payment Form	Payment level	Instruction
Imagine that you see a person looking for someone to help load a sofa into a van.		
Cash	low	Those helping the person load the sofa into the van will receive cash payment ($0.5) in return.
	medium	Those helping the person to load the sofa into the van will receive cash payment ($5) in return.
Candy	low	Those helping the person load the sofa into the van will receive a candy bar in return.
	medium	Those helping the person load the sofa into the van will receive a chocolate box in return.
Monetized candy	low	Those helping the person load the sofa into the van will receive a candy bar that costs $0.5 in return.
	medium	Those helping the person load the sofa into the van will receive a chocolate bar* that costs $5 in return.
Nonpayment-without-mention		No further instruction was given.
Nonpayment-with-mention		Those helping the person to load a sofa into the van will receive **no payment** afterwards. [Extension condition]

Note. We used the nonpayment-without-mention condition as a control group. Thanks to a careful reviewer, we noticed that in both studies, the monetized candy conditions use a chocolate *bar* instead of a chocolate *box*. Given that we successfully replicated and given the results in both studies - in both the between- and within-subject design replications - showed very similar results for candy and monetized candy (see Figure 1), our conclusion is that it mattered very little whether it was a chocolate bar or chocolate box.

Study 1 (Original: Between-subject design)

Method

Extension: Joy

In Study 1, we introduced a new dependent variable, joy, as an extension; previous studies have demonstrated that altruistic helping leads to satisfaction and happiness (Dunn et al., 2008, 2014; Weinstein & Ryan, 2010; Yamaguchi et al., 2016). Heyman and Ariely (2004) originally argued that altruistic motives would underlie helping in the social market relationship, and we predicted that the level of joy individuals in the social market relationship would experience while helping others would not depend on the amount of reward they receive (H3b). Contrastingly, the level of reward would influence the level of joy for those in the money market relationship (H3a). In addition, given that altruistic behavior (i.e., helping) leads to satisfaction, it can be predicted that people expect helpers to experience more joy in the social market condition than in the money market condition (H4). To clarify H3 and H4, we would like to note that we were concerned about the simple effect of payment level in each payment form condition for H3, and we focused on the main effect of the payment form for H4. The inclusion of the new variable, joy, allowed us to test Heyman and Ariely's (2004) claim, using direct emotions assessment. See Table 2 for pre-registered hypotheses.

Procedure

After giving consent, participants were randomly assigned to one of the eight conditions and asked to read a corresponding scenario. Then, they rated how likely others would help the person in the scenario - "*How likely is the average person to help load the sofa into the van in return for ...*" (1 = *Not likely at all*, 11 = *Will help for sure)*. Next, we asked them to indicate how likely a person who helped and did not help in the given scenario would experience joy and regret - "*if a person were to decide [not to help/to help] in that scenario, to what extent do you think that person would experience [regret over not helping / joy over helping]*," respectively (1 = *Not at all*, 11 = *Extremely likely*). We measured regret for exploratory purposes, and auxiliary analyses on regret can be found on the OSF project page.

Results

Based on LeBel et al.'s (2019) criteria for evaluation of replications (see "Lebel's criteria for evaluation of replications" in Supplementary), we compared effect sizes from our hypothesis testing with those in the original study (see Tables 4 and 5). Since Heyman and Ariely (2004) did not report effect sizes, we calculated those using reported F statistics, cell means, and standard errors (see "Effect size calculation" in Supplementary). Following our pre-registration, no data exclusion was performed. Unless explicitly mentioned in the manuscript, we did not deviate from pre-registered analytic plans. We used JAMOVI (version 1.6.3), and R (version 3.6.3) for statistical analyses. We report a 95% CI for *Cohen's d* and a 90% CI for η_p^2. Because the latter can only be positive, and a 90% CI would be equivalent to a 95% CI for *Cohen's d*. Our preregistered analyses were produced using JAMOVI, and we used R codes to compute effect sizes and their CIs that JAMOVI did not calculate (e.g., post hoc comparisons using estimated marginal means). We also used an online effect size calculator where appropriate: https://effect-size-calculator.herokuapp.com/.

Replication

We first conducted a 3 (payment form: cash vs. candy vs. monetized candy) x 2 (payment level: low vs. medium) between-subject ANOVA on the expected willingness to help (see Figure 1 for descriptive statistics). It yielded support for a very weak main effect of the payment form *F*(2, 1643) = 3.33, *p* = .04, η_p^2 = .004, 90% CI [.0001, .01]. The main effect of the payment level was large, *F*(1, 1643) = 189.53, *p* < .001, η_p^2 = .10, 90% CI [.08, .13]. Moreover, there was support for an interaction, *F*(2, 1643) = 31.96, *p* < .001, η_p^2

= .04, 90% CI [.02, .05], supporting H1. We conducted planned pairwise comparisons using estimated marginal means to directly address H1a and H1b (see Tables 4 and 5). Reported p-values were adjusted with the Tukey method.

Table 4

Summary of findings: Replication versus original

Replication	*t*	df	*p*	Mean Difference	*Cohen's d* (replication)	*Cohen's d* (original)
H1a: low-cash condition vs. medium-cash condition on expected willingness to help						
Study 1	-14.45	1643	< .001	-3.52	-1.25 [-1.43, -1.06]	-0.59 [-0.89, -0.29]
Study 2	-41.13	998	< .001	-2.63	-1.30 [-1.39, -1.22]	
H1b: low-candy condition vs. medium-candy condition on expected willingness to help						
Study 1	-4.81	1643	< .001	-1.17	-0.43 [-0.60, -0,26]	0.25 [-0.55, 0.05]
Study 2	-27.37	998	< .001	-1.14	-0.87 [-0.94, -0.79]	
H1c: non-payment-without-mention condition vs. low-cash condition on expected willingness to help						
Study 1	10.56	529.64	< .001	2.52	0.90 [0.73, 1.08]	0.68 [0.38, 0.99]
Study 2	-3.59	998	< .001	-0.21	-0.11 [-0.18, -0.05]	
H2: low-monetized candy condition vs. medium-monetized candy condition on expected willingness to help						
Study 1	-4.56	1643	< .001	-1.11	-0.37 [-0.54, -0.20]	-0.60 [-0.89, -0.29]
Study 2	-25.39	998	< .001	-1.24	-0.80 [-0.87, -0.73]	
Extensions	*t*	df	*p*	Mean Difference	*Cohen's d* (replication)	*Cohen's d* (original)
H3a: low-cash condition vs. medium-cash condition on joy						
	-4.74	1643	< .001	-1.08	-0.40 [-0.57, -0.23]	na
H3a: low-monetized candy condition vs. medium-monetized candy condition on joy						
	-2.25	1643	0.21	-0.51	-0.18 [-0.35, -0.02]	na
H3b: low-candy condition vs. medium-candy condition on joy						
	-2.18	1643	0.25	-0.50	-0.20 [-0.36, -0.03]	na
H4: candy condition vs. cash condition on joy						
	4.54	1643	< .001	0.73	0.28 [0.16, 0.40]	na
H4: candy condition vs. monetized candy condition on joy						
	2.46	1643	0.004	0.34	0.15 [0.03, 0.27]	na

Note. See Table 5 for interpretation of these results.
CI: 95% confidence interval. In this table, hypotheses are simplified and described as "condition X vs. condition Y."
A negative mean difference indicates that participants scored higher in condition Y than in condition X. Likewise, a positive mean difference indicates that they scored higher in condition X than in condition Y.
na = Extensions to the replication, no test conducted in the original study.

Table 5

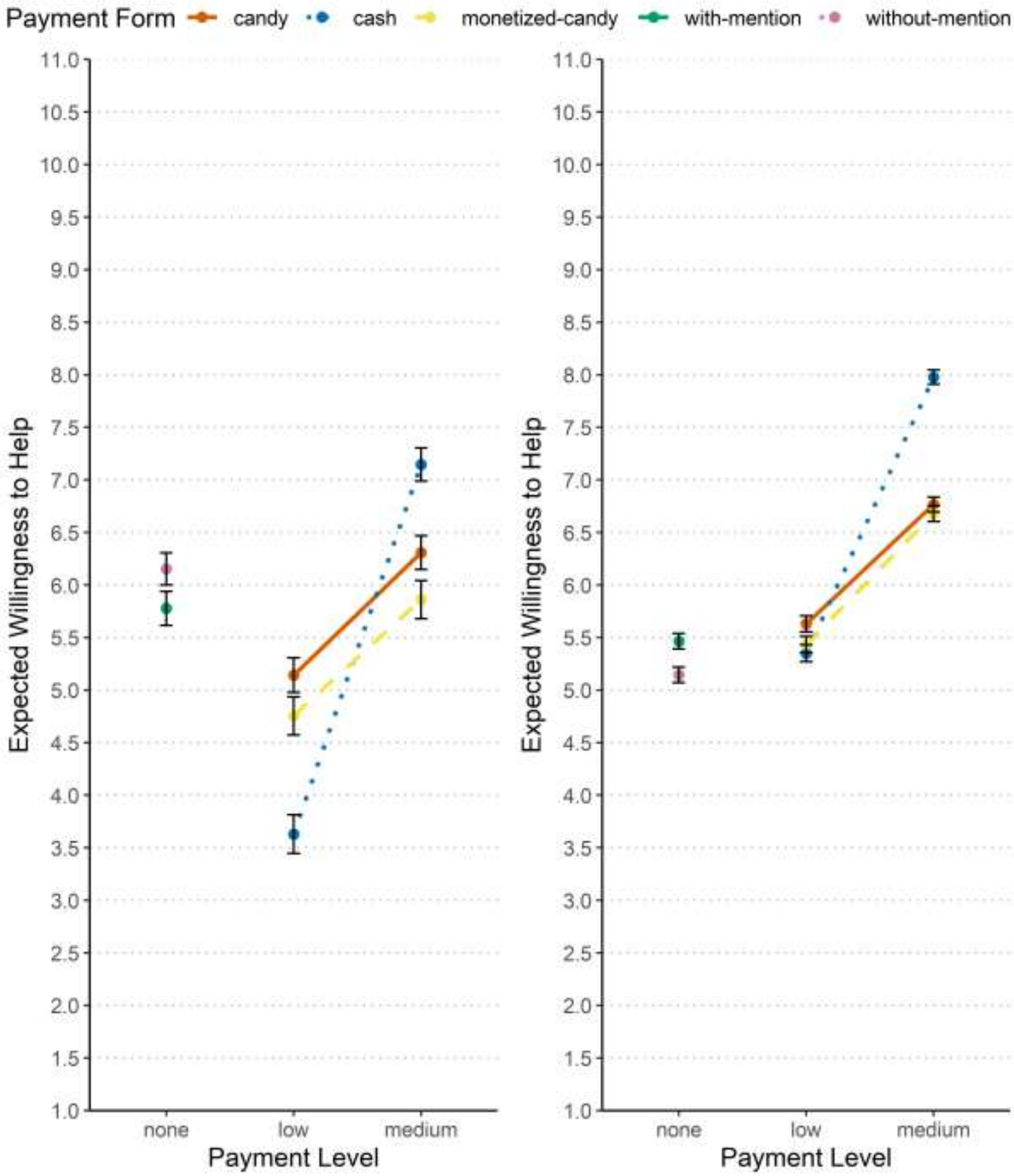
Replication Evaluation based on Lebel et al. (2019)

	Replication Evaluation	Interpretation
H1a: low-cash condition vs. medium-cash condition on expected willingness to help		
Study 1	[signal, inconsistent, larger]	Replication's ES 95% CI excludes 0 but also excludes original's ES. Replication's ES is larger than original's ES.
Study 2	[signal, inconsistent, larger]	Replication's ES 95% CI excludes 0 but also excludes original's ES. Replication's ES is larger than original's ES.
H1b: low-candy condition vs. medium-candy condition on expected willingness to help		
Study 1	[signal, inconsistent, negative]	Replication's ES 95% CI excludes 0 but also excludes original's ES. Replication's ES has a negative effect.
Study 2	[signal, inconsistent, negative]	Replication's ES 95% CI excludes 0 but also excludes original's ES. Replication's ES has a negative effect.
H1c: non-payment-without-mention condition vs. low-cash condition on expected willingness to help		
Study 1	[signal, inconsistent, larger]	Replication's ES 95% CI excludes 0 but also excludes original's ES. Replication's ES is larger than original's ES.
Study 2	[signal, inconsistent, opposite]	Replication's ES 95& CI excludes 0 but also excludes original's ES. Replication's ES is in the opposite direction relative to original's ES.
H2: low-monetized candy condition vs. medium-monetized candy condition on expected willingness to help		
Study 1	[signal, inconsistent, smaller]	Replication's ES 95% CI excludes 0 but also excludes original's ES. Replication's ES is smaller than original's ES.
Study 2	[signal, consistent]	Replication's ES 95% CI excludes 0 and includes original's ES.

Note. ES = effect size. Statistical details provided in Table 4. Further details on the details about the evaluation criteria using LeBel et al. (2019) provided in the supplementary material.

Figure 1

Studies 1 and 2: Expected willingness to help across conditions



Note. The figures on the left and right represent Studies 1 and 2, respectively. Error bars indicate standard errors. Study 1 employed a between-subject design, and Study 2 employed a within-subject design.

We found support for H1a, revealing that the expected willingness to help in the medium-cash condition (*M* = 7.15, *SD* = 2.60, *N* = 274) was higher than that in the low-cash condition (*M* = 3.63, *SD* = 3.02, *N* = 273), and the replication effect size was larger than the original one (see Tables 4 and 5). By contrast, whereas Heyman and Ariely's (2004) findings did not observe a difference between the low-candy (*M* = 5.14, *SD* = 2.74, *N* = 275) and medium-candy conditions (*M* = 6.31, *SD* = 2.66, *N* = 276), the CI for the effect size did not include zero in the present study, and we cannot conclude the absence of an effect (see Tables 4 and 5). Thus, we did not find support for H1b. H1b is a null hypothesis, and we therefore decided to conduct a non-preregistered Bayesian analysis to supplement the hypothesis testing. We report the results of the analysis at the end of the section. The effect of the payment size was larger in the cash condition than in the candy condition (see Tables 4 and 5), and the results suggest that participants in the cash condition (i.e., the money market relationship) were more sensitive to the change in the payment level than those in the candy condition (i.e., the social market relationship), thus, supporting H1.

To test H1c, we carried out a one-way between-subject Welch's ANOVA using the following four conditions: payment-low-cash vs. payment-medium-cash vs. nonpayment-with-notice vs. nonpayment-without-mention. In the pre-registration, we did not explicitly mention whether we would use a conventional Fisher's or Welch's ANOVA. However, given that Levene's test indicated that the assumption of equal variance was violated ($F(3, 1097) = 3.77$, $p = .01$), we opted for Welch's ANOVA as it would correct degrees of freedom for the violation. The analysis revealed a large main effect, $F(3, 608.17) = 72.88$, $p < .001$, $\eta^2 = .26$, 90% CI [.22, .31]. A post hoc comparison showed support for the differences between the nonpayment-without-mention (*M* = 6.15, *SD* = 2.55, *N* = 275) and low-cash (*M* = 3.63, *SD* = 3.02, *N* = 273) conditions (see Tables 4 and 5), with those in the former condition estimating the likelihood of helping higher than those in the latter. Thus, we found support for H1c and replicated the original finding with a larger effect size.

To address H2, we looked at the post hoc comparisons from the 3 x 2 ANOVA and found that the expected willingness to help in the medium monetized candy condition (*M* = 5.86, *SD* = 3.00, *N* = 274) was higher than that in the low monetized candy condition (*M* = 4.75, *SD* = 3.02, *N* = 277, see Tables 4 and 5). However, the effect size was much smaller than the original one, and the effect size of the payment level in the monetized candy condition (*d* = -0.37, 95% CI = [-0.54, -0.20]) was more similar to that in the candy condition (*d* = -0.43, 95% CI = [-0.60, -0,26]) than the cash condition (*d* = -1.25, 95% CI = [-1.43, -1.07]). Thus, we failed to find support for H2, where Heyman and Ariely (2004) predicted that monetized goods would prime the money market relationship, and the expected willingness to help in the cash and monetized goods condition would be similar.

H1b was a null hypothesis, predicting that the expected willingness to help in the low-candy condition would not be different from that in the medium candy condition. In addition to the pre-registered conventional hypothesis testing, we carried out an exploratory Bayesian t-test to directly test H1b. The null hypothesis was that there was no difference in the expected willingness to help between the small-candy and medium candy conditions, and the two-sided alternative hypothesis postulated that the expected willingness to help in the two candy conditions was different. The Bayes factor indicated that the data was in favor of the alternative hypothesis, B_{10} = 20841.04, yielding strong evidence against Heyman and Ariely's (2004) original finding that the level of reward in the candy conditions did not affect the expected willingness to help.

Extensions

For our extension (H3-H4), we conducted a 3 (payment form: cash vs. candy vs. monetized candy) x 2 (payment level: low vs medium) between-subject ANOVA on joy (see Table 6 and Figure 2 for descriptive statistics). We found support for main effects: payment form: *F*(2, 1643) = 10.32, $p < .001$, η_p^2 = .01, 95% CI [.005, .02]; payment level: *F*(1, 1643) = 28.10, $p < .001$, η_p^2 = .02, 95% CI [.01, .03]. However, the 95% CI for the interaction effect included zero, *F*(2, 1643) = 2.15, $p = .12$, η_p^2 = .003, 95% CI [.00, .01]. Thus, we did not find support for H3.

We conducted pre-registered planned comparisons to test H3a, H3b, and H4 (see Table 4). We found that in the cash condition, joy was higher when the payment level was medium (*M* = 7.00, *SD* = 2.46, *N* = 274) compared to when it was low (*M* = 5.92, *SD* = 2.92, *N* = 273, see Table 4). In comparison, the effect of the payment level was much weaker in the candy conditions (low-candy condition: *M* = 6.94, *SD* = 2.69, *N* = 275; medium-candy condition: *M* = 7.44, *SD* = 2.36, *N* = 276), and in the monetized candy conditions (low-monetized candy condition: *M* = 6.54, *SD* = 2.81, *N* = 277; medium-monetized candy condition: *M* = 7.05, *SD* = 2.77, *N* = 274; see Table 4). Consistent with H4, we found that perceived joy was higher in the candy condition than in the cash and monetized candy conditions.

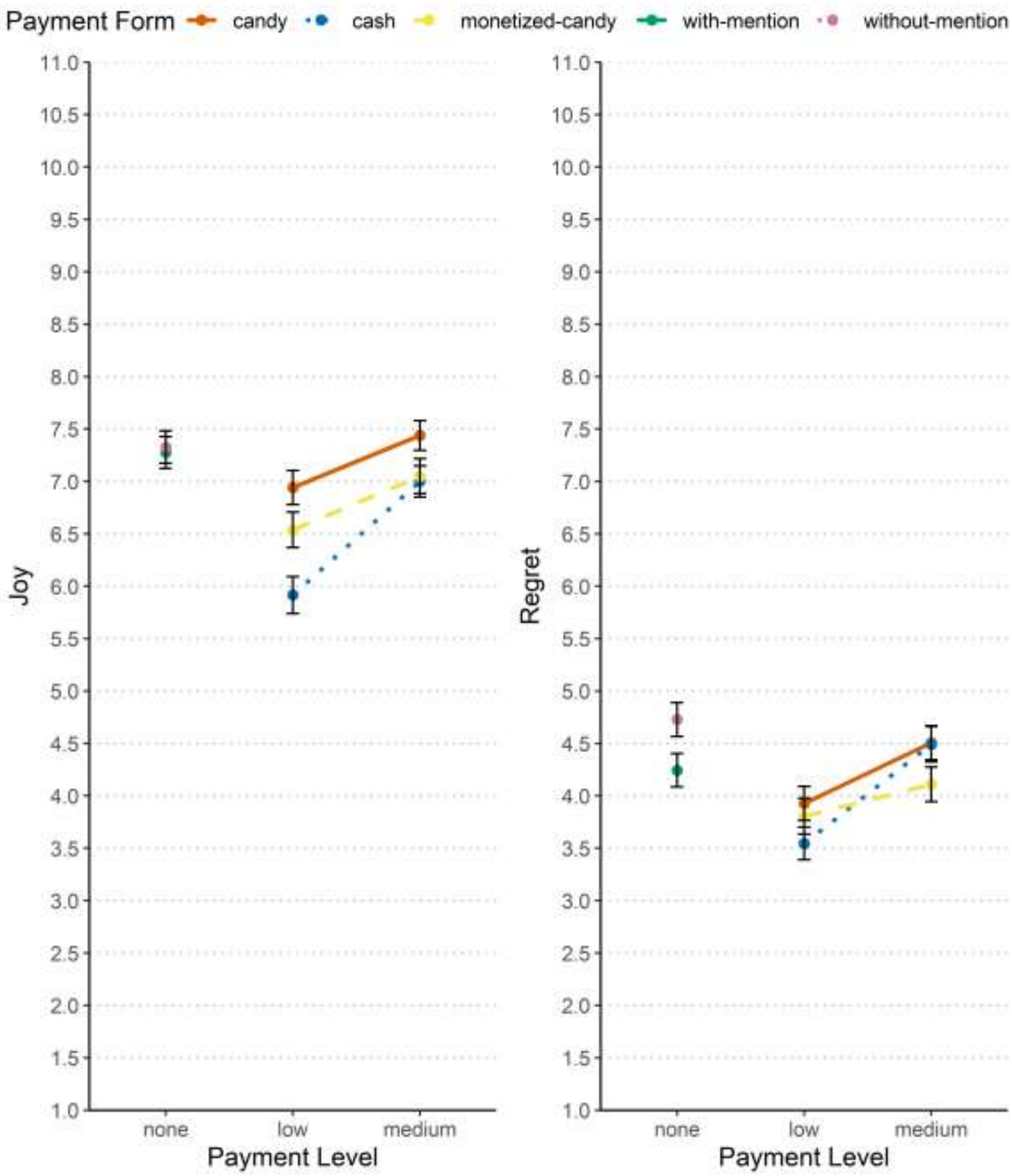
Since H3b was a null hypothesis, we conducted a non-preregistered Bayesian analysis to complement the t-test. The null hypothesis was that there was no difference in expected joy between the small-candy and medium-candy conditions, and the two-sided alternative hypothesis was that expected joy in the two conditions was different. The Bayes factor was B_{10} = 1.25. Thus, the analysis provided anecdotal evidence for the null hypothesis.

Finally, as an exploratory extension (not pre-registered), we compared the nonpayment-without-mention condition with the nonpayment-with-mention condition; we did not find support for differences in the expected willingness to help between the control conditions (nonpayment-without-mention: *M* = 6.15, *SD* = 2.55, *N* = 275; nonpayment-with-mention: *M* = 5.78, *SD* = 2.70, *N* = 279), *t*(551.07) = -1.68, *p* = .09, *d* = -0.14, 95%CI [-0,31, 0.02] (see Table 4).

We report results of other pre-registered analyses in the supplementary material (see Supplementary Results for Study 1 in supplementary material).

Figure 2

Study 1 extensions: Joy and Regret across conditions



Note. The figures on the left and right visualize the perceived joy and regret, respectively. Error bars indicate standard errors.

Discussion

We found support for H1; the effect of the payment level was different across the money and social market relationships. However, whereas Heyman and Ariely (2004) found that individuals in the social market relationship were insensitive to the payment level, our results suggested that they were indeed sensitive in the social market relationship but less so compared to those in the money market relationship. In addition, we successfully replicated H1c with a larger effect size and revealed that the expected willingness to help in the low cash payment condition was lower than the nonpayment control condition, suggesting that the low level of monetary incentive would be counterproductive. However, contrary to Heyman and Ariely (2004), the monetized candy did not induce the money market relationship, as the effect of the payment level in the monetized candy condition was more similar to that in the candy condition than in the cash condition.

As an extension, we measured to what extent participants thought a helper in the scenario would experience joy. As expected, the level of payment did not influence joy in the social market relationship (i.e., the candy condition), where the willingness to help others is primarily driven by internal, altruistic motivations. We predicted that the level of payment should influence joy in the money market relationship (i.e., the cash and monetized candy conditions). We found support for an effect for joy in the cash condition, yet failed to find support for an effect in the monetized candy condition. Thus, this casts doubt on Heyman and Ariely’s (2004) argument that monetized goods would prime the money market relationship. Our results, overall, suggest that monetized goods would fall under the social market relationship.

Study 2 (Extension: Within-subject design)

Method

We conducted Study 2 using a within-subject design; Charness et al. (2012) pointed to the importance of a choice of experimental design (between vs. within) in various decision-making tasks, showing that while some effects and decision-making processes were insensitive to experimental design, others were sensitive. Thus, a comparison of results from between- and within-subject design experiments is a sensible step that can further shed light on the robustness and generalizability of the findings.

Procedure

After giving consent, participants were first presented with a vignette of the nonpayment-without-mention condition (i.e., the control condition) and indicated the likelihood that others would help in the scenario. This design was meant to ensure the control condition is not affected by carryover effects from the other conditions. Then, they were shown the other seven experimental scenarios in a randomized order and answered the dependent measure for each scenario. We did not include joy and regret but otherwise employed the same measures and experimental instructions as in Study 1.

In the pre-registration manuscript, we proposed to use a 7-point scale to measure the expected willingness to help, while Study A and the original study used an 11-point scale (“replication & extension main manuscript – Heyman & Ariely, 2004 – Group B.docx”, page 19). However, in the pre-registered study materials, we planned to use an 11-point scale, and we did conduct the study with the 11-point scale. This was an oversight misalignment between the registered manuscript and the registered survey materials.

Results

Unless explicitly mentioned, we followed the pre-registered analysis plans. We first carried out a 3 (payment form: cash vs. candy vs. monetized candy) x 2 (payment level: low vs. medium) within-subject ANOVA on the expected willingness to help (see Figure 1 for descriptive statistics). Mauchly's test revealed that the assumption of sphericity was violated for the main effect of the payment form and the interaction term, and we employed Greenhouse-Geisser corrected degrees of freedom. We found a large effect for the payment form, $F(1.76, 1752.33) = 161.56$, $p < .001$, $\eta_p^2 = .14$, 90% CI [.12, .16]. The main effect of the payment level was also large, $F(1, 998) = 1679.32$, $p < .001$, $\eta_p^2 = .63$, 90% CI [.60, .65]. Finally, we found an interaction effect, $F(1.90, 1900.34) = 428.58$, $p < .001$, $\eta_p^2 = .30$, 90% CI [.27, .33], supporting H1.

We then carried out planned pairwise comparisons to directly address hypotheses (see Table 4). For Study 1, following our preregistered plan, we conducted ANOVAs with post hoc comparisons using estimated marginal means with the Tukey correction. By contrast, for Study 2, we preregistered that we would run simple paired t-tests using raw means, and that p-values would not be adjusted. As we did not include any covariates in ANOVAs in Study 1 and the number of participants in each cell did not substantially vary, the use of different types of means would not be a problem. Thus, except for the presence of p-value adjustment, these different analytic strategies yielded compatible results. For replication evaluation, we focused on effect sizes and their CIs rather than p values, and we did not deviate from the pre-registered analytic strategies.

First, we found support for H1a; the expected willingness to help was higher in the medium-cash condition (*M* = 7.98, *SD* = 2.17) compared to the low-cash condition (*M* = 5.35, *SD* = 2.54). Moreover, the effect size was bigger than the original one (see Table 4).

However, we found that the increase in the payment level in the candy condition also resulted in higher expected willingness to help (low-candy: *M* = 5.63, *SD* = 2.41; medium-candy: *M* = 6.77, *SD* = 2.24), conflicting with H1b (see Table 4). Regarding H1c, whereas Heyman and Ariely (2004) demonstrated that the expected willingness to help in the control condition was higher than that in the low monetary payment condition, we found support for an effect in the opposite direction (nonpayment-without-mention: *M* = 5.15, *SD* = 2.35; low-cash: *M* = 5.35, *SD* = 2.54). Thus, H1c was not supported (see Table 4).

We further conducted a pairwise comparison to address H2 and found that as in the cash condition, the expected willingness to help in the medium-monetized candy condition (*M* = 6.68, *SD* = 2.36) was higher than that in the low-monetized candy condition (*M* = 5.43, *SD* = 2.42, see Table 4), consistent with Heyman and Ariely (2004). However, as in Study 1, the effect size of the payment level in the monetized candy condition (*d* = -0.80, 95% CI = [-0.89, -0.71]) was more similar to that in the candy condition (*d* = -0.87, 95% CI = [-0.96, -0.78]), rather than the cash condition (*d* = -1.30, 95% CI = [-1.39, -1.20]). Overall, these results did not support H2.

Finally, we compared the expected willingness to help in the nonpayment-without-mention condition (*M* = 5.15, *SD* = 2.35) with that in the nonpayment-with-mention-condition (*M* = 5.47, *SD* = 2.36). We conducted a paired sample *t*-test and revealed that the expected willingness to help was higher in the latter condition than in the former, *t*(998) = 7.73, *p* < .001, *d* = 0.25, 95% CI [0.24, 0.40] (see Table 4). We report results from other pre-registered analyses in the supplementary material (see Supplementary).

Discussion

Overall, our results suggested that the expected willingness to help was higher when the payment level was medium compared to when it was low, regardless of the payment form. This supports H1 and is consistent with Heyman and Ariely’s (2004) core argument that the effect of the payment level would vary depending on the payment form. We replicated the effect of the payment level in the cash condition (H1a) with larger effect size and found that the expected willingness to help was higher in the low cash condition than in the medium cash condition. However, we did not find support for H1b; Heyman and Ariely (2004) found that people were insensitive to the payment level in the candy condition, yet our results showed that participants were sensitive to payment but less so in the candy condition than in the cash condition. Moreover, while we replicated the effect of the payment level in the monetized-candy condition, the effect size was similar to that in the candy condition, failing to support H2. We also failed to find support for H1c. In the original study, they found that the expected willingness to help was higher in the control condition compared to the low cash condition. Nevertheless, our replication revealed that it was the opposite; the expected willingness to help was higher in the low cash condition than in the nonpayment-without-mention condition.

In Studies 1 and 2 we employed different experimental designs, and these yielded mostly converging results, yet a discrepancy emerged in the control conditions; the expected willingness to help in the control conditions was lowest in the present study, whereas it was high in the original study. One possible explanation may be our choice of the experimental design; in our Study 2 using a within-subject design, participants were first presented with the nonpayment-without-mention condition and then shown the remaining seven scenarios in a randomized order. Presumably, participants used their judgment in the control condition as a baseline; they perceived low-level payments of any kind as being more attractive than nonpayment, and this might have inflated the expected willingness to help in the experimental conditions relative to that in the control conditions. Participants in Study 2 perceived even small compensations as more attractive than receiving nothing. This might be explained by individuals' stronger sensitivity to the size of the payment in joint evaluation mode (i.e., the within-subject design) compared to single evaluation mode (i.e., between-subject design) (Anvari et al., 2021; Hsee & Zhang, 2010).

Alternatively, it is possible that the results of Study 2 were affected by demand effects. Participants saw the least attractive option first (nonpayment-without-mention condition) and were then presented with more attractive scenarios in which rewards were given. This might have led to participants guessing the study's goal (i.e., whether rewards would increase people's motivation to help others), and then to their responding in a way that would help achieve the desired outcome. Overall, the discrepancy between Studies 1 and 2 regarding H1c could be attributed to our choice of experimental design and our chosen order of display in Study 2.

Conclusion

Heyman and Ariely (2004) claimed that the effect of the payment level would depend on the payment form. More specifically, they found that individuals expected others to be more willing to help in the money market relationship, but not in the social market relationship. In the two well-powered replication studies, we revealed that the higher payment was more effective in increasing people's perceptions of willingness to help regardless of the market relationship (i.e., regardless of whether rewards were provided as cash, goods, or monetize goods). Notably, we found that the effect of the payment level was much larger when paid in cash than when paid with goods. Thus, while we found support for Heyman and Ariely's (2004) main argument that the effect of the payment level would vary across different market relationships, we did not replicate their finding that the payment level did not matter in the social market relationship. Moreover, Heyman and Ariely (2004) held that monetized goods (goods with a price tag) would prime the money market relationship, yet our results supported an opposite effect. The discrepancy between the original and the replication studies is of practical importance, suggesting that people perceive that more is better, and that this is especially true for cash. These findings help update knowledge regarding how payment form and level are related to expected willingness to help, denoting the value of replication studies.

Finally, we note that our design mirrored that of the original and that the large effects reported in our studies using behavioral intention proxies should not be taken to suggest that the increase in the payment level would make an observable and substantial impact outside of controlled laboratory settings. Moreover, these studies focused on small and medium incentives, and it would be a relevant avenue for future research whether these findings would hold when comparing, for instance, medium and large payment levels. Therefore, now that these findings have been revisited and adjusted, we see promise in further follow-up replications and studies that would extend these to examine higher stakes and practical implications.

References

Anvari, F., Olsen, J., Hung, W. Y., & Feldman, G. (2021). Misprediction of affective outcomes due to different evaluation modes: Replication and extension of two distinction bias experiments by Hsee and Zhang (2004). *Journal of Experimental Social Psychology*, *92*, 104052. https://doi.org/10.1016/j.jesp.2020.104052

Bauer, P. J., & Ariely, D. (2021). Expression of Concern: Effort for Payment: A Tale of Two Markets. *Psychological Science*, https://doi.org/10.1177/09567976211035782

Bowles, S., & Polanía-Reyes, S. (2012). Economic incentives and social preferences: Substitutes or complements? In *Journal of Economic Literature* (Vol. 50, Issue 2, pp. 368–425). https://doi.org/10.1257/jel.50.2.368

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior and Organization*, *81*(1), 1–8. https://doi.org/10.1016/j.jebo.2011.08.009

Cifor. (2005). *Payment for environmental services: some nuts and bolts*. Bogor, Indonesia: CIFOR. https://vtechworks.lib.vt.edu/handle/10919/66932

Dahill-Brown, S. E., Witte, J. F., & Wolfe, B. (2016). Income and access to higher education: Are high quality universities becoming more or less elite? A longitudinal case study of admissions at UW-Madison. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, *2*(1), 69–89. https://doi.org/10.7758/rsf.2016.2.1.04

Dunn, E. W., Aknin, L. B., & Norton, M. I. (2008). Spending money on others promotes happiness. *Science*, *319*(5870), 1687–1688. https://doi.org/10.1126/science.1150952

Dunn, E. W., Aknin, L. B., & Norton, M. I. (2014). Prosocial Spending and Happiness: Using Money to Benefit Others Pays Off. *Current Directions in Psychological Science*, *23*(1), 41–47. https://doi.org/10.1177/0963721413512503

Fiske, A. P. (1992). The four elementary forms of sociality: Framework for a unified theory of social relations. *Psychological Review, 99*(4), 689–723. https://doi.org/10.1037/0033-295X.99.4.689

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and Why Incentives (Don't) Work to Modify Behavior. *Journal of Economic Perspectives*, *25*(4), 191–210. https://doi.org/10.1257/jep.25.4.191

Heyman, J., & Ariely, D. (2004). Effort for payment - A tale of two markets. In *Psychological Science* (Vol. 15, Issue 11, pp. 787–793). Sage CA: Los Angeles, CA. https://doi.org/10.1111/j.0956-7976.2004.00757.x

Hsee, C. K., & Zhang, J. (2010). General evaluability theory. Perspectives on psychological science, 5(4), 343-355. https://doi.org/10.1177/1745691610374586

Lacetera, N., & Macis, M. (2010). Social image concerns and prosocial behavior: Field evidence from a nonlinear incentive scheme. *Journal of Economic Behavior and Organization*, *76*(2), 225–237. https://doi.org/10.1016/j.jebo.2010.08.007

LeBel, Etienne P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). A Unified Framework to Quantify the Credibility of Scientific Findings. *Advances in Methods and Practices in Psychological Science*, *1*(3), 389–402. https://doi.org/10.1177/2515245918787489

LeBel, Etienne Philippe, Vanpaemel, W., Cheung, I., & Campbell, L. (2019). A Brief Guide to Evaluate Replications. *Meta-Psychology, 3*. https://doi.org/10.15626/mp.2018.843

Liu, S. H., Liao, H. L., Sung, Y. H., & Peng, Q. D. (2012). Communal and exchange relationships and the effects of norms on internet participation. *Social Behavior and Personality, 40*(6), 993–1004. https://doi.org/10.2224/sbp.2012.40.6.993

Newman, G. E., & Shen, Y. J. (2011). When Do Gifts Help Charitable Giving and When Do They Hurt? In *ACR North American Advances* (Vol. 39). Association for Consumer Research. http://www.acrwebsite.org/volumes/1009992/volumes/v39/NA-

39http://www.copyright.com/.

Shampanier, K., Mazar, N., & Ariely, D. (2007). Zero as a special price: The true value of free products. *Marketing Science*, *26*(6), 742–757. https://doi.org/10.1287/mksc.1060.0254

Simonsohn, U. (2015). Small Telescopes: Detectability and the Evaluation of Replication Results. *Psychological Science*, *26*(5), 559–569. https://doi.org/10.1177/0956797614567341

Snyder, T. D. S. (2012). Digest of Education Statistics, 2011. National Center for Education Statistics.

Weinstein, N., & Ryan, R. M. (2010). When Helping Helps: Autonomous Motivation for Prosocial Behavior and Its Influence on Well-Being for the Helper and Recipient. *Journal of Personality and Social Psychology*, *98*(2), 222–244. https://doi.org/10.1037/a0016984

Wunder, S. (2007). The efficiency of payments for environmental services in tropical conservation: Essays. In *Conservation Biology* (Vol. 21, Issue 1, pp. 48–58). https://doi.org/10.1111/j.1523-1739.2006.00559.x

Yam, K. C., Bumpus, M. F., & Hill, L. G. (2012). Motivating effort: A theoretical synthesis of the self-sufficiency and two-market theories. *British Journal of Social Psychology*, *51*(4), 709–716. https://doi.org/10.1111/j.2044-8309.2011.02067.x

Yamaguchi, M., Masuchi, A., Nakanishi, D., Suga, S., Konishi, N., Yu, Y. Y., & Ohtsubo, Y. (2016). Experiential purchases and prosocial spending promote happiness by enhancing social relationships. *Journal of Positive Psychology*, *11*(5), 480–488. https://doi.org/10.1080/17439760.2015.1117128

Heyman and Ariely (2004) Replications and extensions: Supplementary Materials

Contents

Erroneous reporting in Heyman and Ariely (2004)

In Heyman and Ariely's (2004) reporting of their results, there are several potential errors. We alerted Psychological Science regarding these issues, and these are now discussed in an official expression of concern by the original authors and Psychological Science editor in chief: Expression of Concern: Effort for Payment: A Tale of Two Markets, DOI: 10.1177/09567976211035782

Oversights detected in our replications' stimuli

Thanks to a careful reviewer we noticed that in both our studies the monetized candy conditions use a chocolate *bar* instead of a chocolate *box*.

Given that we successfully replicated and given the results in both studies - in both the between and within subject design replications - showed very similar results for candy and monetized candy (see Figure 1) then our conclusion is that it mattered very little whether it was a chocolate bar or chocolate box.

Original versus Replication: Adjustments and deviations

Study 1 and Original

	Original	Replication	Reason for change
Analytic approach	Between-participants experimental design with unknown statistical tests	A two-way between-subject ANOVA with follow-up post-hoc pairwise comparisons	The original analytic approaches are ambiguous and cannot be reproduced.
Procedure	Information was inadequate regarding the randomizing procedures. In addition, it is not clear whether an experimenter was blind to conditions.	Participants were randomly assigned into one out of the eight different conditions.	Given the inadequate information about the randomization, we did the best experimental approach.
Conditions	2x3 factorial design plus 1 control condition	2x3 factorial plus 2 control conditions	The original article provided inadequate information about how they phrased their instruction for the control. To disentangle this, as an extension, we added a new control condition.

Study 2 and original

	Original	Replication	Reason for change
Study design, procedure, and analytic approach	between-subject	within-subject	As an extension, we decided to employ a within-subject design. See the manuscript for the rationale. Because of the choice, experimental procedure and analytic approach, correspondingly, differ from those for the original study.
Conditions	2x3 factorial design plus 1 control condition	2x3 factorial plus 2 control conditions	The original article provided inadequate information about how they phrased their instruction for the control. To disentangle this, as an extension, we added a new control condition.

Supplementary Results

Supplementary Results for Study 1

We report the results of four comparisons that Heyman and Ariely also reported. We first conducted a 3 (payment form: cash vs. candy vs. monetized candy) x 2 (payment level: low vs. medium) between-subject ANOVA on the perceived willingness to help (see the main text) and compared the estimated marginal means for the main effect of the payment form. We successfully replicated the original findings, but the effect size for the comparison between the monetized candy and cash condition was smaller than the original effect size.

In addition, we conducted a 1 x 4 (payment: low-candy vs. medium-candy vs. nonpayment-with-mention vs. nonpayment-without- mention) between-subject ANOVA found a significant effect, $F(3, 611.49) = 10.20$, $p < .001$, $\eta_p^2 = .05$. Post hoc comparisons revealed that the difference between the nonpayment-without-mention and low-candy conditions was significant, with those in the former condition estimating the likelihood of helping higher than those in the latter (see Comparison 3 in the table below). This is inconsistent with Heyman and Ariely (2004) that did not find a significant difference between the two conditions.

Finally, we conducted a 1 x 4 (payment: low-monetized candy vs. medium-monetized candy vs. nonpayment-with- mention vs. nonpayment-without- mention) between-subject Welch's ANOVA. The main effect was significant, $F(3, 610.12) = 12.23$, $p < .001$, $\eta_p^2 = .06$. Post hoc comparisons revealed that the expected willingness to help in the nonpayment-without-mention condition was significantly higher than that in the low monetized candy condition, replicating the original finding. Yet, the effect size was smaller than the original effect size.

On a side note, expected willingness to help was positively correlated with perceived joy ($r = .44$, 95% CI [.41, .47], $p < .001$) and regret ($r = .42$, 95% CI [.38, .45], $p < .001$). Perceived joy and regret were also correlated, $r = .27$, 95% CI [.23, .31], $p < .001$.

t	df	*p*	Mean Difference	*Cohen's d* and CI	*Cohen's d* and CI (Heyman and Ariely, 2004)	Replication Evaluation
Comparison 1: monetized candy condition vs. cash condition on expected willingness to help						
-0.47	1643	.89	-0.08	-0.03 [-0.15, 0.09]	-0.10 [-0.30, 0.11]	[no-signal, consistent]
Comparison 2: monetized candy condition vs. candy condition on expected willingness to help						
-2.43	1643	.04	-0.42	-0.15 [-0.27, -0.03]	-0.34 [-0.55, -0.13]	[signal, inconsistent, smaller]
Comparison 3: nonpayment-without-mention condition vs. low-candy condition on expected willingness to help						
4.46	1101	< .001	1.01	0.38 [0.21, 0.55]	0.04 [-0.26, 0.33]	[signal, inconsistent, positive effect]
Comparison 4: nonpayment-without-mention condition vs. low-monetized candy condition on expected willingness to help						
5.88	536	< .001	1.40	0.50 [0.33, 0.67]	0.86 [0.55, 1.17]	[signal, inconsistent, smaller]

Note: CI: 95% confidence interval. In this table, hypotheses are simplified and described as “condition X vs. condition Y.” A negative mean difference indicates that participants scored higher in condition Y than in condition X. Likewise, a positive mean difference indicates that they scored higher in condition X than in condition Y.

Supplementary Results for Study 2

As for Study 1, we performed four pre-registered pairwise comparisons. First, we examined the main effect of the payment form, comparing the monetized candy condition with the cash and candy condition.

We then compared the expected willingness help in the nonpayment-without-mention condition with that in the low candy and low monetized candy conditions. As reported in the main text, we did not replicate these comparisons because the expected willingness to help in the former condition was the lowest amongst all the conditions.

t	df	*p*	Mean Difference	*Cohen's d* and CI	*Cohen's d* and CI (Heyman and Ariely, 2004)	Replication Evaluation
Comparison 1: monetized candy condition vs. cash condition on expected willingness to help						
-17.29	998	< .001	-0.61	-0.55 [-0.61, -0.48]	-0.10 [-0.30, 0.11]	[Signal, inconsistent, negative effect]
Comparison 2: monetized candy condition vs. candy condition on expected willingness to help						
-4.82	998	< .001	-0.14	-0.15 [-0.22, -0.09]	-0.34 [-0.55, -0.13]	[Signal, inconsistent, smaller]
Comparison 3: nonpayment-without-mention condition vs. low-candy condition on expected willingness to help						
-9.51	998	< .001	-0.48	-0.30 [-0.39, -0.21]	0.04 [-0.26, 0.33]	[Signal, inconsistent, negative effect]
Comparison 4: nonpayment-without-mention condition vs. low-monetized candy condition on expected willingness to help						
-5.71	998	< .001	-0.29	-0.18 [-0.27, -0.09]	0.86 [0.55, 1.17]	[Signal, inconsistent, opposite effect]

Note: CI: 95% confidence interval. In this table, hypotheses are simplified and described as “condition X vs. condition Y.” A negative mean difference indicates that participants scored higher in condition Y than in condition X. Likewise, a positive mean difference indicates that they scored higher in condition X than in condition Y.

LeBel's criteria for evaluation of replications

We aimed to evaluate whether the original findings were successfully replicated, using LeBel's et al. (2019) criteria;

For situations where an original study detected a signal (i.e., a significant effect);

(1) Signal consistent: Replication 95% CI for an effect size excludes 0 and includes the original effect size point estimate.

(2) Signal inconsistent: Replication 95% CI for an effect size excludes 0 but also excludes the original effect size point estimate.

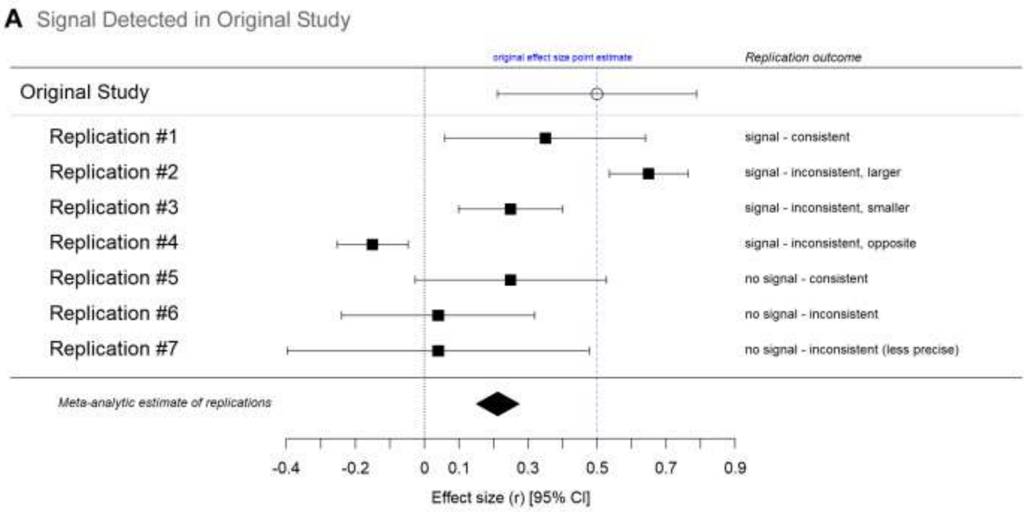
(2-1) Signal inconsistent blarger:

(2-2) Signal inconsistent smaller:

(2-3) Signal inconsistent opposite direction:

(3) No signal consistent: Replication 95% CI for an effect size includes 0 but also includes the original effect size point estimate.

(4) No signal inconsistent: Replication 95% CI for an effect size includes 0 but excludes the original effect size point estimate.



(from LeBel et al., 2019)

For situations where an original study did not detect a signal;

(1) No signal consistent: Replication 95% CI for an effect size includes 0 and the original effect size point estimate.

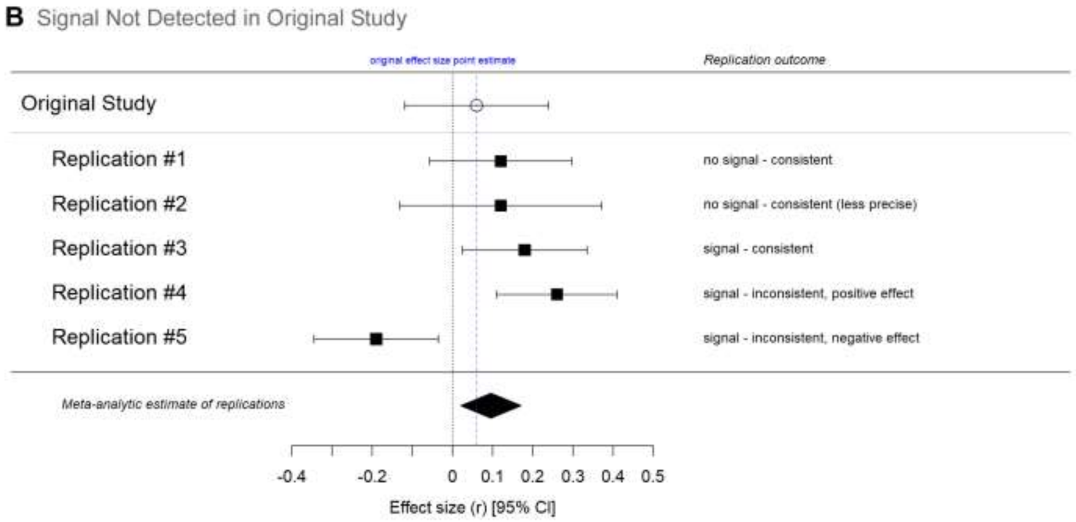
(2) No signal consistent (less precise): Replication 95% CI for an effect size includes 0 and the original effect size point estimate, but the replication effect size is less precise than in the original study.

(3) Signal consistent: Replication 95% CI for an effect size excludes 0 but includes the original effect size point estimate.

(4) Signal inconsistent: Replication 95% CI for an effect size excludes 0 and the original effect size point estimate.

(4-1) Signal inconsistent positive effect

(4-2) Signal inconsistent negative effect

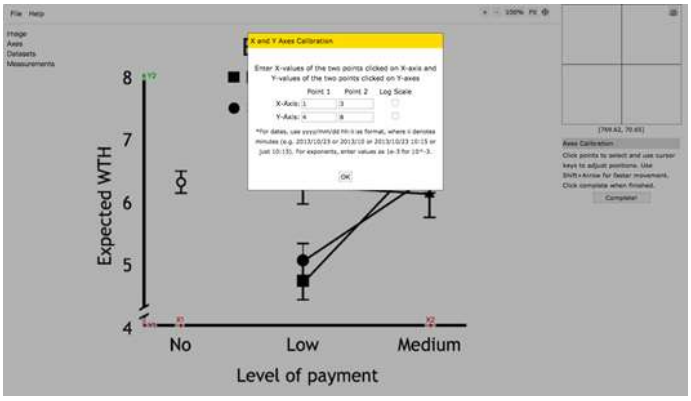


(from LeBel et al., 2019)

Effect size calculations

Original Study

We first computed means and standard errors in the original article, using WebPlotDigitizer;



X and Y Axes Calibration

Enter X-values of the two points clicked on X-axis and Y-values of the two points clicked on Y-axes

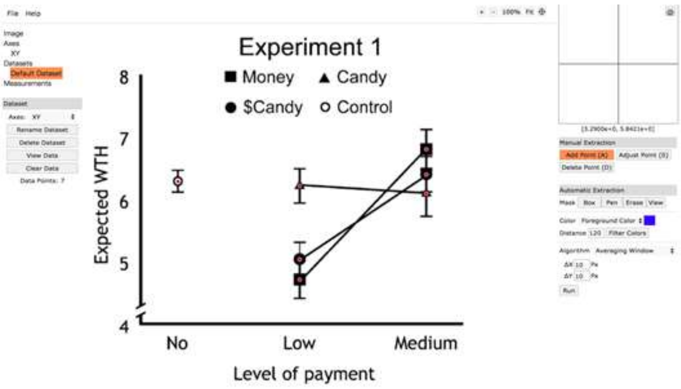
	Point 1	Point 2	Log Scale
X-Axis:	1	3	□
Y-Axis:	4	8	□

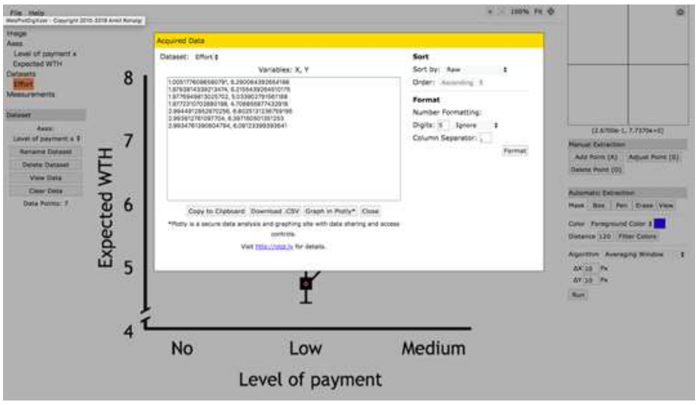
*For dates, use yyyy/mm/dd hh:ii:ss format, where ii denotes minutes (e.g. 2013/10/23 or 2013/10 or 2013/10/23 10:15 or just 10:15). For exponents, enter values as 1e-3 for 10^-3.

OK

Axes Calibration

Click points to select and use cursor keys to adjust positions. Use Shift+Arrow for faster movement. Click complete when finished.





Acquired Data

Dataset: Effort

Variables: X, Y

```
1.008177608680791, 6.290064392654188
1.876381433921347, 6.215543926451017
1.877694981302570, 5.033902791567188
1.877231070388019, 4.708656877432918
2.994491285267025, 6.802513123676919
2.993912761097704, 6.397160501381253
2.993476139080479, 6.091233993936411
```

Sort

Sort by: Raw

Order: Ascending

Format

Number Formatting:

Digits: 5 Ignore

Column Separator: ,

Format

Copy to Clipboard Download .CSV Graph in Plotly* Close

*Plotly is a secure data analysis and graphing site with data sharing and access controls.

Visit http://plot.ly for details.

We used numbers (means and standard errors) from the computation for the subsequent effect size calculations. For its use, visit https://automeris.io/WebPlotDigitizer/. For its reliability and validity, for instance, see Drevon et al. (2016).

Drevon, D., Fursa, S. R., & Malcolm, A. L. (2016). Intercoder Reliability and Validity of WebPlotDigitizer in Extracting Graphed Data: *Http://Dx.Doi.Org/10.1177/0145445516673998*, *41*(2), 323–339. https://doi.org/10.1177/0145445516673998

Low cash vs medium cash (H1a)

```
esc_mean_se(grp1m = 5.03, grp1se = 0.25, grp1n = 88,
            grp2m = 6.4, grp2se =  0.25, grp2n = 88,
            es.type = "d")

##
## Effect Size Calculation for Meta Analysis
##
##      Conversion: mean and se to effect size d
##     Effect Size:  -0.5875
##  Standard Error:   0.1540
##        Variance:   0.0237
##        Lower CI:  -0.8893
##        Upper CI:  -0.2857
##          Weight:  42.1801
```

Low candy vs medium candy (H1b)

```
esc_mean_se(grp1m = 6.22, grp1se = 0.25, grp1n = 88,
            grp2m = 6.8, grp2se =  0.25, grp2n = 88,
            es.type = "d")

##
## Effect Size Calculation for Meta Analysis
##
##      Conversion: mean and se to effect size d
##     Effect Size:  -0.2487
##  Standard Error:   0.1513
##        Variance:   0.0229
##        Lower CI:  -0.5453
##        Upper CI:   0.0479
##          Weight:  43.6623
```

Control vs low cash (H1c)

```
esc_mean_se(grp1m = 6.29, grp1se = 0.125, grp1n = 88,
            grp2m = 5.03, grp2se =  0.25, grp2n = 88,
            es.type = "d")

##
## Effect Size Calculation for Meta Analysis
##
##      Conversion: mean and se to effect size d
##     Effect Size:   0.6835
##  Standard Error:   0.1551
##        Variance:   0.0241
##        Lower CI:   0.3795
##        Upper CI:   0.9875
##          Weight:  41.5724
```

Control vs low candy

```
esc_mean_se(grp1m = 6.29, grp1se = 0.125, grp1n = 88,
            grp2m = 6.22, grp2se =  0.25, grp2n = 88,
            es.type = "d")

##
## Effect Size Calculation for Meta Analysis
##
##      Conversion: mean and se to effect size d
##     Effect Size:   0.0380
##  Standard Error:   0.1508
##        Variance:   0.0227
##        Lower CI:  -0.2575
##        Upper CI:   0.3335
##          Weight:  43.9921
```

Monetized candy vs money

```
esc_mean_se(grp1m = 5.4, grp1se = 0.25, grp1n = 176,
            grp2m = 5.715, grp2se =  0.25, grp2n = 176,
            es.type = "d")

##
## Effect Size Calculation for Meta Analysis
##
##      Conversion: mean and se to effect size d
##     Effect Size:  -0.0952
##  Standard Error:   0.1067
##        Variance:   0.0114
##        Lower CI:  -0.3043
##        Upper CI:   0.1138
##          Weight:  87.9003
```

Monetized candy vs candy

```
esc_mean_se(grp1m = 5.4, grp1se = 0.25, grp1n = 176,
            grp2m = 6.51, grp2se =  0.25, grp2n = 176,
            es.type = "d")

##
## Effect Size Calculation for Meta Analysis
##
##      Conversion: mean and se to effect size d
##     Effect Size:  -0.3356
##  Standard Error:   0.1073
##        Variance:   0.0115
##        Lower CI:  -0.5460
##        Upper CI:  -0.1252
##          Weight:  86.7781
```

Low monetized candy vs medium monetized candy (H2)

```
esc_mean_se(grp1m = 4.71, grp1se = 0.25, grp1n = 88,
            grp2m = 6.09, grp2se =  0.25, grp2n = 88,
            es.type = "d")
```

```
##
## Effect Size Calculation for Meta Analysis
##
##      Conversion: mean and se to effect size d
##     Effect Size:  -0.5918
##  Standard Error:   0.1540
##        Variance:   0.0237
##        Lower CI:  -0.8937
##        Upper CI:  -0.2899
##          Weight:  42.1545
```

Control vs low monetized candy

```
esc_mean_se(grp1m = 6.29, grp1se = 0.125, grp1n = 88,
            grp2m = 4.71, grp2se =  0.25, grp2n = 88,
            es.type = "d")
```

```
##
## Effect Size Calculation for Meta Analysis
##
##      Conversion: mean and se to effect size d
##     Effect Size:   0.8571
##  Standard Error:   0.1575
##        Variance:   0.0248
##        Lower CI:   0.5483
##        Upper CI:   1.1658
##          Weight:  40.2996
```

Study 1

Low cash vs medium cash (H1a)

```
esc_mean_se(grp1m = 3.63, grp1se = 0.183, grp1n = 273,
            grp2m = 7.15, grp2se =  0.157, grp2n = 274,
            es.type = "d")
```

```
##
## Effect Size Calculation for Meta Analysis
##
##      Conversion: mean and se to effect size d
##     Effect Size:  -1.2510
##  Standard Error:   0.0935
##        Variance:   0.0087
##        Lower CI:  -1.4343
##        Upper CI:  -1.0678
##          Weight: 114.3743
```

Low candy vs medium candy (H1b)

```
esc_mean_se(grp1m = 5.14, grp1se = 0.165, grp1n = 275,
            grp2m = 6.31, grp2se =  0.160, grp2n = 276,
            es.type = "d")

## 
## Effect Size Calculation for Meta Analysis
## 
##      Conversion: mean and se to effect size d
##     Effect Size:  -0.4345
##  Standard Error:   0.0862
##        Variance:   0.0074
##        Lower CI:  -0.6035
##        Upper CI:  -0.2656
##          Weight: 134.5732
```

Control vs low cash (H1c)

```
esc_mean_se(grp1m = 6.15, grp1se = 0.154, grp1n = 275,
            grp2m = 3.63, grp2se =  0.183, grp2n = 273,
            es.type = "d")

## 
## Effect Size Calculation for Meta Analysis
## 
##      Conversion: mean and se to effect size d
##     Effect Size:   0.9024
##  Standard Error:   0.0897
##        Variance:   0.0080
##        Lower CI:   0.7266
##        Upper CI:   1.0781
##          Weight: 124.3422
```

Control vs low candy

```
esc_mean_se(grp1m = 6.15, grp1se = 0.154, grp1n = 275,
            grp2m = 5.14, grp2se =  0.165, grp2n = 275,
            es.type = "d")

## 
## Effect Size Calculation for Meta Analysis
## 
##      Conversion: mean and se to effect size d
##     Effect Size:   0.3823
##  Standard Error:   0.0861
##        Variance:   0.0074
##       Lower CI:   0.2137
##       Upper CI:   0.5510
##         Weight: 135.0328
```

Monetized candy vs money

```
esc_mean_se(grp1m = 5.3, grp1se = 0.13, grp1n = 551,
            grp2m = 5.39, grp2se =  0.142, grp2n = 547,
            es.type = "d")

## 
## Effect Size Calculation for Meta Analysis
## 
##      Conversion: mean and se to effect size d
##     Effect Size:  -0.0283
##  Standard Error:   0.0604
##        Variance:   0.0036
##        Lower CI:  -0.1466
##        Upper CI:   0.0901
##          Weight: 274.4690
```

Monetized candy vs candy

```
esc_mean_se(grp1m = 5.3, grp1se = 0.13, grp1n = 551,
            grp2m = 5.73, grp2se =  0.117, grp2n = 551,
            es.type = "d")

## 
## Effect Size Calculation for Meta Analysis
## 
##      Conversion: mean and se to effect size d
##     Effect Size:  -0.1483
##  Standard Error:   0.0603
##        Variance:   0.0036
##        Lower CI:  -0.2665
##        Upper CI:  -0.0300
##          Weight: 274.7451
```

Low monetized candy vs medium monetized candy (H2)

```
esc_mean_se(grp1m = 4.75, grp1se = 0.182, grp1n = 277,
            grp2m = 5.86, grp2se =  0.181, grp2n = 274,
            es.type = "d")

## 
## Effect Size Calculation for Meta Analysis
## 
##      Conversion: mean and se to effect size d
##     Effect Size:  -0.3691
##  Standard Error:   0.0859
##        Variance:   0.0074
##        Lower CI:  -0.5375
##        Upper CI:  -0.2007
##          Weight: 135.4394
```

Control vs low monetized candy

```
esc_mean_se(grp1m = 6.15, grp1se = 0.154, grp1n = 275,
            grp2m = 4.75, grp2se =  0.182, grp2n = 277,
            es.type = "d")

##
## Effect Size Calculation for Meta Analysis
##
##      Conversion: mean and se to effect size d
##     Effect Size:   0.5005
##  Standard Error:   0.0864
##        Variance:   0.0075
##        Lower CI:   0.3310
##        Upper CI:   0.6699
##          Weight: 133.8087
```

Low cash vs medium cash on joy (H3a)

```
esc_mean_se(grp1m = 5.92, grp1se = 0.176, grp1n = 273,
            grp2m = 7.00, grp2se =  0.149, grp2n = 274,
            es.type = "d")

##
## Effect Size Calculation for Meta Analysis
##
##      Conversion: mean and se to effect size d
##     Effect Size:  -0.4014
##  Standard Error:   0.0864
##        Variance:   0.0075
##        Lower CI:  -0.5706
##        Upper CI:  -0.2321
##          Weight: 134.0504
```

Low candy vs medium candy on joy (H3b)

```
esc_mean_se(grp1m = 6.942, grp1se = 0.162, grp1n = 275,
            grp2m = 7.438, grp2se =  0.142, grp2n = 276,
            es.type = "d")

##
## Effect Size Calculation for Meta Analysis
##
##      Conversion: mean and se to effect size d
##     Effect Size:  -0.1966
##  Standard Error:   0.0854
##        Variance:   0.0073
##        Lower CI:  -0.3640
##        Upper CI:  -0.0292
##          Weight: 137.0874
```

Candy vs Cash on joy (H4)

```
esc_mean_se(grp1m = 7.190, grp1se = 0.114, grp1n = 551,
            grp2m = 6.458, grp2se =  0.114, grp2n = 547,
            es.type = "d")

##
## Effect Size Calculation for Meta Analysis
##
##      Conversion: mean and se to effect size d
##     Effect Size:   0.2743
##  Standard Error:   0.0606
##        Variance:   0.0037
##        Lower CI:   0.1554
##        Upper CI:   0.3931
##          Weight: 271.9389
```

Candy vs monetized candy on joy (H4)

```
esc_mean_se(grp1m = 7.190, grp1se = 0.114, grp1n = 551,
            grp2m = 6.795, grp2se =  0.114, grp2n = 551,
            es.type = "d")

##
## Effect Size Calculation for Meta Analysis
##
##      Conversion: mean and se to effect size d
##     Effect Size:   0.1477
##  Standard Error:   0.0603
##        Variance:   0.0036
##        Lower CI:   0.0295
##        Upper CI:   0.2660
##          Weight: 274.7503
```

Study 2

Low cash vs medium cash (H1a)

```
esc_mean_se(grp1m = 5.35, grp1se = 0.081, grp1n = 999,
            grp2m = 7.978, grp2se =  0.069, grp2n = 999,
            es.type = "d", r = 0.644)
```

```
##
## Effect Size Calculation for Meta Analysis
##
##      Conversion: mean and se (within-subject) to effect size d
##     Effect Size:  -1.2955
##  Standard Error:   0.0492
##        Variance:   0.0024
##        Lower CI:  -1.3919
##        Upper CI:  -1.1990
##          Weight: 412.8827
```

Low candy vs medium candy (H1b)

```
esc_mean_se(grp1m = 5.630, grp1se = 0.076, grp1n = 999,
            grp2m = 6.766, grp2se =  0.071, grp2n = 999,
            es.type = "d", r = 0.843)

##
## Effect Size Calculation for Meta Analysis
##
##      Conversion: mean and se (within-subject) to effect size d
##     Effect Size:  -0.8672
##  Standard Error:   0.0468
##        Variance:   0.0022
##        Lower CI:  -0.9590
##        Upper CI:  -0.7755
##          Weight: 456.5772
```

Control vs low cash (H1c)

```
esc_mean_se(grp1m = 5.145, grp1se = 0.074, grp1n = 999,
            grp2m = 5.350, grp2se =  0.081, grp2n = 999,
            es.type = "d", r = 0.73)

##
## Effect Size Calculation for Meta Analysis
##
##      Conversion: mean and se (within-subject) to effect size d
##     Effect Size:  -0.1132
##  Standard Error:   0.0448
##        Variance:   0.0020
##        Lower CI:  -0.2010
##        Upper CI:  -0.0254
##          Weight: 498.7011
```

Control vs low candy

```
esc_mean_se(grp1m = 5.145, grp1se = 0.074, grp1n = 999,
            grp2m = 5.630, grp2se = 0.076, grp2n = 999,
            es.type = "d", r = 0.771)
```

```
##
## Effect Size Calculation for Meta Analysis
##
##      Conversion: mean and se (within-subject) to effect size d
##     Effect Size:  -0.3023
##  Standard Error:   0.0450
##        Variance:   0.0020
##        Lower CI:  -0.3905
##        Upper CI:  -0.2141
##          Weight: 493.8600
```

Monetized candy vs money

```
esc_mean_se(grp1m = 6.057, grp1se = 0.072, grp1n = 999,
            grp2m = 6.664, grp2se =  0.068, grp2n = 999,
            es.type = "d", r = 0.874)

##
## Effect Size Calculation for Meta Analysis
##
##      Conversion: mean and se (within-subject) to effect size d
##     Effect Size:  -0.5435
##  Standard Error:   0.0456
##        Variance:   0.0021
##        Lower CI:  -0.6328
##        Upper CI:  -0.4542
##          Weight: 481.7128
```

Monetized candy vs candy

```
esc_mean_se(grp1m = 6.057, grp1se = 0.072, grp1n = 999,
            grp2m = 6.198, grp2se =  0.071, grp2n = 999,
            es.type = "d", r = 0.915)

##
## Effect Size Calculation for Meta Analysis
##
##      Conversion: mean and se (within-subject) to effect size d
##     Effect Size:  -0.1513
##  Standard Error:   0.0448
##        Variance:   0.0020
##        Lower CI:  -0.2391
##        Upper CI:  -0.0635
##          Weight: 498.0745
```

Low monetized candy vs medium monetized candy (H2)

```
esc_mean_se(grp1m = 5.434, grp1se = 0.077, grp1n = 999,
            grp2m = 6.679, grp2se =  0.075, grp2n = 999,
            es.type = "d", r = 0.791)
```

```
## 
## Effect Size Calculation for Meta Analysis
## 
##      Conversion: mean and se (within-subject) to effect size d
##     Effect Size:  -0.8015
##  Standard Error:   0.0465
##        Variance:   0.0022
##        Lower CI:  -0.8926
##        Upper CI:  -0.7103
##          Weight: 462.3751
```

Control vs low monetized candy

```
esc_mean_se(grp1m = 5.145, grp1se = 0.074, grp1n = 999,
            grp2m = 5.434, grp2se =  0.077, grp2n = 999,
            es.type = "d", r = 0.776)

## 
## Effect Size Calculation for Meta Analysis
## 
##      Conversion: mean and se (within-subject) to effect size d
##     Effect Size:  -0.1807
##  Standard Error:   0.0448
##        Variance:   0.0020
##        Lower CI:  -0.2686
##        Upper CI:  -0.0929
##          Weight: 497.4685
```