Accentuation and compatibility: Replication and extensions of Shafir (1993) to rethink Choosing versus Rejecting paradigms

*Subramanya Prasad Chandrashekar
Institute of International Business and Governance,
Lee Shau Kee School of Business and Administration,
Open University of Hong Kong
spchandr@ouhk.edu.hk

*Jasmin Weber, *Sze Ying Chan, *Won Young Cho, *Tsz Ching Connie Chu
Department of Psychology, University of Hong Kong, Hong Kong SAR
u3553075@connect.hku.hk (s1047897@stud.sbg.ac.at); csy1029@connect.hku.hk;
wycho19@connect.hku.hk (chowonyoung052196@gmail.com); u3519356@connect.hku.hk
(conniechutc@gmail.com)

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

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

In press at Judgment and Decision Making

Accepted for publication on October 3, 2020

*Contributed equally, joint first authors
^Corresponding author
Word: abstract – 167; manuscript - 4747

Corresponding author

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

Author bios:

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

Subramanya Prasad Chandrashekar is a research assistant professor with the Lee Shau Kee School of Business and Administration at the Open University of Hong Kong. His research focuses on social status, lay-beliefs, and judgment and decision-making.

Jasmin Weber, Chan Sze Ying, Won Young Cho, and Chu Tsz Ching Connie were students at the University of Hong Kong during the academic year 2018-19.

Bo Ley Cheng was a teaching assistant at the University of Hong Kong psychology department during the academic year 2018-19.

Declaration of Conflict of Interest:

The author(s) declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

Financial disclosure/funding:

The research was supported by the European Association for Social Psychology corn seed grant awarded to the corresponding author.

Authorship declaration

Gilad led the reported replication effort with the team listed below. Gilad supervised each step of the project, conducted the pre-registration, and ran data collection. Prasad followed up on initial work by the other coauthors to verify and conduct additional analyses, and completed the manuscript draft. Prasad and Gilad jointly finalized the manuscript for submission.

Jasmin Weber, Chan Sze Ying, Won Young Cho, and Chu Tsz Ching (Connie) conducted the replication as part of a university course. They conducted an initial analysis of the paper, designed the replication, initiated the extensions, wrote the pre-registration, conducted initial data analysis, and wrote initial replication reports. Bo Ley Cheng guided and assisted the replication effort in the course.

Acknowledgments:

Subramanya Prasad Chandrashekar would like to thank the Institute of International Business and Governance (IIBG), established with the substantial support of a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (UGC/IDS 16/17), for its support.

In the table below, we 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) on details and definitions of each of the roles listed below.

Role	**Prasad**	**Jasmin Weber, Chan Sze Ying, Won Young Cho, and Chu Tsz Ching**	**Boley Cheng**	**Gilad Feldman**
Conceptualization		X		X
Pre-registration		X		X
Data curation				X
Formal analysis	X	X		
Funding acquisition				X
Investigation	X	X		X
Methodology		X		X
Pre-registration peer review / verification	X	X	X	X
Data analysis peer review / verification	X	X		
Project administration			X	X
Resources				X
Software	X	X		
Supervision			X	X
Validation	X			
Visualization	X			
Writing-original draft	X			X
Writing-review and editing	X			X

Abstract

We conducted a replication of Shafir (1993) who showed that people are inconsistent in their preferences when faced with choosing versus rejecting decision-making scenarios. The effect was demonstrated using an enrichment paradigm, asking subjects to choose between *enriched* and *impoverished* alternatives, with enriched alternatives having more positive and negative features than the impoverished alternative. Using eight different decision scenarios, Shafir found support for a compatibility principle: subjects chose and rejected enriched alternatives in choose and reject decision scenarios (d = 0.32 [0.23, 0.40]), respectively, and indicated greater preference for the enriched alternative in the choice task than in the rejection task (d = 0.38 [0.29, 0.46]). In a preregistered very close replication of the original study (N = 1026), we found no consistent support for the hypotheses across the eight problems: two had similar effects, two had opposite effects, and four showed no effects (overall d: -0.01 [-0.06, 0.03]). Seeking alternative explanations, we tested an extension, and found support for the accentuation hypothesis. Materials, data, and code, are available at https://osf.io/ve9bg/ .

Keywords: pre-registered replication; judgment and decision making; choosing versus rejecting; compatibility principle; replication crisis, accentuation hypothesis

Accentuation and compatibility: Replication and extensions of Shafir (1993) to rethink Choosing versus Rejecting paradigms

Early rational choice theories assumed the principle of invariance in human decision-making (von Neumann & Morgenstern, 1947). The invariance principle states that people's preference does not change when a decision task is described differently (*description invariance*) or when there are variations in the elicitation procedure (*procedural invariance*). Daniel Kahneman, Amos Tversky, and colleagues demonstrated that the assumptions of both description invariance and procedural invariance are often violated in human decision-making. For example, Tversky and Kahneman's (1986) findings on framing effects demonstrated that the invariance principle is violated when decision scenarios are described in a positive or negative frame. Similarly, variations in elicitation procedures were shown to cause preference reversals during the selection of job candidates (Tversky, Sattath, & Slovic, 1988) and in the prediction of others' academic performance (Slovic, Griffin, & Tversky, 1990).

Shafir (1993) was the first to employ the enrichment paradigm to further demonstrate the violations of procedural invariance. His study contrasted two decision-making scenarios that are intuitively equivalent: choosing versus rejecting. Subjects were randomly assigned to either choosing the preferred from two alternatives or rejecting the option not preferred among the two alternatives. The choice sets consisted of an enriched option, with both positive and negative features, and an impoverished alternative that had neutral features. Across eight scenarios, the original study found that the enriched alternative was selected more often in a choice task and was rejected more often in a rejection task. Shafir interpreted the results based on the compatibility principle that predicts that decision outcomes depend on the weighing of positive features during a choice task and negative features during a rejection task. That is, decisionmakers focus their attention on positive features during a choice task as they need positive reasons to justify the choice, whereas they direct their attention to negative features during the rejection task as they need reasons to reject an alternative. We summarized the scenarios used in the original article in Table 1 and the findings in Table 2 and Table 3.

Choice of target article: Shafir (1993)

Shafir's (1993) article has been highly influential with more than 640 citations, and has contributed to an active literature on the relational properties of choice sets. The compatibility principle has formed the theoretical basis for explaining people's decisions when deciding between products and job applicants (Park, Jun, & MacInnis, 2000; Sokolova & Krishna, 2016), and when choosing among products (Chernev, 2009; Nagpal & Krishnamurthy, 2008). Furthermore, the findings of the original article have formed the basis for subsequent theoretical work (e.g., Kahneman, 2003; Morewedge & Kahneman, 2010; Shafir, Simonson, & Tversky, 1993).

Recently, Many Labs 2 (Klein et al., 2018) conducted a partial replication of Problem 1 from the original study. The findings of this partial replication failed to provide support for the compatibility principle and the original findings. In response to the replication, Shafir (2018) noted several limitations in the replication effort. First, Klein and co-authors only attempted to replicate one out of eight decision scenarios reported in the original study. Second, the nature and value of the alternatives presented in the chosen decision-making scenario may have changed in meaning since the publication of the original study due to societal changes over time. Third, unlike the original study, the replication study did not counterbalance the order of presentation of the alternatives. In the current replication, conducted without being aware of the Many Labs effort, we addressed the noted methodological limitations of the earlier replication (Shafir, 2018), and went beyond the replication to add extensions to try and gain further insights about the phenomenon. More on that below.

We choose to conduct a replication of Shafir (1993) due to its impact (Coles, Tiokhin, Scheel, Isager, & Lakens, 2018; Isager, 2019), aiming for a comprehensive independent replication of all problems in the article. Replications are especially relevant following the recent growing recognition of the importance of reproducibility and replicability in psychological science (e.g., Brandt et al., 2014; Open Science Collaboration, 2015; van't Veer & Giner-Sorolla, 2016; Zwaan, Etz, Lucas, & Donnellan, 2018). A comprehensive replication of this target article is needed, given the ongoing discussion regarding the evaluation of replications and the active debate around the findings of the Many Labs 2 and other mass-replication efforts.

Our predictions in the replication followed that of Shafir (1993):

> *Hypothesis 1:* Subjects choose and reject the enriched alternative more often than the impoverished alternative across task frames (choice vs. rejection).
>
> *Hypothesis 2:* Subjects prefer the enriched alternative more often in the choice task frame than during the rejection task frame.

Table 1
Summary of scenarios in Shafir (1993) Experiments 1 to 8.

Problem	Scenario	Impoverished alternative	Enriched alternative
1	Which parent to award/deny the sole custody of the child	Parent A	Parent B
2	Which vacation spot to prefer/cancel	Spot A	Spot B
3	Which course to take immediately/postpone	Course X	Course Y
4	Which lottery to choose/give up	Lottery 1	Lottery 2
5	Which lottery to choose/give up	Lottery 1	Lottery 2
6	Which ice cream flavor to choose/give up	Flavor A	Flavor B
7	Which candidate to vote for/not to vote for	Candidate A	Candidate B
8	Which lottery to choose/Which lottery to reject first, and then reject later	Lottery 1 Lottery 2	Lottery 3

Table 2

Descriptive statistics: The percentages of subjects who Chose/Rejected across all problems in the original study and current replication.

Problem	Options	Original study				Replication (*N* = 1026)		
		N	**Choose-Group**	**Reject-Group**	**Choose + Reject**	**Choose-Group**	**Reject-Group**	**Choose + Reject**
1	Parent A (I)	170	36%	45%	81%	45.70%	52.50%	98.20%
	Parent B (E)		64%	55%	119%	54.30%	47.50%	101.80%
2	Spot A (I)	172	33%	52%	85%	55.60%	48.60%	104.20%
	Spot B (E)		67%	48%	115%	44.40%	51.40%	95.80%
3	Course X (I)	424	25%	65%	90%	45.50%	59.40%	104.90%
	Course Y (E)		75%	35%	110%	54.50%	40.60%	95.10%
4	Lottery 1 (I)	279	25%	50%	75%	18.30%	68.80%	87.10%
	Lottery 2 (E)		75%	50%	125%	81.70%	31.30%	113.00%
5	Lottery 1 (I)	278	23%	60%	83%	14.80%	66.60%	81.40%
	Lottery 2 (E)		77%	40%	117%	85.20%	33.40%	118.60%
6	Flavor A (I)	359	28%	55%	83%	43.60%	53.10%	96.70%
	Flavor B (E)		72%	45%	117%	56.40%	46.90%	103.30%
7	Candidate A (I)	398	79%	8%	87%	90.70%	29.10%	119.80%
	Candidate B (E)		21%	92%	113%	9.30%	70.90%	80.20%
8	Lottery 1 (I)	139	39%	44%	83%	10.90%	41.30%	52.20%
	Lottery 2 (I)					21.00%	36.90%	57.90%
	Lottery 3 (E)		61%	56%	117%	68.10%	21.80%	89.90%

Note. (I) = Impoverished option, (E) = Enriched option. In the replication subjects (*N* = 1026) completed all 8 problems.

Table 3

Summary of findings comparing the original study's and the replication's.

Comparison	**Problem**	**Shafir (1993)**	**Replication**		**Bayes Factor**	**Replication summary**
		Cohen's *d*	***z*-value, *p*-value**	**Cohen's *d***		
Hypothesis 1	1	0.39 [0.08, 0.69]	$z = 0.56$, $p = .287$	0.04 [-0.09, 0.16]	$BF_{10} = 0.13$; $BF_{01} = 7.7$	No signal – inconsistent
	2	0.32 [0.02, 0.62]	$z = -1.37$, $p = .915$	-0.09 [-0.21, 0.04]	$BF_{10} = 0.03$; $BF_{01} = 29.38$	No signal – inconsistent (opposite)
	3	0.21 [0.02, 0.40]	$z = -1.56$, $p = .941$	-0.10 [-0.22, 0.02]	$BF_{10} = 0.03$; $BF_{01} = 31.92$	No signal – inconsistent (opposite)
	4	0.51 [0.27, 0.75]	$z = 4.18$, $p < .001$	0.26 [0.14, 0.39]	$BF_{10} = 951.66$; $BF_{01} = 0.00$	Signal – inconsistent (weaker)
	5	0.34 [0.11, 0.58]	$z = 5.99$, $p < .001$	0.38 [0.26, 0.50]	$BF_{10} = 9.74\times10^{6}$; $BF_{01} = 0.00$	Signal – consistent
	6	0.34 [0.14, 0.62]	$z = 1.06$, $p = .144$	0.07 [-0.06, 0.19]	$BF_{10} = 0.23$; $BF_{01} = 4.29$	No signal – inconsistent
	7	0.23 [0.04, 0.43]	$z = -6.37$, $p = .999$	-0.41 [-0.53, -0.28]	$BF_{10} = 0.01$; $BF_{01} = 104.42$	No signal – inconsistent (opposite)
	8	0.34 [0.01, 0.68]	$z = -4.03$, $p = .999$	-0.21 [-0.31, -0.11]	$BF_{10} = 0.01$; $BF_{01} = 83.42$	No signal – inconsistent (opposite)
Hypothesis 2	1	0.39 [0.09, 0.69]	$z = 0.56$, $p = .288$	0.03 [-0.09, 0.16]	$BF_{10} = 0.21$; $BF_{01} = 4.85$	No signal – inconsistent
	2	0.31 [0.01, 0.61]	$z = -1.37$, $p = .915$	-0.09 [-0.21, 0.04]	$BF_{10} = 0.05$; $BF_{01} = 18.51$	No signal – inconsistent (opposite)
	3	0.22 [0.03, 0.41]	$z = -1.58$, $p = .944$	-0.10 [-0.22, 0.02]	$BF_{10} = 0.05$; $BF_{01} = 19.89$	No signal – inconsistent (opposite)
	4	0.53 [0.30, 0.77]	$z = 4.81$, $p < .001$	0.30 [0.18, 0.43]	$BF_{10} = 2.64\times10^{4}$; $BF_{01} = 0.00$	Signal – consistent
	5	0.37 [0.13, 0.61]	$z = 6.97$, $p < .001$	0.45 [0.32, 0.57]	$BF_{10} = 9.87\times10^{9}$; $BF_{01} = 0.00$	Signal – consistent
	6	0.50 [0.29, 0.71]	$z = 1.06$, $p = .144$	0.07 [-0.06, 0.19]	$BF_{10} = 0.38$; $BF_{01} = 2.65$	No signal – inconsistent
	7	0.38 [0.18, 0.57]	$z = -8.04$, $p = 1.00$	-0.52 [-0.64, -0.39]	$BF_{10} = 0.02$; $BF_{01} = 64.33$	Signal – inconsistent (opposite)
	8	0.34 [0.01, 0.68]	$z = -4.32$, $p = .999$	-0.22 [-0.32, -0.12]	BF10 = 0.02; $BF_{01} = 46.16$	Signal – inconsistent (opposite)

Note. $N = 1026$; Replication summary is using criteria by LeBel, Vanpaemel, Cheung, and Campbell (2019), see section "evaluation criteria for replication design and findings".

Extension: Accentuation hypothesis

There were other findings and theoretical accounts for the choosing versus rejecting paradigm. Ganzach (1995) reported results opposite to that of Shafir (1993) by showing that preference for the enriched alternative was greater in the rejection than in the choice condition. Wedell (1997) proposed a theoretical resolution of the inconsistent findings by Shafir (1993) and Ganzach (1995). Wedell's (1997) accentuation hypothesis stated that there is a greater need for justification in the choice condition than in the reject condition, and attribute differences are weighted more strongly when choosing due to a greater need for justification. If the enriched alternative was overall more attractive than the impoverished alternative, the positive differences were accentuated, and it was preferred more when choosing than when rejecting. If the enriched alternative was overall less attractive, negative differences were accentuated, and it was rejected more when choosing than when rejecting. This was noted by Shafir (2018) as one of the "more interesting possibilities" for the failure to replicate in Many Labs 2 (p. 495).

We extended the original design by examining the attractiveness of each of the alternatives in a choice set. Unlike binary choice, continuous scales for each option allow for higher sensitivity (how far apart are the differences in preferences between alternatives) and advanced analyses to examine the alternative theoretical explanation. Using these measures, we were able to run analyses to test the accentuation hypothesis.

Extension: Choice ability and preferences predictor

We also added an extension to examine the association between trait choice ability and preference and choosing versus rejecting decisions. Previous research on choice has argued that choice mindset, a psychological tendency, is associated with people ascribing agency to themselves and perceiving their own and others' actions through the lens of choice (Savani, Markus, Naidu, Kumar, & Berlia, 2010). People with a choice mindset view mundane actions such as checking emails and reading newspapers not as mere actions, but as choices. Thus, people with a choice mindset are prone to approach decisions with a clear choice framework. Building on the compatibility principle, we expected that individuals who rate themselves high on the ability to choose and indicate a high preference for choice would be more likely to prefer enriched alternatives, because they are more likely to take on the choosing strategy over the rejecting strategy in comparison to people who rate themselves lower on the ability to choose and indicate a low preference for choice.

Method

Pre-registration, power analysis, and open science

We preregistered the experiment on the Open Science Framework (OSF). Disclosures, power analyses, all materials, and additional details and analyses are available in the Supplementary Material. All measures, manipulations, and exclusions are reported, and data collection was completed before analyses. Pre-registration is available at: https://osf.io/r4aku. Data and R/RMarkdown code (R Core Team, 2015) is available at: https://osf.io/ve9bg/. We preregistered with the aim of detecting the smallest effect size (d = 0.21) observed in the original study at power of 0.95, which suggested a sample size of ~1092.

Subjects

A total of 1026 subjects were recruited online through American Amazon Mechanical Turk (MTurk) using the TurkPrime.com platform (Litman, Robinson, & Abberbock, 2017) (M_{age} = 39.39, SD_{age} = 12.47; 540 females). In the pre-registration stage, we noted we will focus on reporting full sample findings and will examine possible exclusion criteria factors such as self-reported seriousness, English proficiency, and failing attention checks. We found that exclusions had no impact on the findings.

Procedure

After consenting to take part in the study, subjects answered measures on their general attitudes towards choice (an extension). Subjects were then randomly assigned to one of two between-subject experimental conditions, either to choose (award or indicate a preference for) an option or to reject (deny or give up) an option. Each of the two experimental conditions consisted of eight decision problems (summarized in Table 1). Seven of the eight problems presented to all the subjects included a choice between two alternatives (binary; Problems 1-7) and one problem consisted of three alternatives (non-binary; Problem 8). Problems with binary alternatives had one option with both more positive and negative aspects (enriched alternative) and one with fewer positive and negative features (impoverished alternative). The problem with non-binary alternatives included one enriched alternative and two impoverished alternatives. In regards to the non-binary problem (Problem 8), half the subjects were asked to choose a lottery that they most preferred among three alternatives, and another half in a two-step decision rejected lotteries that they least preferred, rejecting one at a time. All descriptions and questions were taken from the original article (Shafir, 1993). A comparison of the original study's sample and the replication sample is provided in Table 4 (see Table S1, in which we note the reasons for the chosen differences between original studies and the replication attempt).

Table 4

Comparison between original and the replication study.

	Original study	**Replication**
Number of problems	8 problems that included 7 binary-problems and 1 non-binary problem	8 problems that included 7 binary-problems and 1 non-binary problem
Design	Between-subjects: Design followed two between-subjects conditions for each of the binary and non-binary problems	Between-subjects: Design followed two between-subjects conditions for each of the binary and non-binary problems
Procedure	Conducted in a lab using paper and pencil. Subset of 2-3 problems out of the set, separated by filler items.	Conducted online using Qualtrics. All 8 problems, with no filler items.
Sample size	Ranged between 139 to 424 per problem across 8 problems	1026
Sample population	Undergraduates a university in USA.	Subjects from Amazon Mechanical Turk (MTurk).
Remuneration	Monetary reward	Monetary reward

Measures

Trait choice (ability and preference). Two items measured the subjects' perceived ability to choose: "It's very hard for me to choose between many alternatives." (reversed) and "When faced with an important decision, I prefer that someone else chooses for me." (reversed) (α = .63). Similarly, subjects rated their preference toward choice in two items: "The more choices I have in life, the better" and "In each decision I face, I prefer to have as many alternatives as possible to choose from." (α = .81) On all four items, the scale was from 1 (*Strongly Disagree*) to 7 (*Strongly Agree*). The two scales were adapted from Feldman, Baumeister, and Wong (2014).

Attractiveness. For each of the eight problems, after choosing or rejecting the alternative(s), subjects proceeded to the next page and rated the relative attractiveness of the enriched and impoverished alternatives. As the term "attractive" might be associated with choosing, this may lead to biases in the ratings, thus subjects were asked to rate each alternative with the terms "bad" and "good" to maintain neutrality. The scale for the items ranged from 0 (*Very bad)* to 5 (*Very good*).

Data analysis plan

We employed one-proportion and two-proportion *z*-tests to investigate Hypothesis 1 and Hypothesis 2, respectively. Given the clear directionality of the predictions in the original study, both tests were one-tailed. We then used the obtained *z*-value to calculate the Cohen's *d* effects and 95% confidence intervals. All analyses were performed using R programming environment (R Core Team, 2015). Furthermore, we complemented Null Hypothesis Significance Testing (NHST) analyses with Bayesian analyses to quantify support for the null hypothesis when relevant (Kruschke & Liddell, 2018; Vandekerckhove, Rouder, & Kruschke, 2018) using 'BayesFactor' R package (Version 0.9.12-4.2; Morey & Rouder, 2015) and 'abtest' R package (Version 0.1.3.; based on a model by Kass & Vaidyanathan, 1992). The Bayesian analyses were added after preregistering the data analysis plan.

Evaluation criteria for replication design and findings

Table 5 provides a classification of this replication using the criteria by LeBel, McCarthy, Earp, Elson, and Vanpaemel (2018) (also see Figure S1). We summarized the current replication as a "very close replication". To interpret the replication results, we followed the framework by LeBel et al. (2019). They suggested a replication evaluation using three factors: (a) whether a signal was detected (i.e., confidence interval for the replication effect size (ES) excludes zero), (b) consistency of the replication ES with the original study's ES, and (c) precision of the replications ES estimate (see Figure S2).

Table 5

Classification of the two replication studies based on LeBel et al. 's (2018) taxonomy.

Design facet	**Replication study**
IV operationalization	Same
DV operationalization	Same
IV stimuli	Same
DV stimuli	Same
Procedural details	Similar (minor adjustments)
Physical settings	Different
Contextual variables	Different
Replication classification	Very close replication

Note. Information on this classification is provided in LeBel et al. 2018. See also figure provided in the Supplementary Material.

Results

The proportions of subjects choosing or rejecting the enriched or impoverished alternative in each of the eight problems are detailed in Table 2. The findings of the statistical tests and effect-size estimates are summarized in Table 3 (also see Figures 1 and 2).

The enriched alternatives share exceeded 100% for Problems 1, 4, 5 and 6 (Table 2) across both the choosing and rejecting frames. The results of one-proportion z-test investigating Hypothesis 1 indicated support in Problem 4 (z = 4.18, p < .001, d = 0.26, 95% CI [0.14, 0.39]) and Problem 5 (z = 5.99, p < .001, d = 0.38, 95% CI [0.26, 0.50]). The results were in the opposite direction for Problem 7 (z = – 6.37, p =1.00, d = – 0.41, 95% CI [– 0.53, – 0.28]) and Problem 8 (z = – 4.03, p =.999, d = – 0.21, 95% CI [– 0.31, – 0.11]). The results of Problem 1, 2, 3, and 6 failed to provide empirical support for the compatibility hypothesis (effect sizes ranged from -0.10 95% CI [-0.22, 0.02] to 0.07 95% CI [-0.06, 0.19]). Unlike the original study, these findings do not indicate consistent evidence in support of the Hypothesis 1 prediction that the enriched alternative is selected and rejected more often.

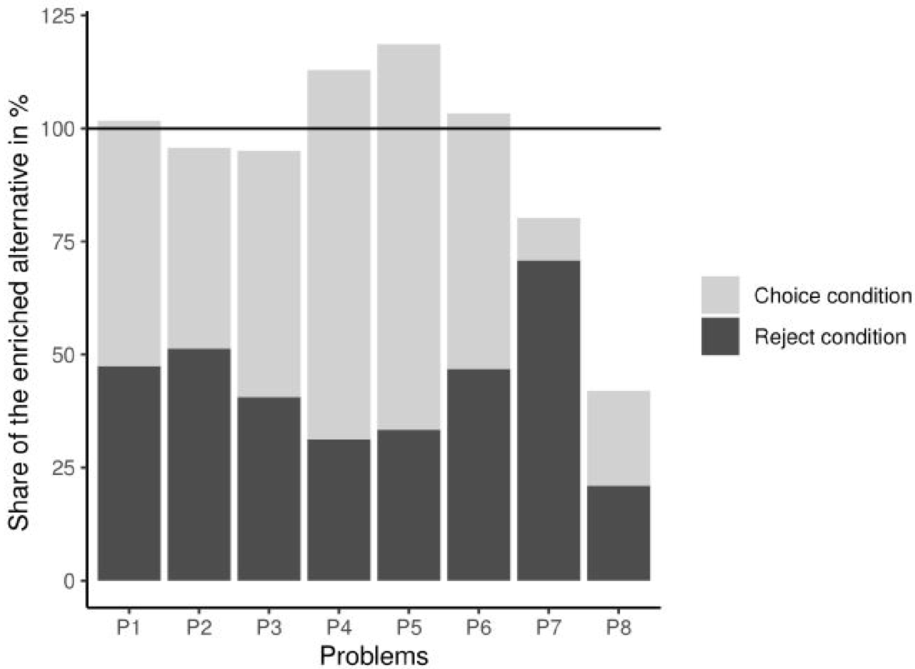

Figure 1. Share of the enriched alternative in % across 'choose' and 'reject' experimental conditions.

We complemented the NHST analyses used in the original article with Bayesian analysis to allow for quantifying the evidence in support of the null hypothesis (see Table 3). We conducted one-sided Bayesian tests of single proportions with a prior *r* scale set at 0.5 (defined as "medium" and considered the more conservative option). The result revealed that Bayes factor (BF) for Problems 1, 2, 3, 6, 7, and 8 was in stronger support of the null: Problem 1: $BF_{10} = 0.13$, $BF_{01}=7.7$; Problem 2: $BF_{10} = 0.03$, $BF_{01} = 29.38$; Problem 3: $BF_{10} = 0.03$, $BF_{01} = 31.92$; Problem 6: $BF_{10} = 0.23$, $BF_{01} = 4.29$; Problem 7: $BF_{10} = 0.01$, $BF_{01} = 104.42$; Problem 8: $BF_{10} = 0.01$, $BF_{01} = 83.42$. For example, Bayes factor (BF_{01}) of 7.7 in Problem 1 suggests that the data were 7 times more likely to be observed under the null hypothesis than the alternative.

To test Hypothesis 2, we conducted a Two-Proportions *z*-test. We then calculated the effect size, Cohen's *d*, with a 95% confidence interval (Table 3). The results of Problem 4 (z = 4.81, p < .001, d = 0.30, 95% CI [0.18, 0.43]) and Problem 5 (z = 6.97, p < .001, d = 0.45, 95% CI [0.32, 0.57]) supported predictions of the original article that more subjects chose the enriched alternative when asked to choose than when asked to reject. However, more subjects chose the enriched alternative when asked to reject than to choose in Problem 7 (z = -8.04, p =1.00, d = -0.52, 95% CI [-0.64, -0.39]) and Problem 8 (z = -4.32, p =.999, d = -0.22, 95% CI [-0.32, -0.12]), which contradicted the findings in the original article. We found no support for differences between the proportions of subjects choosing the enriched alternative in the choosing and rejecting decision frame in Problem 1 (z = 0.56, p = .288, d = 0.03, 95% CI [-0.09, 0.16]), Problem 2 (z = 1.37, p = .915, d = -0.09, 95% CI [-0.21, 0.04]), Problem 3 (z = 1.58, p =.943, d = -0.10, 95% CI [-0.22, 0.02]) and Problem 6 (z = 1.06, p = .144, d = 0.07, 95% CI [-0.06, 0.19]).

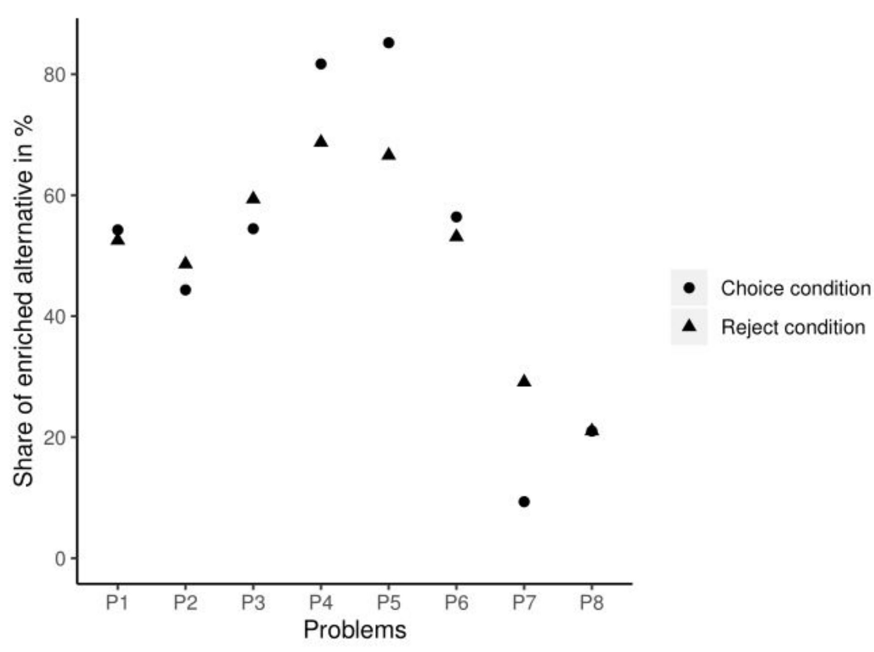

Figure 2. Share of the enriched alternative in % between 'choose' and 'reject' experimental conditions.

Furthermore, we conducted Bayesian A/B testing that mirrors the two-proportion *z*-test based on a model by Kass and Vaidyanathan (1992) using the 'abtest' R package. Mirroring Hypothesis 1, the results for Hypothesis 2 revealed that Bayes factor (BF) for Problems 1, 2, 3, 6, 7, and 8 in are more in favor of the null: Problem 1: $BF_{10} = 0.21$, $BF_{01} = 4.85$; Problem 2: $BF_{10} = 0.05$, $BF_{01} = 18.51$; Problem 3: $BF_{10} = 0.05$, $BF_{01} = 19.89$; Problem 6: $BF_{10} = 0.38$, $BF_{01} = 2.65$; Problem 7: $BF_{10} = 0.02$, $BF_{01} = 64.33$; Problem 8: $BF_{10} = 0.02$, $BF_{01} = 46.16$.

Comparison of the results with the original findings by Shafir (1993)

The evaluation of the replication results by the pairwise comparisons of each of the eight decision scenarios using LeBel et al. (2019) are summarized in Table 3. The findings of the present replication are mostly inconsistent with the results of Shafir's original study. Only two of the eight problems (Problem 4 and Problem 5) are supportive of the compatibility hypothesis. Moreover, two other problems (Problem 7 and Problem 8) showed an effect in the opposite direction. Taken together, the replication findings do not indicate consistent support for the original findings.

General Summary: Mini meta-analysis

The variations in the findings reported across the eight different decision scenarios make it hard to succinctly summarize the overall effect size of the predictions based on the compatibility hypothesis. Therefore, we conducted a mini meta-analysis of the effect sizes observed across eight decision scenarios for each of the predictions (Goh, Hall, & Rosenthal, 2016; Lakens & Etz, 2017). We ran both a within-subject aggregation and a fixed-effects model analysis method using the '*metafor*' package in R (Viechtbauer, 2010; see Figure S3-S6 in the Supplementary Material), and results were near identical.

The mini-meta analysis findings were: Hypothesis 1 *d* = -0.01 [-0.06, 0.03], and Hypothesis 2 *d* =-0.01 [-0.06, 0.03]. The results of the mini-meta analysis are summarized in Table 6.

Table 6

Summary of findings of the original study versus replication, based on mini-meta analysis.

Predictions	**Cohen's *d***		**Replication summary**
	Shafir (1993)	**Replication**	
Hypothesis 1	0.32 [0.23, 0.40]	-0.01 [-0.06, 0.03]	No signal – inconsistent
Hypothesis 2	0.38 [0.29, 0.46]	-0.01 [-0.06, 0.03]	No signal – inconsistent

Extension: Attractiveness ratings

We tested additional variables recorded on a continuous scale that measured the attractiveness of the alternatives. The responses to these additional variables included the attractiveness of the alternatives on a 6-point continuous scale (ranged from 0 to 5). We provide detailed results of the analysis in the Supplementary Materials (see Table S3-S5).

We conducted two sets of independent *t*-tests. First, we compared the attractiveness of the enriched alternative between the choice and reject experimental conditions. Second, we contrasted the relative attractiveness of the enriched alternatives between the choice and reject experimental conditions. The calculation of the relative attractiveness of the enriched alternative involved subtracting the attractiveness score of the enriched alternative from the attractiveness score of the impoverished alternative within each experimental condition. Then we contrasted the relative attractiveness of the enriched alternative between choice and reject experimental conditions. As Problem 8 included a non-binary alternative, we averaged the attractiveness scores of the impoverished alternatives before calculating the relative attractiveness of the enriched alternative. Furthermore, we conducted Bayesian analysis for both the planned contrasts with a prior value set at 0.707 (reflecting expectations for an effect, as it was expected from the original study).

The effect size estimates for the prediction that ratings of the attractiveness of enriched alternative between conditions ranged from 0.00 [-0.12, 0.13] to 0.09 [-0.03, 0.21]. Furthermore, effect size estimates of relative attractiveness of the enriched alternative across conditions ranged from 0.01 [-0.11, 0.14] to 0.14 [0.02, 0.26]. The Bayesian analysis mirrors these effect sizes and indicates support for the null in all the problems except Problem 8.

Extension: Individual-level predictors

We tested the prediction that individuals who rate themselves higher on ability to choose and indicated higher preference for choice are more likely to prefer the enriched alternative. We conducted two separate binary logistic mixed-effects regression analyses which included the experimental condition and individual-level variables as the fixed effect predictors of choosing the enriched alternative (*Yes* = 1; *No* = 0). The regression included subject ID as a random factor on the intercept.

We found no evidence for an association between individual-level predictors' ability to choose (Wald χ^2 (1) = 0.90, p = .343) or preference for choice (Wald χ^2 (1) =0.41, p =.522) and the likelihood of preferring the enriched alternative (see Table S6-S9 for detailed results).

Extension: Testing the accentuation hypothesis

The inconsistent results regarding compatibility hypothesis may have been due to the variation of the overall attractiveness of the enriched alternative relative to the impoverished alternative across the eight problems. The accentuation hypothesis (Wedell, 1997) proposed that if the overall relative attractiveness of the enriched alternative is greater than that of the impoverished alternative in a choice set, the positive attributes are more accentuated in the choice condition compared to the reject condition, because of a greater need for justification in the choice condition. Therefore, people more often prefer the enriched alternative in the choice condition than in the reject condition. In contrast, when the overall relative attractiveness of enriched alternative is lower than that of the impoverished alternative, the negative attributes are more accentuated in the choice condition, again due to greater need for justification. Therefore, in this scenario, people prefer the impoverished alternative in the choice condition more often than in the reject condition.

To test the accentuation hypothesis, we conducted binary logistic mixed-effects regression analysis. In this analysis, we included responses from Problem 1 to 7, as these problems shared the common procedure of choosing between two alternatives (binary choice set). We followed Wedell's (1997) approach to calculate the overall proportion of subjects (across experimental conditions) preferring the enriched alternative for each of the seven problems, as a measure of the overall relative attractiveness of the enriched alternative. We conducted a binary logistic mixed-effects regression analysis in which the experimental condition, the overall proportion preferring the enriched alternative, and the interaction term (overall proportions × experimental condition) were the fixed effects predictors of choosing the enriched alternative (*Yes* = 1; *No* = 0). The regression included subject ID as a random factor on the intercept.

The results of the regression found the main effect of the overall proportion preferring the enriched alternative as Wald χ^2 (1) = 657.28, p < .001), and the interaction effect Wald χ^2 (1) = 127.70, p < .001 (also see Table 7). As can be seen in Figure 3, the proportions preferring the enriched alternative for the choice and reject experimental conditions as a function of the overall proportion preferring the enriched alternative indicate alternate paths. Across 7 problems, the overall proportion preferring the enriched alternative ranged from 19% to 76%, and the results are consistent with the accentuation hypothesis.

Table 7

Results of binary logistic mixed-effects regression following Wedell's (1997) procedure

	Dependent variable: Predicted probability of enriched alternative	
	Main effect	Interaction
Constant	-2.34*** (0.100)	-3.59*** (0.163)
Overall proportion preferring enriched (PEN)	4.69*** (0.168)	6.97*** (0.286)
Experimental condition (EXP) (1 = Choose; 0 = Reject)	-0.05 (0.058)	2.12*** (0.201)
PEN × EXP		−3.95*** (0.350)
Observations	7,182	7,182
Log Likelihood	−4,455.54	−4,387.08
Akaike Inf. Crit.	8,919.07	8,784.15
Bayesian Inf. Crit.	8,946.59	8,818.55

Note: **p* < 0.1; ***p* < 0.05; ***p* < 0.01; ****p* < 0.001

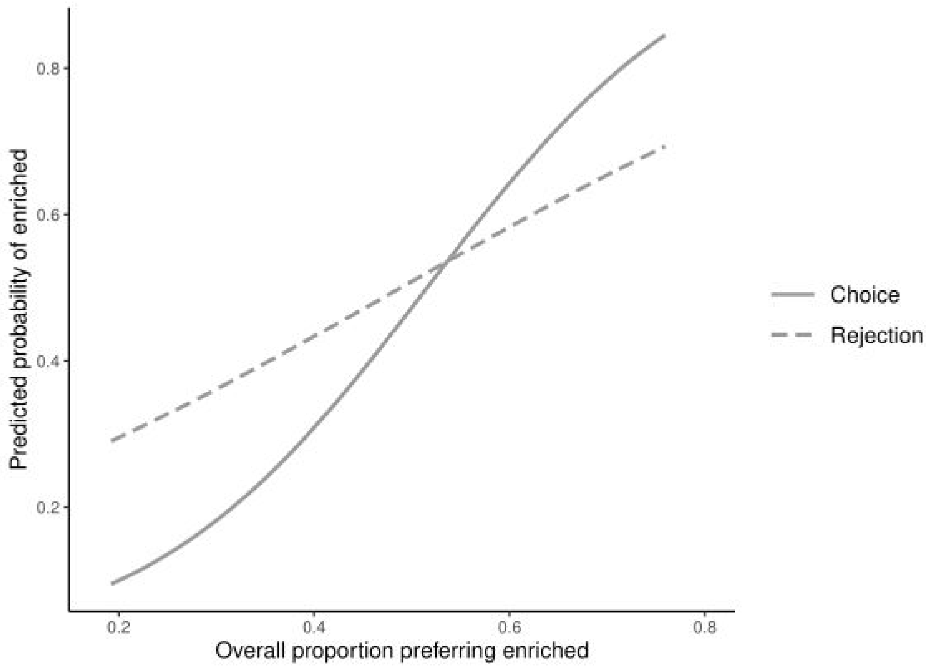

Figure 3. Predicted probability of the enriched alternative in choice and rejection tasks as a function of overall preference for the enriched alternative. Fitted lines are the marginal effects of interaction terms.

Furthermore, we tested for the accentuation hypothesis using the attractiveness measures. For each subject we calculated the relative attractiveness of the enriched alternative by subtracting the attractiveness score of the enriched alternative from the attractiveness score of the impoverished alternative across the seven binary problems. This analysis allowed us to test the accentuation hypothesis in a fine-grained manner by taking into account the relative attractiveness measure at the participant level for each of the seven decision problems.

We then conducted a binary logistic mixed-effects regression analysis in which the experimental condition, the relative attractiveness of the enriched alternative, and the interaction term (relative attractiveness of the enriched alternative × experimental condition) were the fixed effects predictors of choosing the enriched alternative (*Yes* =1 ; *No* = 0). The regression included subject ID as a random factor on the intercept.

The results of the regression showed a main effect of the relative attractiveness of enriched alternative (Wald χ^2 (1) = 980.04, p < .001) and the interaction term (Wald χ^2 (1) = 134.08, p < .001). As can be seen in Figure 4 (also see Table 8), the proportions preferring the enriched alternative for choice and reject experimental conditions as a function of the relative attractiveness of the enriched alternative indicate alternating paths. In summary, the results are consistent with the accentuation hypothesis.

Furthermore, we conducted additional analysis to check the robustness of the results by accounting for the sampling variability of the stimuli (Judd, Westfall, & Kenny, 2012). We conducted the same two sets of mixed-effect regression analyses with additional random intercepts and random condition slopes for stimuli along with other predictors. The results of the additional analysis remain the same (see Table S10-S11).

Table 8

Results of binary logistic mixed-effects regression.

	Dependent variable: Predicted probability of enriched alternative	
	Main effect	Interaction
Constant	0.35*** (0.039)	0.41*** (0.043)
Relative attractiveness of enriched alternative (AEO)	0.71*** (0.021)	1.01*** (0.037)
Experimental condition (EXP) (1 = Choose; 0 = Reject)	-0.12** (0.055)	-0.20*** (0.057)
AEO × XP		−0.51*** (0.044)
Observations	7,182	7,182
Log Likelihood	−4,115.19	−4,043.28
Akaike Inf. Crit.	8,238.37	8,096.55
Bayesian Inf. Crit.	8,265.89	8,130.95

Note: *p < 0.1; **p < 0.05; **p < 0.01. The relative attractiveness variable used in the regression was calculated based on the responses to extension variables.

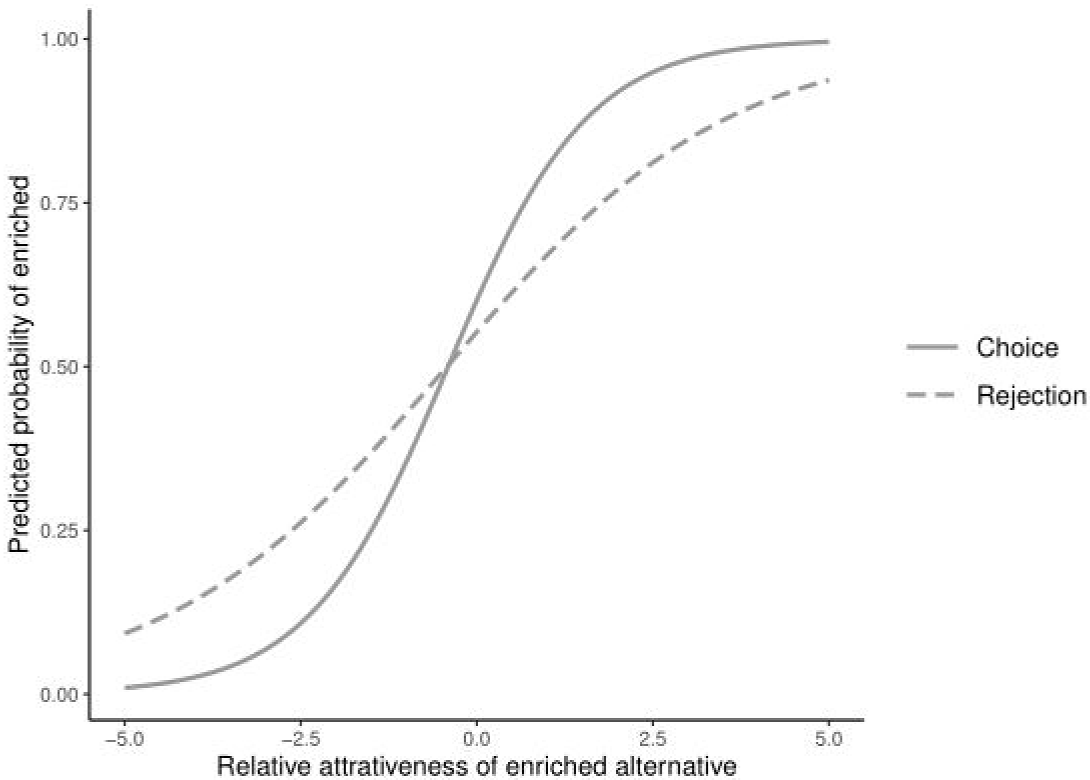

Figure 4. Predicted probability of the enriched alternative in choice and rejection tasks as a function of overall preference for the enriched alternative. Fitted lines are the marginal effects of interaction terms. The relative attractiveness variable used in the regression was calculated based on the responses to extension variables.

Discussion

We conducted a replication of the eight choosing versus rejection problems in Shafir (1993). We successfully replicated the results of Problem 4 and Problem 5 of the original study. However, in Problems 7 and 8 we found effects in the direction opposite to the original findings and our findings for Problems 1, 2, 3, and 6 indicated support for the null hypothesis. Taken together, we failed to find consistent support for the compatibility hypothesis noted in Shafir (1993). Additionally, we conducted supplementary analyses and found support for the accentuation hypothesis.

Replications: Adjustments, implications, and future directions

We aimed for a very close direct replication of the original study, with minimum adjustments, addressing many of the concerns raised over the replication by Many Labs 2, yet our replication still differed from the original studies in several ways. The stimuli used in the original article was targeted at and tested with American undergraduates in the context of the 1990s. We ran the same materials, with no adjustments to the stimuli, online, and with a more diverse population. We also made adjustments to the procedures, by presenting our participants with all eight questions, instead of only two or three in the original study, and with no filler items (see Table 4). Replications are never perfectly exact, and given these changes, it is possible that these factors may have somehow affected the results. However, our findings were not random, but rather demonstrated a pattern of results that replicated findings from a different article and supporting an alternative account, and we believe it is highly unlikely that such a change could be explained by any of the adjustments we made.

We also note limitations that suggest promising directions for future research. First, the extension analysis that use of attractiveness rating in testing compatibility hypothesis is not theoretically precise in making the compatibility hypothesis prediction, that is, that relative attractiveness ratings of enriched will be higher choice condition than for reject condition. We would also like to see further work testing similar extensions testing nuanced preferences (rather than binary choice) yet with more explicit direct integration with the choose/reject framing. Second, it is possible that the inconsistent findings regarding the compatibility hypothesis may be due to deviations of auxiliary theories embedded in the compatibility hypothesis (Meehl, 1990). For example, the compatibility hypothesis spells out the substantive argument that people seek positive reasons to justify choosing an alternative, whereas negative reasons to justify rejecting an alternative. However, auxiliary theories that specify the degree to which justification is a component of the compatibility hypothesis are were not well specified and are not clear, as unfortunately is standard in our field. We call for researchers to specify more precise indicators of the boundary conditions of theory testing, so that if some of contextual factors change, we would be able to directly test and analyze how these affect our findings, rather than engage in posthoc theorizing. Thus our findings may be due to changes in the conjunction of several premises assumed around the compatibility hypothesis's substantive theory, yet we need stronger well defined theories and hypotheses, and continuous testing over time, to be able to truly assess if and to what extend any of these factors are indeed relevant to the theory, and to the empirical test of that theory.

The current study contributes to the theory development by qualifying the theoretical assertions of the compatibility hypothesis. We addressed the methodological issues raised by Shafir (2018) in his commentary on the Many Labs 2. Given our findings we believe that most explanations noted in the commentary are unlikely reasons for the failure to replicate reported in

Many Labs 2 or our failure to find consistent support for the original findings and the compatibility hypothesis. Theoretical accounts need well-defined criteria that would allow for falsification of these accounts, and our replications helps advance theory by testing theoretical assertions of the compatibility hypothesis (Popper, 2002). By improving on the design of Many Labs 2, and by conducting extensions that showed support for plausible alternative accounts, our replication contributes to theory specification and supports further theory development (Glöckner & Betsch, 2011). Researchers conducting research in this domain and future research on this phenomenon can build on insights gained here to advance theory by defining the boundary conditions under which it operates and explore further ways on how it should be tested. Our replication does not rule out the compatibility account, only indicates that it is in need of further elaboration and specification, and further testing, and we see much promise in examining the interaction of the two accounts.

We tested the competing theoretical assertion by Wedell (1997). Our results in support of this account suggest that the stimuli from the 1990s is still of relevance, atleast for testing that account. It is still possible, that other stimuli developed using the choosing versus rejecting paradigm may show support for the compatibility hypothesis reported by Shafir (1993). Yet, given the Many Labs 2 and our findings we recommend that other compatibility hypothesis stimuli be revisited with direct close replications or that new stimuli be developed before further expanding on the compatibility hypothesis. For this phenomenon, and the judgement and decision-making literature overall, we see great value in conducting well-powered, preregistered direct replications, preferably in Registered Reports or blinded outcomes peer review format. Our findings suggest that future work on choosing versus rejecting may benefit from paying closer attention to the accentuation hypothesis (Wedell, 1997).

Importance of direct replications

This replication case study highlights the importance of conducting comprehensive direct replications. Many Labs 2 was one of the largest replication efforts to date, yet such mass collaboration replication efforts cannot and should not be taken as a replacement for singular comprehensive direct replications. These large replication projects are valuable in targeting specific research questions about the overall replicability of a research domain, investigating factors such as heterogeneity and high-level moderators such as culture or setting. Furthermore, large replication projects tend to summarize complex replications in simplified conclusions that fail to capture the complexity inherent in the original articles or the richness of their and the replication's findings. Therefore, we believe that large scale replication projects should be complemented by singular direct replication and extension studies such as the one we conducted here. Combined, they can help better understand the phenomenon of interest and inform future research.

References

Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., ... & Van't Veer, A. (2014). The replication recipe: What makes for a convincing replication?. *Journal of Experimental Social Psychology*, *50*, 217-224. DOI: 10.1016/j.jesp.2013.10.005

Chen, J., Hui, L., Yu, T., Feldman, G., ... (2020). Foregone opportunities and choosing not to act: Replications of Inaction Inertia effect. *Social Psychological and Personality Science*. Manuscript accepted for publication.

Chernev, A. (2009). Choosing versus Rejecting: The Impact of Goal-Task Compatibility on Decision Confidence. *Social Cognition, 27*(2), 249-260. DOI: https://doi.org/10.1521/soco.2009.27.2.249

Coles, N. A., Tiokhin, L., Scheel, A. M., Isager, P. M., & Lakens, D. (2018). The costs and benefits of replication studies. *Behavioral and Brain Sciences,* 41, e124. DOI: https://doi.org/10.1017/S0140525X18000596

Feldman, G., Baumeister, R. F., & Wong, K. F. E. (2014). Free will is about choosing: The link between choice and the belief in free will. Journal of Experimental Social Psychology, 55, 239-245.

Ganzach, Y. (1995). Attribute scatter and decision outcome: Judgment versus choice. *Organizational Behavior & Human Decision Processes, 62,* 113-122. DOI: 10.1006/obhd.1995.1036

Glöckner, A., & Betsch, T. (2011). The empirical content of theories in judgment and decision making: Shortcomings and remedies. *Judgment & Decision Making*, *6*(8), 711-721. DOI:

Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some arguments on why and a primer on how. *Social and Personality Psychology Compass*, *10*, 535-549. DOI: 10.1111/spc3.12267

Isager, P. M. (2019). Quantifying Replication Value: A formula-based approach to study selection in replication research. *Open Science 2019, Trier, Germany*. DOI: 10.23668/psycharchives.2392. See "Quantifying the corroboration of a finding" preprint retrieved from: https://pedermisager.netlify.com/post/quantifying-the-corroboration-of-afinding/

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of personality and social psychology*, *103*(1), 54. DOI: 10.1037/a0028347

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American psychologist*, *58*(9), 697. DOI: 10.1037/0003-066X.58.9.697

Kass, R. E., & Vaidyanathan, S. K. (1992). Approximate Bayes factors and orthogonal parameters, with application to testing equality of two binomial proportions. *Journal of the Royal Statistical Society: Series B (Methodological)*, *54*(1), 129-144. DOI: 10.1111/j.2517-6161.1992.tb01868.x

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Jr., Alper, S., . . . Nosek, B. A. (2018). Many Labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science, 1*(4), 443–490. DOI: 10.1177/2515245918810225

Kutscher, L., & Feldman, G. (2019). The impact of past behaviour normality on regret: replication and extension of three experiments of the exceptionality effect. *Cognition and Emotion*, *33*(5), 901-914. doi: 10.1080/02699931.2018.1504747

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, *25*, 178-206. DOI: 10.3758/s13423-016-1221-4

Lakens, D., & Etz, A. J. (2017). Too True to be Bad: When Sets of Studies with Significant and Nonsignificant Findings Are Probably True. *Social Psychological and Personality Science*, *8*, 875–881. doi:10.1177/1948550617693058

LeBel, E. 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. DOI: 10.1177/2515245918787489

LeBel, E. P., Vanpaemel, W., Cheung, I., & Campbell, L. (2019). A Brief Guide to Evaluate Replications. *Meta Psychology, 541*, 1–17. DOI: 10.31219/osf.io/paxyn

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime. com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*, 433-442. DOI: 10.3758/s13428-016-0727-z

Morewedge, C. K., & Kahneman, D. (2010). Associative processes in intuitive judgment. *Trends in cognitive sciences*, *14*(10), 435-440. DOI: 10.1016/j.tics.2010.07.004

Morey, R. D., & Rouder, J. N. (2015). BayesFactor: Computation of Bayes factors for common designs (R Package Version 0.9.12-2). Retrieved from https://CRAN.R-project.org/package=BayesFactor

Nagpal, A., & Krishnamurthy, P. (2008). Attribute Conflict in Consumer Decision Making: The Role of Task Compatibility. *Journal of Consumer Research, 34,* 696–705. http://doi.org/10.1086/521903

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, *349*, aac4716–aac4716. DOI: 10.1126/science.aac4716

Park, C. W., Jun, S. Y., & MacInnis, D. J. (2000). Choosing what I want versus rejecting what I do not want: An application of decision framing to product option choice decisions. *Journal of Marketing Research*, *37*(2), 187-202. DOI: 10.1509/jmkr.37.2.187.18731

Popper, K . (2002). The logic of scientific discovery (2nd ed.). London, England: Routledge.

R Core Team (2015) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07- 0, URL http://www.R-project.org

Savani, K., Markus, H. R., Naidu, N. V. R., Kumar, S., & Berlia, N. (2010). What counts as a choice? US Americans are more likely than Indians to construe actions as choices. *Psychological Science*, *21*(3), 391-398. DOI: 10.1177/0956797609359908

Shafir, E. (1993). Choosing versus rejecting: Why some options are both better and worse than others. *Memory & cognition*, *21*(4), 546-556.

Shafir, E. (2018). The Workings of Choosing and Rejecting: Commentary on Many Labs 2. *Advances in Methods and Practices in Psychological Science*, *1*(4), 495-496. DOI: 10.1177/2515245918814812

Slovic, P., Griffin, D., & Tversky, A. (1990). Compatibility effects in judgment and choice. *Insights in decision making: Theory and applications* (pp. 5-27). Chicago, IL: University of Chicago Press.

Sokolova, T., & Krishna, A. (2016). Take it or leave it: How choosing versus rejecting alternatives affects information processing. *Journal of Consumer Research*, *43*(4), 614-635. DOI: 10.1093/jcr/ucw049

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business, 59* (4), 251-278.

Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological review*, *95*(3), 371-384. DOI: 10.1037/0033-295X.95.3.371

van't Veer, A.E., & Giner-Sorolla, R. (2016). Pre-registration in social psychology—A discussion and suggested template. *Journal of Experimental Social Psychology*, *67*, 2-12. DOI: 10.1016/j.jesp.2016.03.004

Vandekerckhove, J., Rouder, J. N., & Kruschke, J. K. (2018). Bayesian methods for advancing psychological science. *25*, 1-4. DOI: 10.3758/s13423-018-1443-8

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

von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior.* Princeton, NJ: Princeton University Press.

Wedell, D. H. (1997). Another look at reasons for choosing and rejecting. *Memory & Cognition, 25*(6), 873-887.

Ziano, I., Wang, Y., Sany, S., Feldman, G., Ngai, L., ... (2020). Perceived morality of direct versus indirect harm: Replications of the preference for indirect harm effect. *Meta Psychology*. Manuscript accepted for publication. https://doi.org/10.31234/osf.io/bs7jf https://psyarxiv.com/bs7jf/

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, *41, E120*. DOI: 10.1017/S0140525X17001972

Choosing versus Rejecting (Shafir, 1993): A pre-registered replication: Supplementary

Contents

Power analyses

We conducted a power analysis of the results described in Shafir (1993) (α = .05, power = .95, G*Power 3.1.9.3). Based on the smallest effect size reported in the Shafir (1993) a required sample size of 1092 subjects was determined. Please refer to Appendix A for detailed power analysis for each of contrasts in the original study.

Open Science

Data and code

Data and code are shared using the Open Science Framework. Review link for data and code of the study: https://osf.io/ve9bg/

Pre-registrations and Qualtrics study designs

link: https://osf.io/r4aku

Procedure and data disclosures

Data collection

Data collection was completed before analyzing the data.

Conditions reporting

All collected conditions are reported.

Data exclusions

Details are reported in the materials section of this document

Variables reporting

All variables collected for this study are reported and included in the provided data.

Formulas employed in the R functions

R does not yet have functions and packages that allow us to conduct one proportions test and two-proportions test directly. Therefore, we built functions to calculate that calculated z-statistic based on the formulas noted below:

Two proportions test:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where:

p1= proportion of subjects with the characteristic of interest in the 1st group (x1/n1)
p2= proportion of subjects with the characteristic of interest in the 2nd group (x2/n2)

and:

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

The R function calculated the estimate of the two proportions, the p-value based on the above formula.

One proportions test:

Again for one proportion test the normal approximation to the binomial distribution to calculate a test statistic ***z.***

The formula for a z-statistic is:

$$z = \frac{\bar{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}$$

Where

- n = Sample size
- p_o = Null hypothesized value
- p-dash = Observed proportion

The R function calculated the estimate of the one proportion z-test, the p-value based on the above formula.

Calculation of effect size:

The effect size was calculated by converting the standard normal deviate (z) into the strength of association (r) using Rosenthal (1984, p.25) and then to the standardized mean difference (Cohen's d) using the equation from Friedman (1968, p.246)

$$r = \frac{z}{\sqrt{N}}$$ ---------------- Rosenthal (1984, p.25)

$$d = \frac{2r}{\sqrt{1-r^2}}$$ ---------------- Friedman (1968, p.246)

Project Process Outline

The current replication is part of the mass pre-registered replication project, with the aim of revisiting well-known research findings in the area of judgment and decision making (JDM) and examining the reproducibility and replicability of these findings.

For each of the replication projects, researchers completed full pre-registrations, data analysis, and APA style submission-ready reports. Each of these four researchers (second to fifth author) independently reproduced the materials and designed the replication experiment, with a separate pre-registration document. The researchers then peer-reviewed one another to try and arrive at the best possible design. Then, then the last two authors reviewed the integrated work and the last corresponding author made final adjustments and conducted the pre-registration and data collection.

The OSF page of the project contains one Qualtrics survey design used for data collection with four pre-registration documents submitted by each of the researchers. In the manuscript, we followed the most conservative of the four pre-registrations.

Verification of Analyses

Initial analyses were conducted by the independent researchers, who were used JAMOVI (jamovi project, 2018) in the analyses. In preparing this manuscript, the lead and corresponding authors verified the analyses in R. One proportions test, two-proportions test, and T-tests were conducted using base R package, point estimates and confidence intervals for Cohen's d were calculated using 'esc' or 'lsr' R package.

Materials and scales used in the experiment

Procedure

Subjects were randomly assigned to one of the two conditions, and in each condition, read eight problems. The survey followed the following sequence:

- Subjects signed the consent form. Then were given instructions, and then were randomly assign to one of the two conditions.
- Demographics questions.
- After that, subjects filled the funneling section that checked if they are seriously filling in the survey, and if they can guess the purpose of the study.

Exclusion criteria

In the pre-registration we included the following:
"We will focus on our analyses on the full sample. However, as a supplementary analysis and to examine any potential issues, we will also determine further findings reports with exclusions. In any case, we will report exclusions in detail with results for full sample and results following exclusions (in either the manuscript or the supplementary).
General criteria:

1. Subjects indicating a low proficiency of English (self-report<5, on a 1-7 scale)
2. Subjects who self-report not being serious about filling in the survey (self-report<4, on a 1-5 scale).
3. Subjects who correctly guessed the hypothesis of this study in the funneling section.
4. Have seen or done the survey before
5. Subjects who failed to complete the survey. (duration = 0, leave question blank)
6. Not from the United States"

Instructions and experimental material

All subjects first read the instruction:

This survey consists of a scale, followed by 8 decision-making problems with 2-3 items each. In each problem, you will first make your decision on two or more options in each problem, followed by rating your feelings about each option on a 6-point scale.

Read the questions and choice options carefully.
There are no right or wrong answers, please answer to the best of your understanding, based on your own preferences and intuition.

After that, subjects answered 6 survey items (order randomized) on a scale that ranged between *1 (Strongly disagree) and 7 (Strongly agree)*. Two of the six items were attention check items. The read:

Before we begin with the scenarios, please answer these short questions about your general attitudes towards choice by indicating your agreement with the following statements.

- It's very hard for me to choose between many alternatives.
- When faced with an important decision, I prefer that someone else chooses for me.
- The more choices I have in life, the better.

- In each decision I face, I prefer to have as many options as possible to choose from.
- Fifty is more than one hundred.
- I am human and I read each item carefully.

After that, subjects were randomly assigned to one of two experimental conditions and in each Answered 8 problems. The order of the problems within each condition was randomized.

Experimental condition: Choosing

i. Problem 1

Scenario - Imagine that you serve on the jury of an only-child sole-custody case following a relatively messy divorce. The facts of the case are complicated by ambiguous economic, social, and emo-tional considerations, and you decide to base your decision entirely on the following few observations.

Question - To which parent would you **award** sole custody of the child?

Options:

Parent A

average income

average health

average working hours

reasonable rapport with the child

relatively stable social life

Parent B

above-average income

very close relationship with the child

extremely active social lira5bfe

lots of work-related travel

minor health problems

ii. Problem 2

Scenario - Imagine that you are planning a week vacation in a warm spot over spring break. You currently have two options that are reasonably priced. The travel brochure gives only a limited amount of information about the two options.

Question - Given the information available, which vacation spot would you **prefer**?

Options:

Spot A

average weather

average beaches

medium-quality hotel

medium-temperature water

average nightlife

Spot B

lots of sunshine

gorgeous beaches and coral reefs

ultra-modern hotel

very cold water

very strong winds

no nightlife

iii. Problem 3

Scenario - Assume that you are an undergraduate student and would eventually need to take two courses to fulfill your graduation requirements. But you could only take one in the coming semester, and the other at some time later. You need to make your decision based on the characteristics of the courses.

Question - With the information below, which course would you take in the coming semester?

Options:

Course X is considered an average course, with a reasonable reading list, and with an average work load.

Course Y has an extremely interesting reading list and is taught by a professor who is supposed to be very good. It has the reputation of a tough course, slow-going at times, and it meets more hours per week than the usual.

iv. Problem 4

Scenario - Imagine that you were invited to play one of the following two lotteries.

Question - Which one would you **prefer**?

Options:

Lottery 1 You have a 50% chance to win $50, otherwise nothing.

Lottery 2 You have an 80% chance to win $150, and a 20% chance to lose $10.

v. Problem 5

Scenario - Imagine that you were invited to play one of the following two lotteries.

Question - Which one would you **prefer**?

Lottery 1 You have a 20% chance to win $50, otherwise nothing.

Lottery 2 You have a 60% chance to win $100, and a 40% chance to lose $5.

vi. Problem 6

Scenario - You go to your favorite ice-cream parlor, and have to decide between two flavors: Flavor A is good; Flavor B is excellent, but is high in cholesterol.

Question - Which do you **choose**?

Flavor A Good

Flavor B Excellent, but is high in cholesterol

vii. Problem 7

Scenario - Imagine that you are voting for the president of your town council. You are now considering two final candidates. A friend who is knowledgeable in the area of local politics gives you the following information about them. You find the choice difficult and are trying to decide which candidate to vote for.

Question - Based on the information below, which candidate would you decide to **vote for**?

Candidate A

Enjoys camping and other outdoor activities

Is a local businessman

Was voted "Most Enthusiastic" in high school

Has two children enrolled in the local elementary school

Majored in history in college

Candidate B

Served honorably as the vice president of the council last term

Organized a fund raiser to support the local children's hospital

Was voted "Best Looking" in high school

Has bragged about his promiscuity in the past

Refused to disclose income tax records despite repeated requests

viii. Problem 8

Scenario - Imagine that you were invited to play one of the following three lotteries.

Question - Which one would you **<u>prefer</u>**?

Lottery 1 You have a 50% chance to win $50, otherwise nothing.

Lottery 2 You have a 60% chance to win $50, otherwise nothing.

Lottery 3 You have an 80% chance to win $150, and a 20% chance to lose $20.

Extension question followed by the decision for every problem

Question - Please rate each option from 0 (very bad) to (very good)

0. Very bad
1. Bad
2. Slightly bad
3. Slightly good
4. Good
5. Very good

Experimental condition: Rejecting

b. Independent variable manipulation

i. Problem 1

Scenario - Imagine that you serve on the jury of an only-child sole-custody case following a relatively messy divorce. The facts of the case are complicated by ambiguous economic, social, and emo-tional considerations, and you decide to base your decision entirely on the following few observations.

Question - To which parent would you **<u>deny</u>** sole custody of the child?

Options:

Parent A

average income

average health

average working hours

reasonable rapport with the child

relatively stable social life

Parent B

above-average income

very close relationship with the child

extremely active social lira5bfe

lots of work-related travel

minor health problems

ii. Problem 2

Scenario - Imagine that you are planning a week vacation in a warm spot over spring break. You currently have two options that are reasonably priced. The travel brochure gives only a limited amount of information about the two options.

Question - Given the information available, which vacation spot would you **<u>cancel</u>**?

Options:

Spot A

average weather

average beaches

medium-quality hotel

medium-temperature water

average nightlife

Spot B

lots of sunshine

gorgeous beaches and coral reefs

ultra-modern hotel

very cold water

very strong winds

no nightlife

iii. Problem 3

Scenario - Assume that you are an undergraduate student and would eventually need to take two courses to fulfill your graduation requirements. But you could only take one in the coming semester, and the other at some time later. You need to make your decision based on the characteristics of the courses.

Question - With the information below, which course would you **leave for later**?

Options:

Course X is considered an average course, with a reasonable reading list, and with an average work load.

Course Y has an extremely interesting reading list and is taught by a professor who is supposed to be very good. It has the reputation of a tough course, slow-going at times, and it meets more hours per week than the usual.

iv. Problem 4

Scenario - Imagine that you were invited to play one of the following two lotteries.

Question - Which one would you **give up**?

Options:

Lottery 1 You have a 50% chance to win $50, otherwise nothing.

Lottery 2 You have an 80% chance to win $150, and a 20% chance to lose $10.

v. Problem 5

Scenario - Imagine that you were invited to play one of the following two lotteries.

Question - Which one would you **give up**?

Lottery 1 You have a 20% chance to win $50, otherwise nothing.

Lottery 2 You have a 60% chance to win $100, and a 40% chance to lose $5.

vi. Problem 6

Scenario - You go to your favorite ice-cream parlor, and have to decide between two flavors: Flavor A is good; Flavor B is excellent, but is high in cholesterol.

Question - Which do you **give up**?

Flavor A Good

Flavor B Excellent, but is high in cholesterol

vii. Problem 7

Scenario - Imagine that you are voting for the president of your town council. You are now considering two final candidates. A friend who is knowledgeable in the area of local politics gives you the following information about them. You find the choice difficult and are trying to decide which candidate to vote for.

Question - Based on the information below, which candidate would you decide to **<u>not vote for</u>**?

Candidate A

Enjoys camping and other outdoor activities

Is a local businessman

Was voted "Most Enthusiastic" in high school

Has two children enrolled in the local elementary school

Majored in history in college

Candidate B

Served honorably as the vice president of the council last term

Organized a fund raiser to support the local children's hospital

Was voted "Best Looking" in high school

Has bragged about his promiscuity in the past

Refused to disclose income tax records despite repeated requests

viii. Problem 8

Scenario - Imagine that you were invited to play one of the following three lotteries.

Question - Which one would you **<u>give up</u>**?

Lottery 1 You have a 50% chance to win $50, otherwise nothing.

Lottery 2 You have a 60% chance to win $50, otherwise nothing.

Lottery 3 You have an 80% chance to win $150, and a 20% chance to lose $20.

Extension question followed by the decision for every problem

Question - Please rate each option from 0 (very bad) to (very good)

0. Very bad
1. Bad
2. Slightly bad
3. Slightly good
4. Good

5. Very good

Funneling section

Three funneling questions:

- *What do you think the purpose of the last part was?*
- *Have you ever seen the materials used in this study or similar before? If yes - please indicate where*
- *Did you spot any errors? Anything missing or wrong? Something we should pay attention to in next runs? (Briefly, up to one sentence, write "none" if not relevant)*

Finally, subjects were asked to fill in demographics and were debriefed. No filler items were included.

Additional Tables and Figures

Moved from the main manuscript to keep manuscript short and concise.

Table S1

Difference and similarities between original studies and the replication attempt

	Original Study	Replication Study	Reason of changes
Number of problems per subject	Author of the original study notes that on an average 2 or 3 problems reported in the original study were present to a subject	Subjects provided a response to all the 8 problems	The current study with an aim to replicate the effects of the original study included all the 8 problems.
Filler items	Included	Not included	Including filler items (e.g., unrelated) along with 8 problems have caused respondent fatigue could lead to deterioration of the quality of the responses.
Procedure	Problems were presented in a booklet format	An online survey (Qualtrics) was used.	Allows minimal error in data collection and entry, and useful in faster data collection.
Sample population	Undergraduates from American university.	The online marketplace Amazon Mechanical Turk (mTurk) from an expected that the recruited subjects varied on demographic variables.	To recruit more subjects
Sample Size	Ranged between 139 to 424 across 8 problems	1028 across four experimental conditions (average of 257 subjects/condition)	See power analysis in Supplementary material Part

Table S2

Preregistration planning and deviation documentation

Components of pre-registration	Were there deviations?	If yes describe the details of the deviation(s)	Rationale for deviation	How might the results be different if had not deviated
Procedures	No	N/A	N/A	N/A
Power analysis	No	N/A	N/A	N/A
Exclusion rules	No	N/A	N/A	N/A
Evaluation criteria	Minor	N/A	N/A	N/A
Analysis	Minor additions	Bayesian analysis was performed in addition to null-hypothesis significance tests (NHST).	Bayesian analysis is useful in testing for and quantifying an absence of an effect	With the additional tests along with NHST tests, we are not only able to falsify predictions about the presence of effects, but also declare the absence of meaningful effects.
		Additional analysis that tested for the Accentuation hypothesis.	The collected that closely matched the replication also allowed for testing of the alternate predictions based on the Accentuation hypothesis (Wedell, 1997)	The results of the replication do not change with the additional analysis that tested Accentuation hypothesis

Figures

Target similarity	Highly similar				Highly dissimilar
Category	Direct replication				Conceptual replication
Design facet	**Exact replication**	**Very close replication**	**Close replication**	**Far replication**	**Very far replication**
IV operationalization	Same	Same	Same	Different	
DV operationalization	Same	Same	Same	Different	
IV stimuli	Same	Same	Different		
DV stimuli	Same	Same	Different		
Procedural details	Same	Different			
Physical setting	Same	Different			
Contextual variables	Different				

Figure S1. Criteria for evaluation of replications by LeBel et al. (2018). A classification of relative methodological similarity of a replication study to an original study. “Same” (“different”) indicates the design facet in question is the same (different) compared to an original study. IV = independent variable. DV = dependent variable. “Everything controllable” indicates design facets over which a researcher has control. Procedural details involve minor experimental particulars (e.g., task instruction wording, font, font size, etc.).

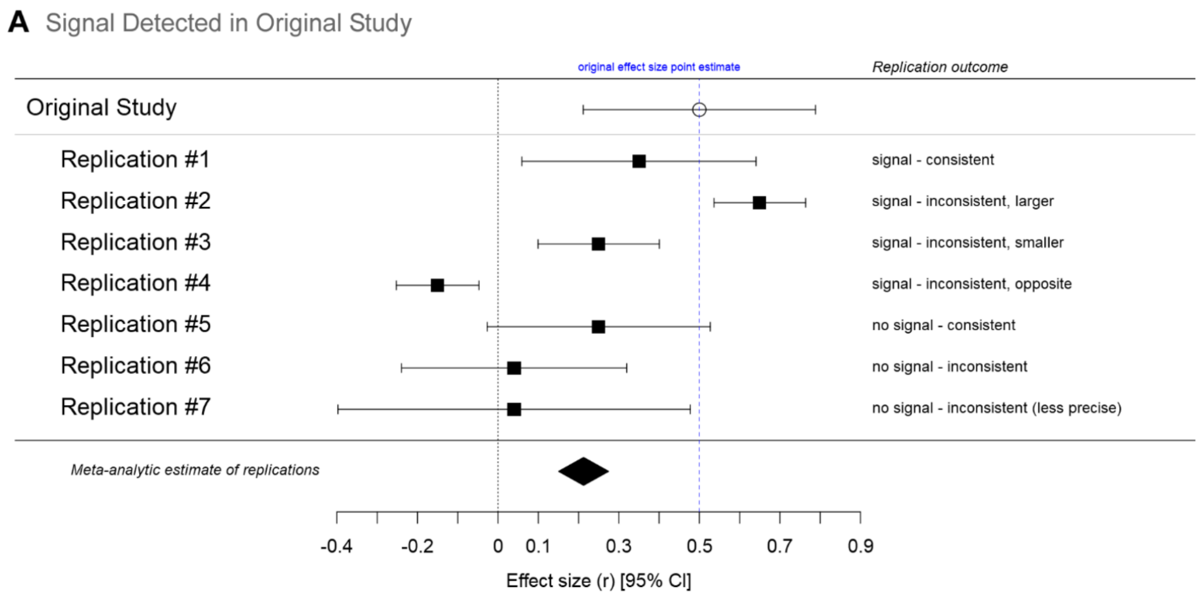

Figure S2. Criteria for evaluation of replications by LeBel et al. (2019). A taxonomy for comparing replication effects to target article original findings.

Table S3

Descriptive statistics of the additional measures

Variable	**M**	**SD**	**Skewness**	**Kurtosis**	**n**
Perceived ability to choose	5.00	1.35	-0.54	-0.23	1026
Preference for choice	5.07	1.25	-0.52	-0.04	1026
Choose condition					
Attractiveness of ‘Parent A’ option	3.55	0.78	-0.51	0.94	514
Attractiveness of ‘Parent B’ option	3.52	0.96	-0.56	0.63	514
Attractiveness of ‘Spot A’ option	3.39	0.84	-0.62	1.19	514
Attractiveness of ‘Spot B’ option	3.25	1.12	-0.37	-0.25	514
Attractiveness of ‘Course X’ option	3.51	0.77	-0.54	1.54	514
Attractiveness of ‘Course Y’ option	3.65	1.05	-0.71	0.28	514
Attractiveness of ‘Lottery 1’ option	3.37	1.04	-0.75	0.88	514
Attractiveness of ‘Lottery 2’ option	3.76	1.06	-0.93	0.74	514
Attractiveness of ‘Lottery 1’ option	2.75	1.22	-0.29	-0.33	514
Attractiveness of ‘Lottery 2’ option	3.48	1.05	-0.73	0.53	514
Attractiveness of ‘Flavor A’ option	3.74	0.76	-1.42	4.55	514
Attractiveness of ‘Flavor B’ option	3.26	1.39	-0.43	-0.69	514
Attractiveness of ‘Candidate A’ option	3.81	0.96	-1.14	1.86	514
Attractiveness of ‘Candidate B’ option	2.03	1.26	0.25	-0.56	514
Attractiveness of ‘Lottery 1’ option	3.24	0.96	-0.41	0.76	514
Attractiveness of ‘Lottery 2’ option	3.39	0.95	-0.69	0.98	514
Attractiveness of ‘Lottery 3’ option	3.58	1.20	-0.71	-0.07	514

Note: M = Mean; SD = Standard deviation.

Table S4

Descriptive statistics of the additional measures

Variable	M	SD	Skewness	Kurtosis	n
Reject condition					
Attractiveness of ‘Parent A’ option	3.50	0.80	-0.57	1.38	512
Attractiveness of ‘Parent B’ option	3.51	0.94	-0.58	0.36	512
Attractiveness of ‘Spot A’ option	3.36	0.85	-0.56	1.05	512
Attractiveness of ‘Spot B’ option	3.23	1.06	-0.30	-0.37	512
Attractiveness of ‘Course X’ option	3.40	0.79	-0.13	0.18	512
Attractiveness of ‘Course Y’ option	3.61	1.02	-0.66	0.34	512
Attractiveness of ‘Lottery 1’ option	3.20	0.94	-0.75	1.52	512
Attractiveness of ‘Lottery 2’ option	3.77	1.10	-0.97	0.74	512
Attractiveness of ‘Lottery 1’ option	2.63	1.11	-0.24	-0.41	512
Attractiveness of ‘Lottery 2’ option	3.39	1.02	-0.61	0.28	512
Attractiveness of ‘Flavor A’ option	3.75	0.74	-1.17	3.00	512
Attractiveness of ‘Flavor B’ option	3.32	1.35	-0.49	-0.61	512
Attractiveness of ‘Candidate A’ option	3.65	1.04	-0.93	0.98	512
Attractiveness of ‘Candidate B’ option	2.13	1.29	0.24	-0.62	512
Attractiveness of ‘Lottery 1’ option	3.17	0.91	-0.65	1.11	512
Attractiveness of ‘Lottery 2’ option	3.41	0.86	-0.43	0.48	512
Attractiveness of ‘Lottery 3’ option	3.66	1.17	-0.77	0.02	512

Note: M = Mean; SD = Standard deviation.

Table S5

Summary of findings of compatibility hypothesis based on additional variables

Comparison	Problem	Replication		Bayes Factor
		T-statistic	**Cohen's d**	
Attractiveness of enriched option between choose and reject experimental conditions	1	*t* (1023.86) = 0.10, *p* = .92	0.01 [-0.12, 0.13]	BF_{10} = 0.08; BF_{01}=13.22
	2	*t* (1021) = 0.27, *p* = .79	0.02 [-0.11, 0.14]	BF_{10} = 0.09; BF_{01}=11.37
	3	*t* (1023.51) = 0.72, *p* = .47	0.04 [-0.08, 0.17]	BF_{10} = 0.14; BF_{01}=7.28
	4	*t* (1022.53) = -0.07, *p* = .94	0.00 [-0.12, 0.13]	BF_{10} = 0.07; BF_{01}=15.14
	5	*t* (1023.73) = 1.42, *p* = .16	0.09 [-0.03, 0.21]	BF_{10} = 0.35; BF_{01}=2.87
	6	*t* (1023.13) = -0.72, *p* = .47	0.04 [-0.08, 0.17]	BF_{10} = 0.04; BF_{01}=23.31
	7	*t* (1023.13) = -1.30, *p* = .19	0.08 [-0.04, 0.20]	BF_{10} = 0.03; BF_{01}=31.61
	8	*t* (1023.42) = -1.09, *p* = .28	0.07 [-0.05, 0.19]	BF_{10} = 0.04; BF_{01}=28.50
Relative Attractiveness of enriched option between choose and reject experimental conditions	1	*t* (1022.91) = -0.60, *p* = .55	0.04 [-0.09, 0.16]	BF_{10} = 0.05; BF_{01}=21.68
	2	*t* (1023.69) = -0.19, *p* = .85	0.01 [-0.11, 0.13]	BF_{10} = 0.06; BF_{01}=16.56
	3	*t* (1020.46) = -0.69, *p* = .49	0.04 [-0.08, 0.17]	BF_{10} = 0.04; BF_{01}=22.91
	4	*t* (1019.23) = -1.97, *p* = .05	0.12 [0.00, 0.25]	BF_{10} = 0.02; BF_{01}=41.73
	5	*t* (1020.76) = -0.23, *p* = .82	0.01 [-0.11, 0.14]	BF_{10} = 0.06; BF_{01}=17.03
	6	*t* (1023.62) = -0.53, *p* = .60	0.03 [-0.09, 0.16]	BF_{10} =0.05; BF_{01}=20.78
	7	*t* (1020.49) = -2.26, *p* = .02	0.14 [0.02, 0.26]	BF_{10} = 0.02; BF_{01}=46.27
	8	*t* (603.90) = 2.22, *p* = .03	0.14 [0.02, 0.26]	BF_{10} =1.54 ; BF_{01}=0.65

Note. N = 1026;

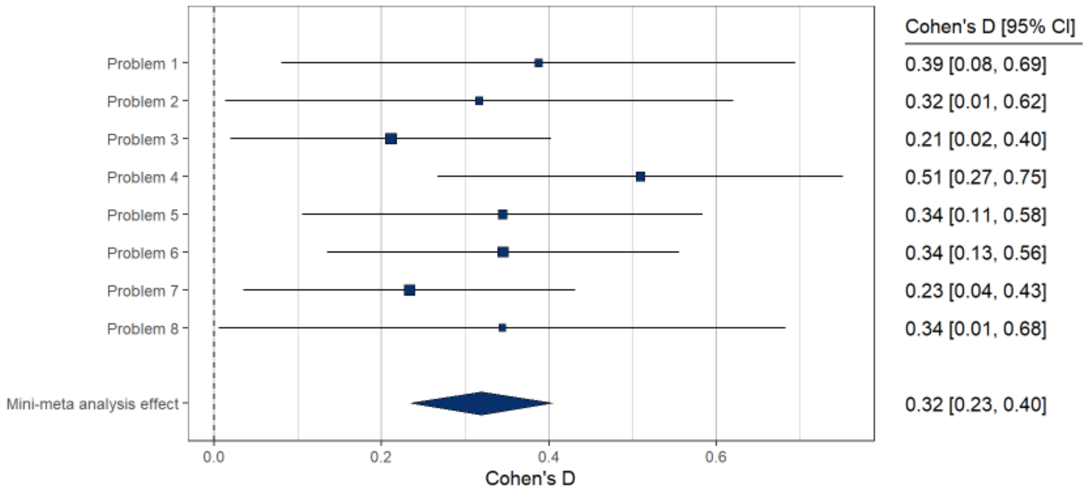

Figure S3.

Forest plots of the mini meta-analytic effect sizes for Hypothesis 1 across eight decision problems in the original study. CI = confidence interval.

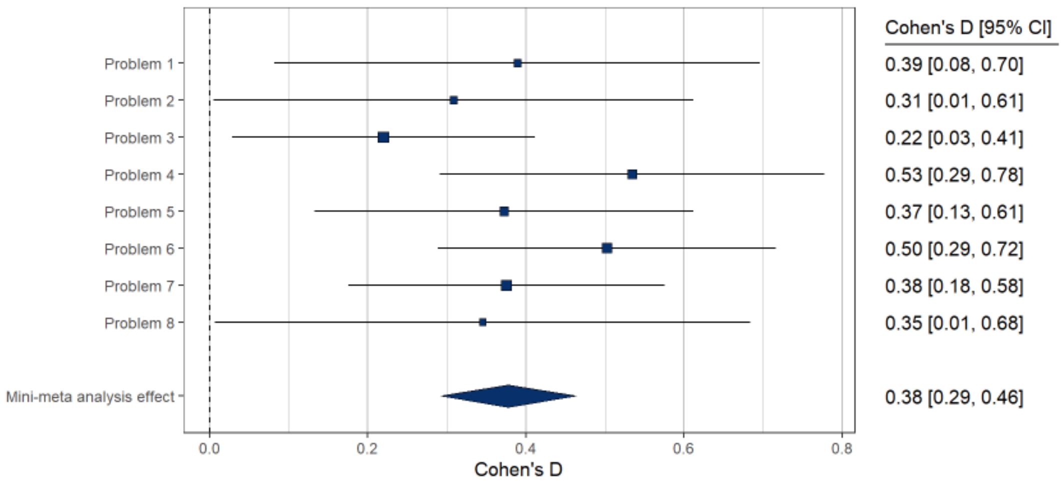

Figure S4

Forest plots of the mini meta-analytic effect sizes for Hypothesis 2 across eight decision problems in the original study. CI = confidence interval.

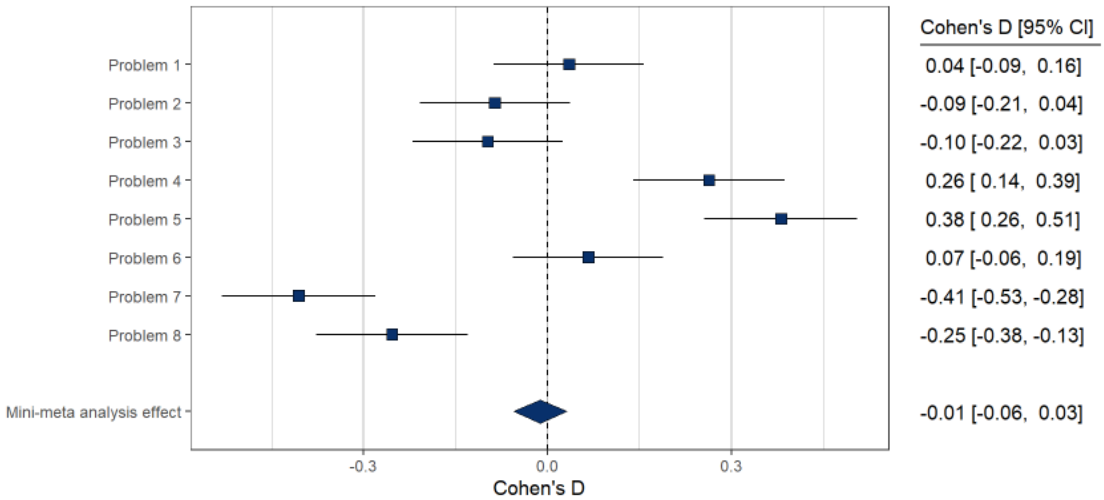

Figure S5

Forest plots of the mini meta-analytic effect sizes for Hypothesis 1 across eight decision problems in the replication study. CI = confidence interval.

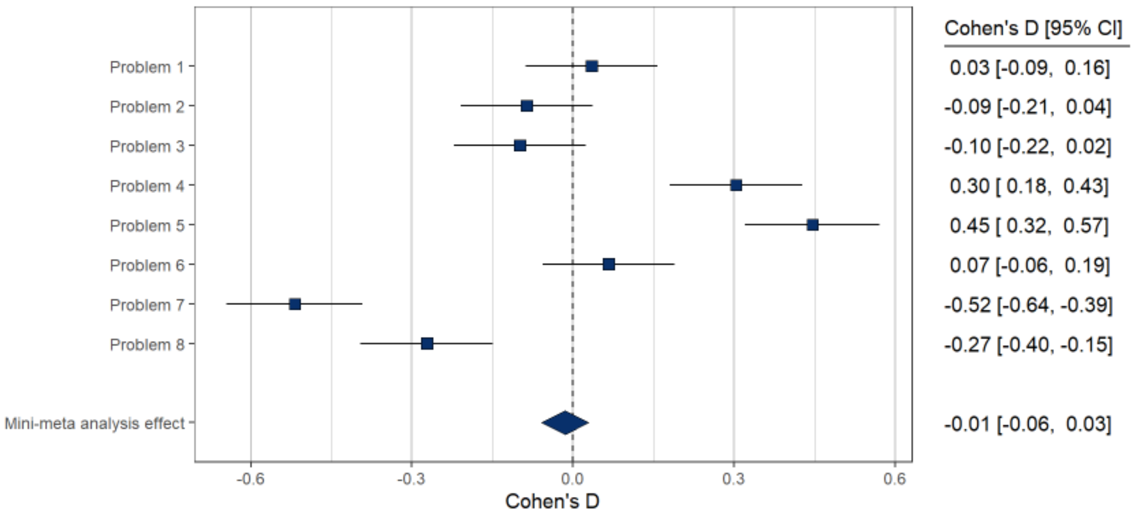

Figure S6

Forest plots of the mini meta-analytic effect sizes for Hypothesis 2 across eight decision problems in the replication study. CI = confidence interval.

Individual differences in preference for enriched alternative.

We examined how individual differences influence the way we make choices. We looked at the exploratory hypotheses that tested if an individual's perceived ability to choose and preference for choice could influence choices.

Prediction: Individuals that rate themselves with a high ability to choose and prefer to have choices focus more on the positive aspects of options. Given the enriched option endowed with more positive features than the impoverished option we expected these individuals more often select the enriched option in both choosing- and rejecting-condition.

We tested the prediction using two separate (one with 'ability to choose' as IV and other with 'preference for choice' as IV) binary logistic mixed-effects regression analysis. In this analysis, we included responses from Problem 1 to 7, as these problems shared the common procedure of choosing between two alternatives (binary).

'Ability to choose' on preference for enriched:

We conducted a binary logistic mixed-effects regression analysis in which experimental condition, ability to choose, and the interaction term (Experiment condition x Ability to choose) were the fixed effects predictors of choosing enriched option (Yes =1; No = 0). The regression included subject ID as a random effect predictor.

The results of the regression revealed the main effect of 'ability to choose' was not significant Wald χ^2 (1) = 0.90, *p* = .343). The interaction term introduced in step 2 was not significant either: χ^2 (1) = 0.49, *p* =.485). See the results in Table S6.

We also tested for the correlations between 'ability to choose' measure and attractiveness of choice in each of the problems (See Table S8)

'Preference for choice' on preference for enriched:

We conducted a binary logistic mixed-effects regression analysis in which experimental condition, preference for choice, and the interaction term (Experiment condition x preference for choice) was the fixed effects predictors of choosing enriched option (Yes =1; No = 0). The regression included subject ID as a random effect predictor.

The results of the regression revealed the main effect of 'preference for choice' was not significant Wald χ^2 (1) = 0.41, *p* =.522). The interaction term introduced in step 2 was not significant either: χ^2 (1) = 1.24, *p* =.266). See the results in Table S7.

We also tested for the correlations between 'preference for choice' measure and attractiveness of choice in each of the problems (See Table S9)

Table S6

Results of binary logistic mixed-effects regression

	Dependent variable: Predicted probability of enriched	
	Main effect	Interaction
Constant	0.12 (0.098)	0.185 (0.135)
Experimental condition (EXP) (1=Choose; 0=Reject)	−0.043 (0.049)	−0.171 (0.189)
Ability to choose (AC)	0.017 (0.018)	0.004(0.026)
EXP x AC		0.026 (0.036)
Observations	7,182	7,182
Log Likelihood	−4,946.272	−4,946.028
Akaike Inf. Crit.	9,900.55	9,902.06
Bayesian Inf. Crit.	9,928.06	9,936.45

Note: ∗ p<0.1; ∗∗ p<0.05; ∗∗∗ p<0.01

Table S7

Results of binary logistic mixed-effects regression

	Dependent variable: Predicted probability of enriched	
	Main effect	Interaction
Constant	0.143 (0.106)	0.037 (0.142)
Experimental condition (EXP) (1=Choose; 0=Reject)	−0.045 (0.049)	0.177 (0.205)
Ability to choose (AC)	0.013 (0.020)	0.034 (0.027)
EXP x AC		−0.044 (0.039)
Observations	7,182	7,182
Log Likelihood	−4,946.516	−4,945.899
Akaike Inf. Crit.	9,901.03	9,901.80
Bayesian Inf. Crit.	9,928.55	9,936.19

Table S8

The correlation between perceived '*ability to choose'* and variables listed in the table.

Variable	**n**	**r**	**p**	**LL**	**UL**
Preference for choice	1024	0.28	0.000	0.22	0.33
Choose condition					
Attractiveness of 'Parent A' option	512	0.01	0.884	-0.08	0.09
Attractiveness of 'Parent B' option	512	-0.07	0.105	-0.16	0.01
Attractiveness of 'Spot A' option	512	0.00	0.982	-0.09	0.09
Attractiveness of 'Spot B' option	512	-0.08	0.089	-0.16	0.01
Attractiveness of 'Course X' option	512	0.01	0.776	-0.07	0.10
Attractiveness of 'Course Y' option	512	0.05	0.306	-0.04	0.13
Attractiveness of 'Lottery 1' option	512	0.00	0.997	-0.09	0.09
Attractiveness of 'Lottery 2' option	512	0.05	0.288	-0.04	0.13
Attractiveness of 'Lottery 1' option	512	-0.05	0.222	-0.14	0.03
Attractiveness of 'Lottery 2' option	512	0.03	0.558	-0.06	0.11
Attractiveness of 'Flavor A' option	512	0.05	0.219	-0.03	0.14
Attractiveness of 'Flavor B' option	512	-0.01	0.861	-0.09	0.08
Attractiveness of 'Candidate A' option	512	0.07	0.093	-0.01	0.16
Attractiveness of 'Candidate B' option	512	-0.13	0.002	-0.22	-0.05
Attractiveness of 'Lottery 1' option	512	-0.02	0.584	-0.11	0.06
Attractiveness of 'Lottery 2' option	512	0.03	0.455	-0.05	0.12
Attractiveness of 'Lottery 3' option	512	0.09	0.031	0.01	0.18
Reject condition					
Attractiveness of 'Parent A' option	510	0.06	0.180	-0.03	0.15
Attractiveness of 'Parent B' option	510	0.03	0.478	-0.06	0.12
Attractiveness of 'Spot A' option	510	0.05	0.226	-0.03	0.14
Attractiveness of 'Spot B' option	510	0.00	0.949	-0.09	0.08
Attractiveness of 'Course X' option	510	0.01	0.800	-0.08	0.10
Attractiveness of 'Course Y' option	510	0.14	0.001	0.06	0.23
Attractiveness of 'Lottery 1' option	510	0.03	0.552	-0.06	0.11
Attractiveness of 'Lottery 2' option	510	0.13	0.005	0.04	0.21
Attractiveness of 'Lottery 1' option	510	-0.04	0.314	-0.13	0.04
Attractiveness of 'Lottery 2' option	510	0.05	0.301	-0.04	0.13
Attractiveness of 'Flavor A' option	510	0.04	0.421	-0.05	0.12
Attractiveness of 'Flavor B' option	510	0.01	0.735	-0.07	0.10
Attractiveness of 'Candidate A' option	510	0.06	0.208	-0.03	0.14
Attractiveness of 'Candidate B' option	510	-0.08	0.089	-0.16	0.01
Attractiveness of 'Lottery 1' option	510	-0.05	0.306	-0.13	0.04
Attractiveness of 'Lottery 2' option	510	0.01	0.754	-0.07	0.10
Attractiveness of 'Lottery 3' option	510	0.14	0.002	0.05	0.22

Note: r = Pearson correlation coefficient; LL= lower limit of r estimate; UL= upper limit of r estimate;

Table S9

The correlation between *'Preference for choice'* and variables listed in the table.

Variable	n	r	p	LL	UL
perceived ability to choose	0.28	0.000	0.22	0.33	0.28
Choose condition					
Attractiveness of 'Parent A' option	512	-0.02	0.711	-0.10	0.07
Attractiveness of 'Parent B' option	512	0.12	0.006	0.04	0.21
Attractiveness of 'Spot A' option	512	0.10	0.020	0.02	0.19
Attractiveness of 'Spot B' option	512	-0.01	0.856	-0.09	0.08
Attractiveness of 'Course X' option	512	0.03	0.435	-0.05	0.12
Attractiveness of 'Course Y' option	512	0.12	0.008	0.03	0.20
Attractiveness of 'Lottery 1' option	512	-0.02	0.620	-0.11	0.06
Attractiveness of 'Lottery 2' option	512	0.09	0.051	0.00	0.17
Attractiveness of 'Lottery 1' option	512	-0.01	0.900	-0.09	0.08
Attractiveness of 'Lottery 2' option	512	0.15	0.000	0.07	0.24
Attractiveness of 'Flavor A' option	512	0.09	0.036	0.01	0.18
Attractiveness of 'Flavor B' option	512	-0.03	0.520	-0.11	0.06
Attractiveness of 'Candidate A' option	512	0.14	0.001	0.06	0.23
Attractiveness of 'Candidate B' option	512	0.02	0.650	-0.07	0.11
Attractiveness of 'Lottery 1' option	512	0.00	0.978	-0.09	0.09
Attractiveness of 'Lottery 2' option	512	0.01	0.755	-0.07	0.10
Attractiveness of 'Lottery 3' option	512	0.17	0.000	0.08	0.25
Reject condition					
Attractiveness of 'Parent A' option	510	0.02	0.613	-0.06	0.11
Attractiveness of 'Parent B' option	510	0.04	0.405	-0.05	0.12
Attractiveness of 'Spot A' option	510	0.07	0.092	-0.01	0.16
Attractiveness of 'Spot B' option	510	-0.02	0.612	-0.11	0.06
Attractiveness of 'Course X' option	510	-0.01	0.841	-0.10	0.08
Attractiveness of 'Course Y' option	510	0.14	0.002	0.05	0.22
Attractiveness of 'Lottery 1' option	510	-0.01	0.841	-0.10	0.08
Attractiveness of 'Lottery 2' option	510	0.09	0.048	0.00	0.17
Attractiveness of 'Lottery 1' option	510	0.02	0.707	-0.07	0.10
Attractiveness of 'Lottery 2' option	510	0.04	0.342	-0.04	0.13
Attractiveness of 'Flavor A' option	510	-0.02	0.678	-0.10	0.07
Attractiveness of 'Flavor B' option	510	-0.01	0.900	-0.09	0.08
Attractiveness of 'Candidate A' option	510	0.09	0.046	0.00	0.17
Attractiveness of 'Candidate B' option	510	-0.07	0.098	-0.16	0.01
Attractiveness of 'Lottery 1' option	510	0.02	0.613	-0.06	0.11
Attractiveness of 'Lottery 2' option	510	0.04	0.404	-0.05	0.12
Attractiveness of 'Lottery 3' option	510	0.14	0.002	0.05	0.22

Note: r = Pearson correlation coefficient; LL= lower limit of r estimate; UL= upper limit of r estimate;

Table S10

Results of binary logistic mixed-effects regression with two simultaneous random factors: 1) random intercepts for participant; 2) random intercepts and random conditions slopes for stimuli (i.e. problem numbers)

	Dependent variable: Predicted probability of enriched alternative	
	Main effect	Interaction
Constant	-3.51*** (0.164)	-3.61*** (0.186)
Overall proportion preferring enriched (PEN)	6.82*** (0.287)	7.00*** (0.325)
Experimental condition (EXP) (1 = Choose; 0 = Reject)	-0.04 (0.261)	2.13*** (0.246)
PEN × EXP		−3.98*** (0.426)
Observations	7,182	7,182
Log Likelihood	−4,402.48	−4,386.45
Akaike Inf. Crit.	8,818.96	8,788.91
Bayesian Inf. Crit.	8,847.12	8,843.94

Note: *$p < 0.1$; **$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

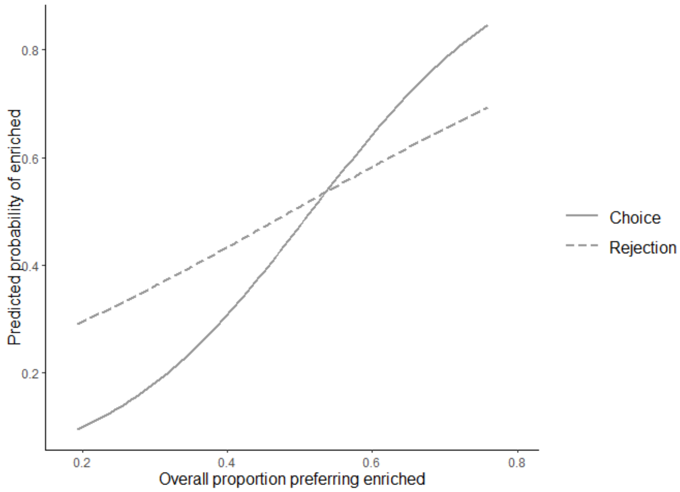

Figure S7

Predicted probability of the enriched alternative in choice and rejection tasks as a function of overall preference for the enriched alternative. Fitted lines are the marginal effects of interaction terms. The relative attractiveness variable used in the regression was calculated based on the responses to extension variables. The model specification included two random factors: 1)random intercepts for participants, 2) both random intercepts and random conditions slopes for stimuli (i.e. problem numbers)

Table S11

Results of binary logistic mixed-effects regression with two simultaneous random factors: 1) random intercepts for participant; 2) random intercepts and random conditions slopes for stimuli (i.e. problem numbers)

	Dependent variable: Predicted probability of enriched alternative	
	Main effect	Interaction
Constant	0.32 (0.391)	0.37 (0.401)
Relative attractiveness of enriched alternative (AEO)	0.64*** (0.022)	0.95*** (0.039)
Experimental condition (EXP) (1 = Choose; 0 = Reject)	-0.10** (0.319)	-0.17 (0.300)
AEO ✕ XP		−0.51*** (0.047)
Observations	7,182	7,182
Log Likelihood	−3,886.71	−3,825.46
Akaike Inf. Crit.	7,787.48	7,666.92
Bayesian Inf. Crit.	7,835.64	7,721.95

Note: *p < 0.1; **p < 0.05; **p < 0.01. The relative attractiveness variable used in the regression was calculated based on the responses to extension variables.

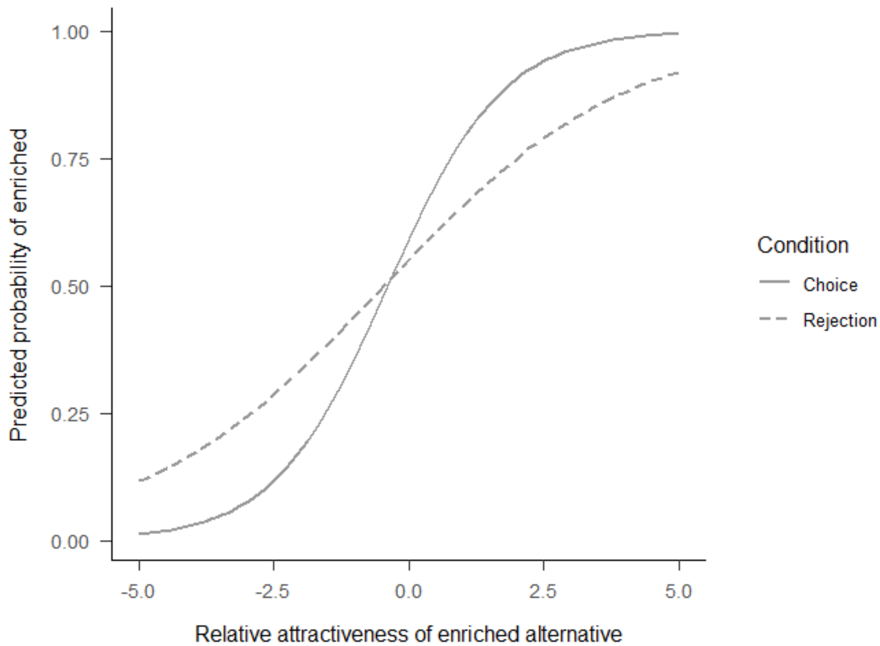

Figure S8

Predicted probability of the enriched alternative in choice and rejection tasks as a function of overall preference for the enriched alternative. Fitted lines are the marginal effects of interaction terms. The relative attractiveness variable used in the regression was calculated based on the responses to extension variables. The model specification included two random factors: 1)random intercepts for participants, 2) both random intercepts and random conditions slopes for stimuli (i.e. problem numbers)

References

Friedman, H. (1968). "Magnitude of experimental effect and a table for its rapid estimation," *Psychological Bulletin*, 70(4): 245-251.

Rosenthal, R. (1984), *Meta-Analytic Procedures for Social Research*. Newbury Park: Sage

Shafir, E. (1993). Choosing versus rejecting: Why some options are both better and worse than others. *Memory & cognition*, *21*(4), 546-556. DOI: https://doi.org/10.3758/BF03197186

Appendix A

Power analysis was run by G*Power 3.1 (Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G., 2009) to calculate the sample size needed for each question. For Hypothesis 1, the generic binomial test was used; p1 stays 0.5 as it is the null value while p2 is calculated by the proportion of the enriched option over the proportion of both enriched and impoverished options (Credits to Chu Tsz Ching Connie). For Hypothesis 2, z test of comparing two proportions was used; p1 is the proportion of choosing the enriched option while p2 is the proportion of rejecting the enriched option.

Hypothesis 1

Problem 1

Input:	Tail(s)	=	One
	Proportion p2	=	0.595
	α err prob	=	0.05
	Power (1-β err prob)	=	0.95
	Proportion p1	=	0.5
Output:	Lower critical N	=	166
	Upper critical N	=	166
	Total sample size	=	302
	Actual power	=	0.9513149
	Actual α	=	0.0475007

Problem 2

Input:	Tail(s)	=	One
	Proportion p2	=	0.575
	α err prob	=	0.05
	Power (1-β err prob)	=	0.95
	Proportion p1	=	0.5
Output:	Lower critical N	=	258
	Upper critical N	=	258
	Total sample size	=	479
	Actual power	=	0.9508633
	Actual α	=	0.0499467

Problem 3

Input:	Tail(s)	=	One
	Proportion p2	=	0.55
	α err prob	=	0.05
	Power (1-β err prob)	=	0.95
	Proportion p1	=	0.5
Output:	Lower critical N	=	574
	Upper critical N	=	574
	Total sample size	=	1092
	Actual power	=	0.9502119

		Actual α	=	0.0479962

Problem 4

Input:		Tail(s) =	One	
		Proportion p2	=	0.625
		α err prob	=	0.05
		Power (1-β err prob)	=	0.95
		Proportion p1	=	0.5
Output:	Lower critical N =		98.0000000	
		Upper critical N	=	98.0000000
		Total sample size	=	173
		Actual power	=	0.9514084
		Actual α	=	0.0470574

Problem 5

Input:		Tail(s) =	One	
		Proportion p2	=	0.585
		α err prob	=	0.05
		Power (1-β err prob)	=	0.95
		Proportion p1	=	0.5
Output:	Lower critical N =		203	
		Upper critical N	=	203
		Total sample size	=	373
		Actual power	=	0.9501117
		Actual α	=	0.0487048

Problem 6

Input:		Tail(s) =	One	
		Proportion p2	=	0.585
		α err prob	=	0.05
		Power (1-β err prob)	=	0.95
		Proportion p1	=	0.5
Output:	Lower critical N =		203	
		Upper critical N	=	203
		Total sample size	=	373
		Actual power	=	0.9501117
		Actual α	=	0.0487048

Problem 7

Input:		Tail(s) =	One	
		Proportion p2	=	0.565
		α err prob	=	0.05
		Power (1-β err prob)	=	0.95
		Proportion p1	=	0.5
Output:	Lower critical N =		343	
		Upper critical N	=	343
		Total sample size	=	643
		Actual power	=	0.9506907

Actual α = 0.0487908

Problem 8

Input: Tail(s) = One
Proportion p2 = 0.585
α err prob = 0.05
Power (1-β err prob) = 0.95
Proportion p1 = 0.5

Output: Lower critical N = 203
Upper critical N = 203
Total sample size = 373
Actual power = 0.9501117
Actual α = 0.0487048

Hypothesis 2

Problem 1

Input: Tail(s) = One
Proportion p2 = 0.45
Proportion p1 = 0.64
α err prob = 0.05
Power (1-β err prob) = 0.95
Allocation ratio N2/N1 = 1

Output: Critical z = -1.6448536
Sample size group 1 = 146
Sample size group 2 = 146
Total sample size = 292
Actual power = 0.9500476

Problem 2

Input: Tail(s) = One
Proportion p2 = 0.52
Proportion p1 = 0.67
α err prob = 0.05
Power (1-β err prob) = 0.95
Allocation ratio N2/N1 = 1

Output: Critical z = -1.6448536
Sample size group 1 = 230
Sample size group 2 = 230
Total sample size = 460
Actual power = 0.9506681

Problem 3

Input: Tail(s) = One
Proportion p2 = 0.65
Proportion p1 = 0.75
α err prob = 0.05
Power (1-β err prob) = 0.95

		Allocation ratio N2/N1	=	1
Output:	Critical z	=	-1.6448536	
		Sample size group 1	=	452
		Sample size group 2	=	452
		Total sample size	=	904
		Actual power	=	0.9500671

Problem 4

Input:		Tail(s)	=	One
		Proportion p2	=	0.50
		Proportion p1	=	0.75
		α err prob	=	0.05
		Power (1-β err prob)	=	0.95
		Allocation ratio N2/N1	=	1
Output:	Critical z	=	-1.6448536	
		Sample size group 1	=	79
		Sample size group 2	=	79
		Total sample size	=	158
		Actual power	=	0.9512237

Problem 5

Input:		Tail(s)	=	One
		Proportion p2	=	0.60
		Proportion p1	=	0.77
		α err prob	=	0.05
		Power (1-β err prob)	=	0.95
		Allocation ratio N2/N1	=	1
Output:	Critical z	=	-1.6448536	
		Sample size group 1	=	159
		Sample size group 2	=	159
		Total sample size	=	318
		Actual power	=	0.9501233

Problem 6

Input:		Tail(s)	=	One
		Proportion p2	=	0.55
		Proportion p1	=	0.72
		α err prob	=	0.05
		Power (1-β err prob)	=	0.95
		Allocation ratio N2/N1	=	1
Output:	Critical z	=	-1.6448536	
		Sample size group 1	=	171
		Sample size group 2	=	171
		Total sample size	=	342
		Actual power	=	0.9501308

Problem 7

Input:		Tail(s)	=	One

	Proportion p2	=	0.08	
	Proportion p1	=	0.21	
	α err prob	=	0.05	
	Power (1-β err prob)	=	0.95	
	Allocation ratio N2/N1	=	1	
Output:	Critical z	=	-1.6448536	
	Sample size group 1	=	157	
	Sample size group 2	=	157	
	Total sample size	=	314	
	Actual power	=	0.9510194	

<u>Problem 8</u>

Input:	Tail(s)	=	One
	Proportion p2	=	0.44
	Proportion p1	=	0.61
	α err prob	=	0.05
	Power (1-β err prob)	=	0.95
	Allocation ratio N2/N1	=	1
Output:	Critical z	=	-1.6448536
	Sample size group 1	=	185
	Sample size group 2	=	185
	Total sample size	=	370
	Actual power	=	0.9508737