Replication: Revisiting Tversky and Shafir's (1992) Disjunction Effect with extension comparing between and within subject designs

*Ignazio Ziano
Grenoble Ecole de Management, F-38000 Grenoble (France)
Ignazio.ZIANO@grenoble-em.com

*Man Fai Kong, *Hong Joo Kim, *Chit Yu Liu, *Sze Chai Wong
Department of Psychology, University of Hong Kong, Hong Kong SAR
manfaikong@gmail.com ;khj114324@gmail.com; lchityu@gmail.com; nicolewong2210123@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 Journal of Economic Psychology

Accepted for publication on December 17, 2020

*Contributed equally, joint first authors

^Corresponding author: Gilad Feldman

Word: abstract – 140; manuscript - 3754 (without tables or figures)

Corresponding author

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

Author bios:

Ignazio Ziano is an assistant professor with the Grenoble Ecole de Management marketing department, F-38000 Grenoble (France). His research focuses on judgment and decision-making and consumer behavior.

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

Man Fai Kong, Hong Joo Kim, Chit Yu Liu, Sze Chai Wong were students at the University of Hong Kong during academic year 2018.

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

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:
This research was supported by the *European Association for Social Psychology seedcorn grant*.

Authorship declaration:

Gilad led the reported replication effort in advanced social psychology and judgment and decision-making courses (PSYC2071/3052). Gilad supervised each step in the project, conducted the pre-registration, and ran data collection. Ignazio reanalyzed and validated all findings, added additional analyses and reports, and integrated all reports into a manuscript. Ignazio and Gilad jointly finalized the manuscript for submission.

Man Fai Kong, Hong Joo Kim, Chit Yu Liu, Sze Chai Wong designed the replication, wrote the pre-registrations, analyzed the findings and wrote an initial report of the findings as part of their course. Man Fai Kong designed and initiated the between-within design contrast extension.

Bo Ley Cheng guided and assisted the replication effort.

Abstract

Does uncertainty about an outcome influence decisions? The sure-thing principle (Savage, 1954) posits that it should not, but Tversky and Shafir (1992) found that people regularly violate it in hypothetical gambling and vacation decisions, a phenomenon they termed "disjunction effect". Very close replications and extensions of Tversky and Shafir (1992) were conducted in this paper (N = 890, MTurk). The target article demonstrated the effect using two paradigms in a between-subject design: here, an extension also testing a within-subject design, with design being randomly assigned was added. These results were consistent with the original findings for the "paying to know" problem (original: Cramer's V = .22, 95% (CI) [.14, .32]; replication: Cramer's V = .30, 95% CI [.24, .37]), yet not for the "choice under risk" problem (original: Cramer's V = .26, 95% CI [.14, .39]; replication: Cramer's V = .11, 95% CI [-.07, .20]). The within-subject extension showed very similar results. Implications for the disjunction effect and judgment and decision-making theory are discussed, and a call for improvements on the statistical understanding of comparisons of between-subject and within-subject designs is introduced. All materials, data, and code are available on *https://osf.io/gu58m/.*

Keywords: Disjunction effect; replication; judgment and decision-making; uncertainty; risk; between versus within subject design

Replication: Revisiting Tversky and Shafir's (1992) Disjunction Effect with extension comparing between and within subject designs

The sure-thing principle (STP; Savage, 1954) is an axiom of rational choice theory. It posits that if decision-makers are willing to make the same decision regardless of whether an external event happens or not, then decision-makers should also be willing to make the same decision when the outcome of the event is uncertain. Tversky and Shafir (1992), however, found that people regularly violate the STP. In a “paying-to-know” paradigm they found that participants were willing to pay a small fee to postpone a decision about a vacation package promotion when outcome of an exam was uncertain, despite preferences to purchase the package regardless of exam outcome. Using a “choice under risk” problem, they found that facing uncertainty about the outcome of an initial bet led to less willingness to again accept the exact same bet, compared to when having learned the outcome of the first bet.

Tversky and Shafir (1992) attributed this effect – coined “disjunction effect” – to the relative ease of coming up with reasons for making definitive choices that definitive outcomes provide, compared to uncertain ones. They argued the following: when people envision that they have passed an exam, they could easily come up with reasons to go on vacation (“let’s celebrate!”); when people envision they have failed an exam, they could easily find opposite reasons to go on vacation (“let’s live a little!”); yet, an uncertain outcome does not elicit good reasons to make a definitive decision.

Chosen target for replication: Tversky and Shafir (1992)

We chose Tversky and Shafir (1992) due to the impact the article has had, the lack of direct close replications, and open questions regarding the findings (Coles, Tiokhin, Scheel, Isager, & Lakens, 2018; Isager, 2019; Lambdin & Burdsal, 2007; Li, Jiang, Dunn, & Wang, 2012). We identified several potential contributions and clarifications that could be achieved by revisiting this classic, and we discuss those further below.

Impact wise, the article has been highly influential across disciplines because it provided a new model of decision-makers, one that is based on rationalization and not on expected value. At the time of writing, the article has been cited 664 times according to Google Scholar. Furthermore, highly influential theoretical papers of decision-making in psychology (Shafir, Simonson, & Tversky, 1993), marketing (Simonson & Tversky, 1992) and management (Tversky & Simonson, 1993) were directly based on this empirical finding.

Tversky and Shafir claimed support for the disjunction effect in both “choice under risk” and “paying to know" paradigms, and for these to hold for both between-subject and within-subject experimental designs. Tversky and Shafir did not report any inferential statistics in their paper, limiting the discussion of their results to descriptives.

The “choice under risk” results are not without controversy. Kühberger et al. (2001) failed to replicate the “choice under risk” problem four times, and Lambdin and Burdsal (2007) also failed to find support for a disjunction effect (as conceptualized by the original authors). However, it may be that neither replication team had sufficient power to detect a disjunction effect in two-step gambles. Moreover, Li et al. (2012) found support for the disjunction effect in a conceptual replication involving a World Cup scenario, and mixed support for the disjunction effect in a variation of the two-steps gambles problem. Further, there are no known direct replications of the “paying to know” problem. Given the paper's influence across fields and the controversy surrounding the findings, we decided to attempt a pre-registered well-powered replication using a between-subject design resembling the original study. We summarized our review of the current findings in the literature in Tables 1 and 2.

Extension: Testing both between-subject and within-subject designs

We decided to also test the robustness of the disjunction effect by conducting an extension, adding a conceptual replication of both the "choice under risk" and the "paying to know" paradigms in a within-subject design (joint evaluation), in which all participants are exposed to all experimental conditions. There is some evidence that people make different judgments and decisions when evaluating different options jointly compared to when they are in separate evaluation (Hsee, 1996). Such differences are interesting for both theoretical and practical reasons, as they highlight the “on-the-fly” nature of preference construction, and may give indications on how to construct choice menus in order to achieve desired goals (Sunstein, 2018). It is not entirely clear which problems in judgments and decision-making are affected by evaluation mode, and to what extent (Lambdin & Shaffer, 2009). Note that in the original paper, results were very similar and in support of the disjunction effect when using either within-subject or the between-subject experimental designs. This extension would therefore provide theoretically interesting insights into the nature of the disjunction effect and the impact of study design on a classic problem in judgment and decision-making.

Table 1

Descriptive and omnibus inferential statistics, across original studies and replications

	Paying to know							Choice under risk						
	N	Choice	Win	Loss	Uncertain	Inferential Statistics	ES [95% CI]	N	Choice	Win	Loss	Uncertain	Inferential Statistics	ES [95% CI]
Tversky and Shafir 1992, original (within-subject)	/	/	/	/	/	/	/	98	Accept (%)	68 (69%)	58 (59%)	35 (34%)		
	/	/	/	/	/	/	/		Reject (%)	30 (31%)	40 (41%)	63 (66%)		
Tversky and Shafir 1992, original (between-subject)	199	Buy (%)	36 (54%)	38 (57%)	21 (32%)	χ^2 (4) = 19.02, p < .001	Cramer's V = .218 [.137, .317]	213	Accept (%)		49 (69%)[1]	40 (57%)[1]	χ^2 (2) = 13.89, p < .001	Cramer's V = .255 [.144, .394]
		Not buy (%)	11 (16%)	8 (12%)	4 (7%)				Reject (%)		22 (31%)[1]	31 (43%)[1]		
		Pay $5 (%)	20 (30%)	21 (31%)	41 (61%)									
Tversky and Shafir 1992, modified gambles (between-subject)	/	/	/	/	/	/	/	171	Accept (%)	42 (73%)[1]	39 (69%)[1]	43 (75%)[1]	χ^2 (2) = .76, p = .68	Cramer's V = .067 [-.108, .218]
	/	/	/	/	/	/	/		Reject (%)	15 (27%)[1]	18 (31%)[1]	14 (25%)[1]		
Kühberger et al 2001, exp. 1 (between-subject)	/	/	/	/	/	/	/	177	Accept (%)	(60%)[2]	(47%)[2]	(47%)[2]	…	…

	Paying to know							Choice under risk						
	N	Choice	Win	Loss	Uncertain	Inferential Statistics	ES [95% CI]	N	Choice	Win	Loss	Uncertain	Inferential Statistics	ES [95% CI]
	/	/	/	/	/	/	/		Reject (%)	(40%)[2]	(53%)[2]	(53%)[2]	…	…
Kühberger et al 2001, exp. 2 (between-subject)	/	/	/	/	/	/	/	184	Accept (%)	(83%)[2]	(70%)[2]	(62%)[2]	…	…
	/	/	/	/	/	/	/		Reject (%)	(17%)[2]	(30%)[2]	(38%)[2]	…	…
Kühberger et al 2001, exp. 3 (within-subject)	/	/	/	/	/	/	/	35	Accept (%)	28 (80%)[1]	13 (37%)[1]	15 (43%)[1]	…	…
	/	/	/	/	/	/	/		Reject (%)	7 (20%)[1]	22 (63%)[1]	20 (57%)[1]	…	…
Kühberger et al 2001, exp. 4 (between-subject)	/	/	/	/	/	/	/	97	Accept (%)	(68%)[2]	(32%)[2]	(38%)[2]	…	…
	/	/	/	/	/	/	/		Reject (%)	(32%)[2]	(68%)[2]	(62%)[2]	…	…
Lambdin and Burdsal, 2007 (within-subject)	/	/	/	/	/	/	/	55	Accept (%)	35 (64%)	26 (47%)	21 (38%)	…	…
	/	/	/	/	/	/	/		Reject (%)	20 (36%)	31 (53%)	34 (62%)	…	…
Present work (within-subject)	445	Buy (%)	256 (58%)	127 (29%)	99 (22%)	Friedman χ^2 (2) = 132.678, $p < .001$	†	445	Accept (%)	164 (37%)	187 (42%)	165 (37%)	Cochran's Q (2) = 4.63, $p = .099$	†

	Paying to know							Choice under risk						
	N	Choice	Win	Loss	Uncertain	Inferential Statistics	ES [95% CI]	N	Choice	Win	Loss	Uncertain	Inferential Statistics	ES [95% CI]
		Not buy (%)	97 (22%)	247 (56%)	168 (38%)				Reject (%)	281 (63%)	258 (58%)	280 (63%)		
		Pay $5 (%)	92 (21%)	71 (16%)	178 (40%)									
Present work (between-subject)	445	Buy (%)	58 (39%)	61 (42%)	25 (16%)	χ^2 (4) = 81.00, *p* < .001	Cramer's V = .302 [.239, .368]	445	Accept (%)	46 (31%)	56 (38%)	65 (44%)	χ^2 (2) = 4.99, *p* = .082	Cramer's V = .106 [-.067, .202]
		Not buy (%)	38 (26%)	61 (42%)	29 (19%)				Reject (%)	102 (69%)	92 (62%)	84 (56%)		
		Pay $5 (%)	52 (35%)	22 (16%)	99 (65%)									

[1] reconstructed cell Ns
[2] impossible to recover cell N because no cell size is specified
- - - impossible to calculate without original data
†no appropriate omnibus effect size
/ absent

Table 2

Comparison of differences across conditions

	Paying to know, , difference in % Pay $5 across conditions				Choice under risk, difference in % Accept across conditions			
	N	Pass-Fail	Pass-Uncertain	Fail-Uncertain	N	Win-Loss	Win-:Uncertain	Loss-Uncertain
Tversky and Shafir 1992 (within-subjects)	/	/	/	/	98	10	35	25
Inferential statistics	/	/	/	/		…	…	…
Effect size [95% CI]	/	/	/	/		†	†	†
Tversky and Shafir 1992 (between-subjects)	199	-1	-31	-30	213	14	31	17
Inferential statistics		χ^2 (2) = .552, p = .759	χ^2 (2) = 14.437, p < .001	χ^2 (2) = 12.676, p = .001		χ^2 (1) = 1.927, p = .165	χ^2 (1) = 12.484, p < .001	χ^2 (1) = 4.07, p = .04
Effect size [95% CI]		Cramer's V = .064 [-.122, .231]	Cramer's V = .329 [.188, .505]	Cramer's V = .308 [.171, .484]		Cramer's V = .131 [-.083, .307]	Cramer's V = .31 [.168, .482]	Cramer's V = .183 [.08, .357]
Tversky and Shafir 1992 , modified gambles (between-subjects)	/	/	/	/	171	-2	1	-4
Inferential statistics	/	/	/	/		χ^2 (1) = .171, p = .68	χ^2 (1) < .001, p > .99	χ^2 (1) = .391, p = .531
Effect size [95% CI]	/	/	/	/		Cramer's V = .058 [-.094, .258]	Cramer's V = .02 [-.093, .207]	Cramer's V = .078 [-.094, .278]
Kühberger et al 2001, exp. 1 (between-subject)	/	/	/	/	177	13	13	0
Inferential statistics	/	/	/	/		$\chi^2 < 2.14$, $p > .14$	$\chi^2 < 2.14$, $p > .14$	$\chi^2 < 2.14$, $p > .14$
Effect size [95% CI]	/	/	/	/		…	…	…
Kühberger et al 2001, exp. 2 (between-subject)	/	/	/	/	171	18	26	6

	Paying to know, , difference in % Pay $5 across conditions				Choice under risk, difference in % Accept across conditions			
	N	Pass-Fail	Pass-Uncertain	Fail-Uncertain	N	Win-Loss	Win-:Uncertain	Loss-Uncertain
Inferential statistics	/	/	/	/		χ^2 (1) = 2.76, p = .10	χ^2 (1) = 6.50, p = .01	χ^2 (1) = 0.88, p = .35
Effect size [95% CI]	/	/	/	/		…	…	…
Kühberger et al 2001, exp. 3 (within-subject)	/	/	/	/	184	44	39	5
Inferential statistics	/	/	/	/		p < .001	p < .001	p = .73
Effect size [95% CI]	/	/	/	/				
Kühberger et al 2001, exp. 4 (between-subject)	/	/	/	/	97	35	30	5
Inferential statistics	/	/	/	/		χ^2 (1) = 8.02, p = .005	χ^2 (1) = 6.24, p = .01	χ^2 (1) = 0.19, p = .66
Effect size [95% CI]	/	/	/	/		…	…	…
Lambdin and Burdsal, 2007 (within-subject)	35	17	26	9		/	/	/
Inferential statistics		…	…	…		/	/	/
Effect size [95% CI]		…	…	…		/	/	/
Present work (within-subject)	445	-5	-19	-24	445	-5	0	-5
Inferential statistics		χ^2 (3) = 138.38, p < .001	χ^2 (3) = 152.08, p < .001	χ^2 (3) = 85.72, p < .001		χ^2 (1) = 2.989, p = .084	χ^2 (1) = .007, p = .936	χ^2 (1) = 4.481, p = .034
Effect size [95% CI]		†	†	†		†	†	†
Present work (between-subject)	445	-20	-30	-50	445	-7	-13	-6
Inferential statistics		χ^2 (2) = 17.53, p < .001	χ^2 (2) = 28.88, p < .001	χ^2 (2) = 75.24, p < .001		χ^2 (1) = 1.496, p = .221	χ^2 (1) = 4.991, p = .025	χ^2 (1) = 1.03, p = .31
Effect size [95% CI]		Cramer's V = .245 [.146, .363]	Cramer's V = .31 [.207, .426]	Cramer's V = .503 [.394, .619]		Cramer's V = .071 [-.058, .194]	Cramer's V = .13 [-.058, .25]	Cramer's V = .059 [-.058, .182]

†no appropriate omnibus effect size

/absent

… impossible to recalculate from original paper

Method

Pre-registrations and open data

We first pre-registered the experiment on the Open Science Framework (OSF) and data collection was launched later that week. Pre-registrations, disclosures, power analyses, and all materials are available in the supplementary materials. These together with datasets and code were made available on the OSF at https://osf.io/gu58m/. All measures, manipulations, and exclusions for this investigation are reported, and data collection was completed before analyses. Pre-registrations are available on the OSF: https://osf.io/fzchj.

Procedure and participants

We recruited a total of 890 participants from Mechanical Turk (405 males, 483 females, 2 other/prefer not to disclose, M_{age} = 40, SD_{age} = 11.35), who were paid $1.38 for this task, administered as part of a multi-study replication effort. We ran the replications both using a between-subject design as in the original paper, and using a within-subject design, randomly assigned. Specifically, half of participants completed the “choice under risk” problem between-subject and the “paying to know” problem within-subject; the other half completed the “paying to know” problem between-subject and the “choice under risk” problem within-subject.

In the between-subject replication of choice under risk and the within-subject replication of “paying to know”, 445 participants (194 male, 250 female, 1 other/would rather not disclose, M_{age} = 39.2, SD_{age} = 11.32) were randomly assigned to one of the three conditions of the “choice under risk” scenario (Win, Loss, or Uncertain) and all conditions in the “paying to know” scenario (Pass, Fail, Uncertain) presented in randomized order.

In the within-subject replication of “choice under risk” and the between-subject replication of “paying to know”, 445 participants (211 males, 233 females, 1 other/would rather not disclose, M_{age} = 40.1, SD_{age} = 11.38) were randomly assigned to one of the three conditions of the “paying to know” scenario (Pass, Fail, Uncertain) and all conditions in the “choice under risk” scenario (Win, Loss, or Uncertain) presented in randomized order.

We employed two checks, which indicated that participants were very attentive (Table 3). Following our pre-registered plan, we report analyses below based on data from all participants, maximizing statistical power.

Table 3

Attention check results

Response alternative	“Never answer scales in online studies seriously”*		“Always carefully read and answer each item on online surveys”**	
	Counts	% of total	Counts	% of total
1 (*Not at all characteristic of me*)	834	93.7 %	1	0.1 %
2 (*A little characteristic of me*)	19	2.1 %	9	1.0 %
3 (*Somewhat characteristic of me*)	19	2.1 %	19	2.1 %
4 (*Very characteristic of me*)	14	1.6 %	81	9.1 %
5 (*Entirely characteristic of me*)	4	0.4 %	780	87.6 %

*M = 1.13; SD = 0.55 (here, lower numbers indicate higher attentiveness)
**M = 4.83, SD = 0.51 (here, higher numbers indicate higher attentiveness)

How to analyze the disjunction effect?

Lambdin and Burdsal (2007) argued that disjunction effects can only be observed using within-subject designs, i.e., by observing how participants change their choice of a bet or of a vacation in uncertain situations compared to certain situations, and then classifying them as displaying a disjunction effect. This approach certainly has merits, because of its granularity and precision. Our goal for this replication was to compare our findings with the original findings. Using Lambdin and Burdsal (2007) approach is unfeasible, as it would require the original data and to limit the comparison to only a within-subject design. Further, using Lambdin and Burdsal (2007)'s method is uninformative for our goals, as Tversky and Shafir (1992) measured the disjunction effect at the group level in between-subjects studies, and at the condition level in within-subjects. For both these reasons (unfeasibility and impossibility of comparison), we decided to compare group proportions as in the original paper.

Scenarios

"Paying to know"

In the "paying to know" paradigm, participants read the following scenarios (differences between the scenarios are underlined):

> [Pass/Fail Version]
>
> "Imagine that you have just taken a tough qualifying examination. It is the end of the semester, you feel tired and run-down, and you find out that you [passed the exam / failed the exam. You will have to take it again in a couple of months—after the Christmas holidays.]
>
> You now have an opportunity to buy a very attractive 5-day Christmas vacation package to Hawaii at an exceptionally low price. The special offer expires tomorrow.
>
> [Uncertain Version]

> "Imagine that you have just taken a tough qualifying examination. It is the end of the fall quarter, you feel tired and run-down, and you are not sure that you passed the exam. In case you failed you have to take the exam again in a couple of months—after the Christmas holidays.
>
> You now have an opportunity to buy a very attractive 5-day Christmas vacation package to Hawaii at an exceptionally low price. The special offer expires tomorrow, while the exam grade will not be available until the following day.

Once presented with a scenario, participants had to make a choice between three options: 1) "I would buy the vacation package", 2) "I would not buy the vacation package", and 3) "I would pay a $5 nonrefundable fee in order to retain the rights to buy the vacation package at the same exceptional price the day after tomorrow".

"Choice under risk"

In the "choice under risk" scenario, participants were assigned to one of the following scenarios:

> [Win/Loss version]
>
> "Imagine that you have just played a game of chance that gave you a 50% chance to win $200 and a 50% chance to lose $100. The coin was tossed and you have [won $200 / lost $100].
>
> You are now offered a second identical gamble:
>
> 50% chance to win $200 and
>
> 50% chance to lose $100
>
> [Uncertain version]
>
> "Imagine that the coin has already been tossed, but that you will not know whether you have won $200 or lost $100 until you make your decision concerning a second, identical gamble:
>
> 50% chance to win $200 and
>
> 50% chance to lose $100

Once presented with a scenario, participants then indicated whether they would accept or reject the second bet.

Clarifications about effect sizes

Across between-subject scenarios, we used Cramer's V as a standardized effect size. However, Cramer's V is bounded at 0 and 1. One could therefore find similar Cramer's V in two studies, but a completely different pattern of results. Further, the calculation of 95% CIs around Cramer's V is problematic for the same reason. We calculated 95% CIs with the R package DescTools (Signorell, 2016) that provides with negative pseudo-lower bounds. Finally, Cramer's V cannot be used for within-subject designs. We chose to include it to give a broader indication of an unstandardized effect size, but given these limitations, we caution against the over-reliance on Cramer's V and instead invite the reader to give more weight to descriptive statistics.

Results

Descriptives and inferential statistics are provided in Tables 1 and 2, and findings are plotted in Figure 1.

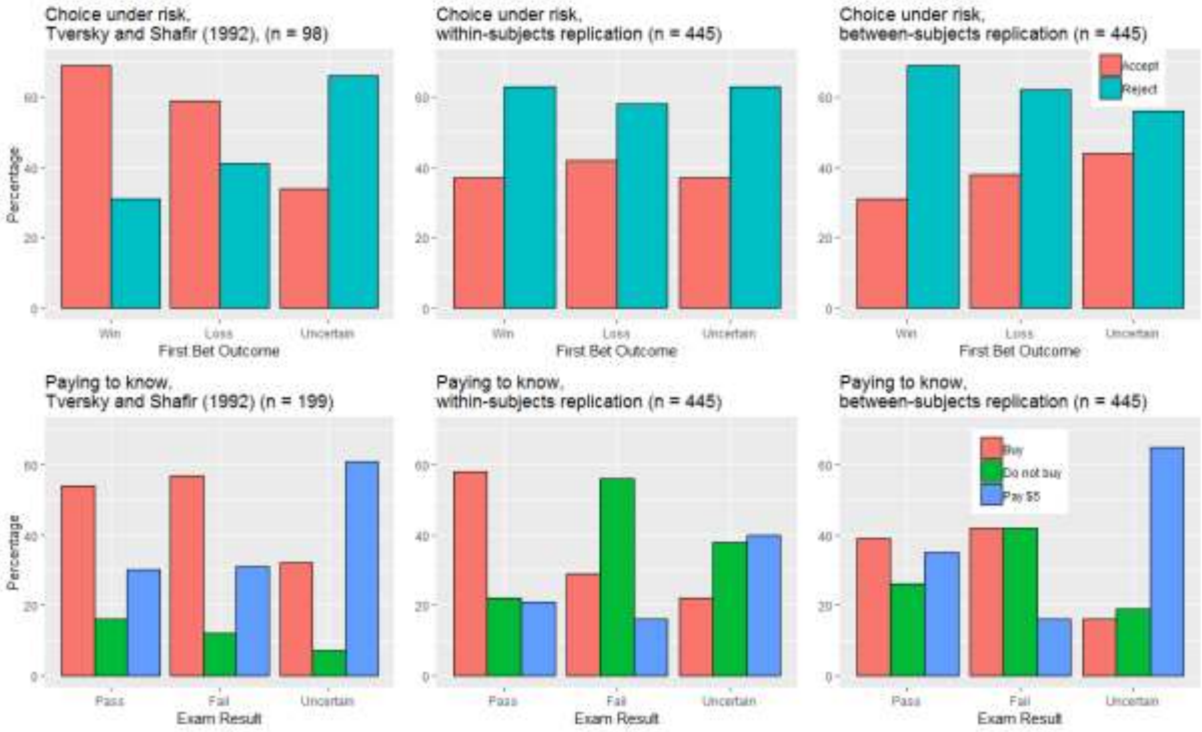


Figure 1. Tversky and Shafir (1992) original studies' results and present replications results

"Paying to know"

Between-subject design replication

In the Fail condition, only 22/144 (15%) participants chose to pay the $5 to reserve the vacation price, in the Pass condition, this proportion increased to 52/148 (35%), and in the Uncertain condition 99/153 (65%) participants indicated that they would pay the $5. This pattern was largely consistent with the original results, with a sharp increase in the proportion of participants choosing to pay $5 to reserve in the Uncertain condition compared to the Pass and the Fail conditions.

We conducted a test for equality of proportions and found support for an omnibus effect of condition on decision (χ^2 (4) = 81.00, p < .001, Cramer's V = .302, [.239, .368]). We proceeded to conduct three pairwise tests for equality of proportion. We found support for differences between the Pass and the Fail conditions (χ^2 (2) = 17.53, p < .001, Cramer's V =

.245, [.146, .363]), support for differences between the Fail and the Uncertain conditions (χ^2 (2) = 75.24, $p < .001$, Cramer's V = .503, [.394, .619]), and support for differences between the Pass and the Uncertain conditions (χ^2 (2) = 28.88, $p < .001$, Cramer's V = .31, [.207, .426]). Dwass-Steel-Critchlow-Fligner comparisons in a Kruskal-Wallis ANOVA, which control for multiple comparisons, showed no support for differences between the Pass and the Fail conditions (W = 3.153, $p = .066$), and support for differences between the Fail and the Uncertain conditions (W = 11.34, $p < .001$) and for the Pass and the Uncertain conditions (W = 7.58, $p < .001$).

Within-subject design replication

In the Fail condition, 71/445 (16%) participants chose to pay the $5 to reserve the vacation price, in the Pass condition this proportion increased to 92/445 (21%), and in the Uncertain condition 178/445 (40%) participants indicated that they would pay the $5. As in the between-subject replication, this pattern of results was consistent with original findings.

We conducted three pairwise multiple comparisons using McNemar's test for repeated measures. We found support for differences between the Pass and Fail conditions (χ^2 (3) = 138.38, $p < .001$), support for differences between the Fail and the Uncertain condition (χ^2 (3) = 85.72, $p < .001$), and support for difference between the Pass and the Uncertain conditions (χ^2 (3) = 152.08, $p < .001$). In a Friedman test and series of Durbin-Conover comparisons, which correct for multiple comparisons, we found support for an omnibus effect of condition (χ^2 (2) = 132.678, $p < .001$; Uncertain – Pass statistic = 12.436, $p < .001$; Uncertain – Fail statistic = 7.05, $p < .001$; Pass – Fail statistic = 5.386, $p < .001$).

"Paying to know" summary: Comparing between and within designs

Overall, in both the within-subject and the between-subject replications we found effects consistent with the original findings. We found an increase in the share of participants reporting that they would pay $5 to reserve the price of the vacation in the Uncertain condition, compared to the two other conditions. The share of participants who decided not to buy the vacation was higher across our replications in all conditions.

"Choice under risk "

Between-subject replication

In the "Win" condition, 46/148 (31%) participants chose to accept the gamble, in the "Loss" condition, 56/148 participants (38%) chose to accept the gamble, and in the "Uncertain" condition 65/149 (44%) participants chose to accept the gamble. This pattern was inconsistent with the original findings, and in direct contrast to original results. We expected the proportion of participants who chose to accept the bet to decrease in the Uncertain condition compared to the other two conditions, and yet we found that only a minority of participants accepted the bet across all conditions. We conducted a test of equality of proportion with condition (win, loss, uncertain) as the independent variable and choice (accept; reject) as the dependent variable and indeed failed to find support for the effect (χ^2 (2) = 4.99, p = .082, Cramer's V = .106, 95% CI [-.067, .202].)

We followed by conducting three pairwise tests for equality of proportions. We found support for differences between the Win and the Uncertain conditions (χ^2 (1) = 4.991, p = .025, Cramer's V = .13 [-.058, .25]), albeit in a direction opposite to the original findings. We found no support for differences between the Win and the Loss conditions (χ^2 (1) = 1.496, p = .221, Cramer's V = .071 [-.058, .194]) or for differences between the Loss and the Uncertain conditions (χ^2 (1) = 1.03, *p* = .31, Cramer's V = .059 [-.058, .182]). Dwass-Steel-Critchlow-Fligner comparisons in a Kruskal-Wallis ANOVA, which correct for multiple comparisons (Douglas & Michael, 2007), and again found no evidence for any differences between conditions (Loss-Win: W = 1.73, *p* = .441; Loss-Uncertain: W = -1.43, *p* = .569; Win - Uncertain: W = -3.15, *p* = .066).

Within-subject replication

In the Win condition 164/445 (37%) participants chose to accept the gamble, in the Loss condition 187/445 (42%) participants chose to accept the gamble, and in the Uncertain condition 165/445 (37%) chose to accept the gamble. Comparing the certain conditions (Win, Loss) with the Uncertain condition, we failed to find support for a disjunction effect. Again, as in the between-subject design findings, this pattern was not consistent with the original findings. Whereas original findings pointed to the majority of participants accepting the bet in both the Win and the Loss conditions, and a minority accepting the bet in the Uncertain condition, we found that the minority accepted the bet across all conditions.

We ran a Cochran test for equality of outcomes in a repeated-measures design and found no support for an effect (Cochran's Q (2) = 4.63, *p* = .099). We conducted three pairwise McNemar test for repeated-measures equality of proportions, and found no support for differences between the Win and the Loss conditions (χ^2 (1) = 2.989, *p* = .084), some support for differences between the Loss and the Uncertain condition (χ^2 (1) = 4.481, *p* = .034), and no support for differences between the Win and the Uncertain condition (χ^2 (1) = .007, *p* = .936). Similar results were obtained using the Durbin-Conover pairwise comparisons, which correct for multiple comparisons (Conover & Iman, 1979) (Uncertain – Win statistic = .083, *p* = .934; Uncertain – Loss statistic = 1.823, *p* = .069; Win – Loss statistic = 1.906, *p* = .057).

"Choice under risk" summary: Comparing between and within designs

In both replications using different designs only a minority of participants accepted the second bet, whereas in the original studies a majority of participants chose to accept the bet in the Win and the Loss conditions, but only a minority chose to accept it in the Uncertain condition.

General Discussion

We conducted a replication of disjunction effect (Tversky & Shafir, 1992), testing two paradigms. Our results were consistent with original findings for the "paying to know" paradigm, but inconsistent with a much weaker effect than original findings in the "choice under risk" paradigm. We ran each of the two paradigms using two designs, between-subject and within-subject, and results were very consistent across designs.

Replications results

Two and a half decades after the publication of the original findings, we were able to successfully replicate the findings regarding "paying to know" scenario, regardless of research design, showing support for the robustness and reliability of the disjunction effect. With that said, we identified a caveat in a failed replication for the "choice under risk" scenario. Moving forward, those who aim to study the disjunction effect further may want to base their follow-ups on what was successfully replicated, or to investigate factors that led to the differences between the two paradigms.

What may explain differences between original and replication? An immediate suspect is the sample, if of different demographics. The original experiment employed Stanford undergraduates, and we employed online MTurk samples, which have been shown reliable (Buhrmester, Kwang, & Gosling, 2011; Coppock, 2017; Coppock, Leeper, & Mullinix, 2018;

Zwaan et al., 2018), especially so in the domain of judgement and decision making replications, with replications from the economic psychology and judgment and decision-making yielding highly similar results even more than 20 years later (Chandrashekar et al., 2020; Ziano, Jie, et al., 2020; Ziano, Wang, et al., 2020; Ziano, Mok, & Feldman, 2020). Yet, we consider it unlikely that the sample is to blame for the failed replication of the "choice under risk" problem, when at the same time demonstrating a successful replication of the "paying to know" problem.

Second, some may argue that the passing of time may have affected replication results. The original studies were conducted on or before 1992. It is possible that the meaning of the "choice under risk" problem factors has changed during that time. Again, this account does not explain why the "paying to know" problem was successfully replicated. It is possible that the passing of time has affected the two problems differently, yet given the broad context-less descriptions of the gambles in that scenario, we find this argument unconvincing.

Third, it is possible that the "choice under risk" problem was a false-positive finding (given the smaller effect size we found compared to the original paper), whereas the "paying to know" problem was a true positive finding. We provide two arguments in support of this explanation. First, previous research failed to find a disjunction effect in two-steps gambles, using either a between-subject design or a within-subject design (Kühberger et al., 2001; Lambdin & Burdsal, 2007), or found mixed results in conceptual replications (Li et al., 2012). Second, Tversky and Shafir (1992) report two successful replications of the "choice under risk" problem (p. 307), yet also report that increasing the stakes in the initial gamble (but leaving the second gamble unchanged) led to no disjunction effect, presumably because additional gambles did not provide strong enough reasons since the stakes were lower in comparison to first one. Possibly, this account of Tversky and Shafir (1992)'s failed two-step gambles rerun with modified amounts may be an indication of their own first failed replication of the disjunction effect.

An additional possibility, suggested by a recent paper (Broekaert, Busemeyer, & Pothos, 2020), is that risk-aversion may moderate the extent to which people exhibit the disjunction effect, such that less risk-averse people do not exhibit the effect, and that a quantum-dynamic model can reconcile opposing results from the original paper and from unsuccessful replication. This investigation falls outside the purview of this paper, but it seems a potentially fruitful avenue for future research.

Overall, these results pose a challenge for research based on the disjunction effect. With inconsistent evidence for the two problems, which of the problems is to be associated with the disjunction effect? Though we now have fairly clear criteria for summarizing a replication for a single hypothesis with a single association between two variables, we still lack the criteria to evaluate complex replications with mixed findings, and then relate that back underlying theory. Further research is needed to disentangle when and why supposedly irrelevant uncertain outcomes cause preference reversals. There is also much need for to establishing clearer criteria in evaluating complex replication efforts, of multiple studies, multiple hypotheses, and multiple independent and dependent variables, all representing a single theory or article.

Comparing research designs

We found consistent results across within-subject and between-subject designs. We did not find larger differences in the within-subjects condition compared to the between-subjects condition in the Paying to Know scenario. While there was pattern of choices more pronounced and more similar to the original results in the Win and Loss condition for the within-subjects condition, there was a more pronounced pattern for the Uncertain condition in the between-subjects condition. Comparisons of evaluation modes is highly relevant for both theoretical and practical purposes, as it highlights the fickle nature of preferences and choices that people make in different situations (Sunstein, 2018). This is an important contribution, as there are conflicting findings in judgment and decision-making, some showing differences between joint evaluations (within-subject) and separate evaluations (between-subject) (e.g., Hsee, 1996; Hsee, Loewenstein, Blount, & Bazerman, 1999; Paharia, Kassam, Greene, & Bazerman, 2009) whereas others show effects robust to evaluation mode change (Lambdin & Shaffer, 2009; Ziano, Lembregts, & Pandelaere, 2019; Ziano & Pandelaere, 2020).

We identified a methods gap regarding comparisons of within- and between- subject experiments. Although there are methods for such comparisons for frequentist linear dependent variables (e.g., Sezer, Zhang, Gino, & Bazerman, 2016), methods are still lacking regarding similar analyses for binomial or multinomial dependent variables. This poses a challenge for comparisons of joint and separate evaluations from an inferential point of view (beyond descriptives in Tversky & Shafir, 1992), and it is a promising issue to tackle in future research.

References

Broekaert, J. B., Busemeyer, J. R., & Pothos, E. M. (2020). The Disjunction Effect in two-stage simulated gambles. An experimental study and comparison of a heuristic logistic, Markov and quantum-like model. *Cognitive Psychology*, *117*(November 2019), 101262. https://doi.org/10.1016/j.cogpsych.2019.101262

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science*, *6*(1), 3–5. https://doi.org/10.1177/1745691610393980

Chandrashekar, P., Yeung, W., Yau, K. C., Feldman, G., Yuen Ting Julie, M., Ho, J. C., … Chan, C. (2019). Agency and self-other asymmetries in perceived bias and shortcomings: Replications of the Bias Blind Spot and extensions linking to free will beliefs. *Preprint*, (March). https://doi.org/10.13140/RG.2.2.19878.16961

Conover, W. J., & Iman, R. L. (1979). On multiple-comparisons procedures. *Los Alamos Sci. Lab. Tech. Rep. LA-7677-MS*, 1–14.

Coppock, A. (2017). Generalizing from Survey Experiments Conducted on Mechanical Turk: A Replication Approach. *Political Science Research and Methods*. https://doi.org/http://alexandercoppock.com/papers/Coppock_generalizability.pdf

Coppock, A., Leeper, T. J., & Mullinix, K. J. (2018). Generalizability of heterogeneous treatment effect estimates across samples. *Proceedings of the National Academy of Sciences*, 201808083. https://doi.org/10.1073/pnas.1808083115

Douglas, C. E., & Michael, F. A. (2007). On distribution-free multiple comparisons in the one-way analysis of variance. *Communications in Statistics - Theory and Methods*, *20*(1), 127–139. https://doi.org/10.1080/03610929108830487

Hsee, C. K. (1996). The Evaluability Hypothesis: An Explanation for Preference Reversals between Joint and Separate Evaluations of Alternatives. *Organizational Behavior and Human Decision Processes*, *67*(3), 247–257. https://doi.org/10.1006/obhd.1996.0077

Hsee, C. K., Loewenstein, G. F., Blount, S., & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin*, *125*(5), 576–590. https://doi.org/10.1037/0033-2909.125.5.576

Kühberger, A., Komunska, D., & Perner, J. (2001). The Disjunction Effect: Does It Exist for Two-Step Gambles? *Organizational Behavior and Human Decision Processes*, *85*(2), 250–264. https://doi.org/10.1006/obhd.2000.2942

Lambdin, C., & Burdsal, C. (2007). The disjunction effect reexamined: Relevant methodological issues and the fallacy of unspecified percentage comparisons. *Organizational Behavior and Human Decision Processes*, *103*(2), 268–276. https://doi.org/10.1016/j.obhdp.2006.04.001

Lambdin, C., & Shaffer, V. a. (2009). Are within-subjects designs transparent ? *Judgment and Decision Making*, *4*(7), 554–566. https://doi.org/10.1037/e722352011-194

Li, S., Jiang, C., Dunn, J. C., & Wang, Z. (2012). A test of " reason-based " and " reluctance-to-think " accounts of the disjunction effect . A test of '" reason-based "' and '" reluctance-to-think "' accounts of the disjunction effect, (January). https://doi.org/10.1016/j.ins.2011.09.002

Paharia, N., Kassam, K. S., Greene, J. D., & Bazerman, M. H. (2009). Dirty work, clean hands: The moral psychology of indirect agency. *Organizational Behavior and Human*

Decision Processes, *109*(2), 134–141. https://doi.org/10.1016/j.obhdp.2009.03.002

Savage, L. J. (1954). *The foundations of statistics*. New York, New York, USA: Wiley.

Sezer, O., Zhang, T., Gino, F., & Bazerman, M. H. (2016). Overcoming the outcome bias: Making intentions matter. *Organizational Behavior and Human Decision Processes*, *137*, 13–26. https://doi.org/10.1016/j.obhdp.2016.07.001

Shafir, E., Simonson, I., & Tversky, A. (1993). Reason-based choice. *Cognition*, *49*(1), 11–36. https://doi.org/10.1016/0010-0277(93)90034-S

Signorell, A. (2016). DescTools: Tools for descriptive statistics. *R Package Version 0.99*, *18*.

Simonson, I., & Tversky, A. (1992). Choice in Context : Tradoff Contrast and Extremeness Aversion. *Journal of Marketing Research*, *29*(3), 281–295. https://doi.org/10.2307/3172740

Sunstein, C. R. (2018). On Preferring A to B, While Also Preferring B to A. *Rationality and Society*, *30*(3), 305–331.

Tversky, A., & Shafir, E. (1992). The disjunction effect in choice under uncertainty. *Psychological Science*, *3*(5), 305–309. Retrieved from https://0-journals-sagepub-com.wam.city.ac.uk/doi/pdf/10.1111/j.1467-9280.1992.tb00678.x

Tversky, A., & Simonson, I. (1993). Context- dependent Preferences. *Management Science*, *39*(10), 1179–1189. https://doi.org/10.1287/mnsc.39.10.1179

Ziano, I., Lembregts, C., & Pandelaere, M. (2019). *Perceived Income Inequality: Why Pay Ratios Are Less Effective Than Median Incomes*. Retrieved from https://psyarxiv.com/2zngf/

Ziano, I., & Pandelaere, M. (2020). *Late-Action Bias: Perceived Outcome Reversibility and Heightened Counterfactual Thinking Make Actions Closer to a Definitive Outcome Seem More Impactful*. https://doi.org/10.1017/CBO9781107415324.004

Ziano, I., Tsun, S. M., Lei, H. C., & Kamath, A. A. (2020). *Revisiting " Money Illusion ": Replication and Extension of Shafir et al . (1997)*. https://doi.org/10.13140/RG.2.2.10679.47525

Ziano, I., Wang, Y. J., Sany, S. S., Feldman, G., Ngai, L. H., Lau, Y. K., … Chan, C. (2020). Perceived morality of direct versus indirect harm : Replications of the preference for indirect harm effect In press at Meta Psychology . Accepted for publication on Jan 30 , 2020. *Meta-Psychology*. Retrieved from https://psyarxiv.com/bs7jf

Zwaan, R. A., Pecher, D., Paolacci, G., Bouwmeester, S., Verkoeijen, P., Dijkstra, K., & Zeelenberg, R. (2018). Participant Nonnaiveté and the reproducibility of cognitive psychology. *Psychonomic Bulletin and Review*, *25*(5), 1968–1972. https://doi.org/10.3758/s13423-017-1348-y

Tversky & Shafir (1992) replication & extension: Supplementary

Table of contents

Open Science

Data and code

Data and code are shared using the Open Science Framework: https://osf.io/gu58m/

Pre-registrations and Qualtrics study designs

Pre-registration available on: https://osf.io/fzchj

We note that several coauthored worked on this manuscript independently, peer reviewing one another, with each writing their own pre-registration, yet with one Qualtrics survey design. Pre-registrations were very similar, and we included all of those. We conducted our data analyses based on the most conservative/restrictive of those.

Procedure and data disclosures

Data collection

Data collection was completed before conducting an analysis of the data.

Conditions reporting

All collected conditions are reported.

Data exclusions

Details are reported in each of the two studies in the materials section of this document

Variables reporting

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

Disclosures and clarifications

Deviations from preregistration

Components in your preregistration	Were there deviations?	If yes - describe details of deviation(s)	Rationale for deviation	How might the results be different if you had not deviated
Within-subjects analysis of the Choice under risk problem	Additional analyses	Friedman omnibus test and Conover-Durbin comparisons	This analysis allows for an omnibus p-value and test statistics, and corrects for multiple comparisons	Lower p-value, since the preregistered pairwise comparisons do not correct for multiple comparisons
Within-subjects analysis of the Paying to know problem	Additional analyses	Friedman omnibus test and Conover-Durbin comparisons	This analysis allows for an omnibus p-value and test statistics, and corrects for multiple comparisons	Lower p-value, since the preregistered pairwise comparisons do not correct for multiple comparisons
Between-subjects analysis of the Choice under risk problem	Additional analyses	Dwass-Steel-Critchlow-Fligner comparison	These comparison correct for multiple comparison	Lower p-value, since the preregistered pairwise comparisons do not correct for multiple comparisons
Between-subjects analysis of the Paying to know problem	Additional analyses	Dwass-Steel-Critchlow-Fligner comparison	These comparison correct for multiple comparison	Lower p-value, since the preregistered pairwise comparisons do not correct for multiple comparisons

Classification of the replications presented here, based on LeBel et al. 2017

Problem	Design facet	Within-subjects replication	Between-subjects replication
Choice under risk	IV operationalization	same	same
	DV operationalization	same	same
	IV stimuli	same	same
	DV stimuli	same	same
	Procedural details	different	same
	Physical settings	different	different
	Contextual variables	different	different
	Replication classification	Close replication	Very close replication
Paying to know	IV operationalization	same	same
	DV operationalization	same	same
	IV stimuli	same	same
	DV stimuli	same	same
	Procedural details	different	same
	Physical settings	different	different
	Contextual variables	different	different
	Replication classification	Close replication	Very close replication

Comparison of procedural details between original and replication studies

	Choice under risk			Paying to know		
	Original	Replication between-subjects	Replication within-subjects	Original	Replication between-subjects	Replication within-subjects
Sample size	98	445	445	199	445	445
Geographic origin	US American	US American	US American	US American (Stanford undergraduate students)	US American	US American
Gender	Not reported	194 male, 250 female, 1 other	211 male, 233 female, 1 other	« about half male, half female »	211 male, 233 female, 1 other	194 male, 250 female, 1 other
Average age (years)	Not reported	39.2	40	Not reported	40	39.2
Medium (location)	Paper sheets (classroom)	Computer (online)	Computer (online)	Paper sheets (classroom)	Computer (online)	Computer (online)
Compensation	Not reported	Nominal payment	Nominal payment	Not reported	Nominal payment	Nominal payment
Year	Not reported	2018	2018	Not reported	2018	2018

Documenting Differences between the Original and Replication Study

For each part of the study indicate whether the replication study is Exact, Close, or Conceptually Different compared to the original study. Then, justify the rating.

<table>
<tr><td colspan="2">Rating</td><td>Original Study</td><td>Replication</td></tr>
<tr><td>17. instructions</td><td>Exact</td><td colspan="2">Experiment 1:
Decision to purchase a travel package
Item: (1) buy, (2) not buy or (3) pay $5 nonrefundable fee to retain the rights to purchase the vacation package
Experiment 2:
Decision to accept or reject the second game
Item: (1) accept, or (2) reject the second gamble.</td></tr>
<tr><td>18. measures</td><td>Exact</td><td colspan="2">Proportion of choice in two experiments</td></tr>
<tr><td rowspan="2">19. stimuli</td><td rowspan="2">Exact</td><td>Same questions on paper survey</td><td>Same questions on online survey</td></tr>
<tr><td colspan="2">English medium</td></tr>
<tr><td>20. procedure</td><td>Different</td><td>Experiment 1:
Between subject: three groups of were given three scenarios including the pass, fail and disjunctive scenario
Experiment 2:
Within subject:
The same group of participants were given the won, lost and disjunctive conditions concerning the first gamble to decide whether to accept or reject the second game.</td><td>Conducted on the Mechanical Turk, an online <u>Survey Research Center (https://psrc.princeton.edu/our-services/using-mturk) that</u> recruits participants for surveys and experiments mainly from the United States.

Combined survey:
1. Hindsight bias
2. Insensitivity to Sample Bias
3. Disjunction Effect
Experiment 1
Within subject: same group of participants in the same three conditions
Experiment 2:
Within subject:
Within subject: same group of participants in the same three conditions</td></tr>
<tr><td rowspan="2">21. location</td><td rowspan="2">Different</td><td>Stanford University</td><td>Mechanical Turk online</td></tr>
<tr><td colspan="2">Both recruited from the US</td></tr>
<tr><td>22. remuneration</td><td>Different</td><td>Voluntary</td><td>Minimum hourly pay in 2018 of the US counted in minutes for completing the survey</td></tr>
</table>

23. participant populations	**Different**	Experiment 1: 200 ; Experiment 2: 98 Equal number of females and males Age unreported	Experiment 1: 445 (Female: 250; Male: 194; Other:1) Experiment 2: 445 (Female: 233; Male: 211; Other:1) Age 20-75
24. Expected influence on size and/or direction of the effect		- More focused sample of undergraduate (limited age range in a university) and convenient sampling - No exclusion - Not randomized	- More diverse demographics - Exclusion based on self-reported English proficiency and seriousness in completing the survey
25. Steps taken to test whether differences listed above will influence the outcome of replication attempt		- Comparing the result between original study, previous studies and replication to address potential difference in presence and degree of effect with different demographics.	

Additional extension moved from the main manuscript

In our replication we also included an additional exploratory extension in our design. We failed to find any support for this extension, and as it secondary to our main replication and the design extension, we report the extension and the findings here in the supplementary.

Background

As an additional extension, before the problems were presented, participants completed a seven-item version of the Intolerance to Uncertainty Scale (IUS) (α = .867; Carleton, Norton, & Asmundson, 2007). IUS measures the extent to which people are averse to uncertainty, and we predicted that higher levels of IUS would be associated with a higher likelihood of rejecting the bet in the Choice under risk scenario and of selecting to pay $5 to reserve the price of the vacation package in the Paying to know scenario.

Results

Paying to know

Between-subject design

The Intolerance to Uncertainty Scale showed good reliability for the participants that completed the "paying to know" task within-subject (Cronbach's α = .856). In a multinomial regression analysis, we found no support for Intolerance to Uncertainty Scale being associated with participants' choices (χ^2 (2) = .889, p = .641). Contrast analysis showed no support for an association between IUS and comparison of participants' choice of "Buy" to "Do not Buy" (b(SE) = .021 (.023), 955 CI [-.024, .066], z = .928, p = .353) or for "Buy" against "Pay $5 to reserve the vacation price" (b (SE) = .007 (.022), 95% CI [-.035, .050], z = .336, p = .737).

Within-subject design

The Intolerance to Uncertainty Scale showed good reliability for these participants, Cronbach's α = .877. A multinomial repeated-measures regression using the GENLIN procedure in SPSS found no support for an association between IUS and choice (B (SE) = .003 (.011) ,Wald χ^2 (1) = .08, p = .778).

Choice under risk

Between -subject design

We ran a binomial regression analysis with the gamlj package on jamovi (Gallucci, 2018; The jamovi Project, 2019) and found no support for an association between uncertainty intolerance (α = .877) and participants' choice of accepting or rejecting the bet (B (SE) = .003, (.004), 95% CI [-.006, .011], Wald χ^2 (1) = .404, p = .525).

Within-subject design

We ran a repeated-measures logistic regression conducted with the GENLIN procedure on SPSS with uncertainty intolerance as the independent variable (α = .856) and choice as the dependent variable and found support for an effect (B (SE) = -.016 (.011), Wald χ^2 (1) = 1.425, p = .233).

Attention checks results

We report here the frequencies of the attention checks. Given the high rate of compliance, we decided to include all participants in the analyses in order to maximize statistical power.

- Q8. Never answer scales in online studies seriously (M = 1.13; SD = 0.55).

Levels	Counts	% of Total	Cumulative %
Not at all characteristic of me	834	93.7 %	93.7 %
A little characteristic of me	19	2.1 %	95.8 %
Somewhat characteristic of me	19	2.1 %	98.0 %
Very characteristic of me	14	1.6 %	99.6 %
Entirely characteristic of me	4	0.4 %	100.0 %

- Q9. Always carefully read and answer each item on online surveys (M = 4.83, SD = 0.51).

Levels	Counts	% of Total	Cumulative %
Not at all characteristic of me	1	0.1 %	0.1 %
A little characteristic of me	9	1.0 %	1.1 %
Somewhat characteristic of me	19	2.1 %	3.3 %
Very characteristic of me	81	9.1 %	12.4 %
Entirely characteristic of me	780	87.6 %	100.0 %

All materials/stimuli used in the experiment

For both scenarios:
Individual Difference Scale: Intolerance of Uncertainty Scale (7 items)
a. Instructions: Please select the answer choice that best corresponds to how much you agree with each item.

Questions:

- Q1. Unforeseen events upset me greatly.
- Q2. It frustrates me not having all the information I need.
- Q3. One should always look ahead so as to avoid surprises.
- Q4. A small, unforeseen event can spoil everything, even with the best of planning.
- Q5. I always want to know what the future has in store for me.
- Q6. I can’t stand being taken by surprise.
- Q7. I should be able to organize everything in advance.

Attention checks:

- Q8. Never answer scales in online studies seriously.
- Q9. Always carefully read and answer each item on online surveys.

Scales (1-5):

- 1 – Not at all characteristic of me
- 2 – A little characteristic of me
- 3 – Somewhat characteristic of me
- 4 – Very characteristic of me
- 5 – Entirely characteristic of me

Scenario 1 - Paying to Know

1. Condition 1: Pass exam

    a. Independent Variable:
    Imagine that you have just taken a tough qualifying examination. It is the end of the fall quarter, you feel tired and run-down, and you find out that you passed the exam. You now have an opportunity to buy a very attractive 5-day Christmas vacation package to Hawaii at an exceptionally low price. The special offer expires tomorrow, while the exam grade will not be available until the following day.
    b. Dependent Variable:
    Choose one of the following:
    x. buy the vacation package.
    y. not buy the vacation package.

z. pay a $5 nonrefundable fee in order to retain the rights to buy the vacation package at the same exceptional price the day after tomorrow.

2. Condition 2: Fail exam
    a. Independent Variable:
    Imagine that you have just taken a tough qualifying examination. It is the end of the fall quarter, you feel tired and run-down, and you find out that you failed the exam. You now have an opportunity to buy a very attractive 5-day Christmas vacation package to Hawaii at an exceptionally low price. The special offer expires tomorrow, while the exam grade will not be available until the following day.
    b. Dependent variable: same as condition 1

3. Condition 3: Disjunctive (Exam result unknown)
    a. Independent Variable:
    Imagine that you have just taken a tough qualifying examination. It is the end of the fall quarter, you feel tired and run-down, and you find out that you failed the exam. You now have an opportunity to buy a very attractive 5-day Christmas vacation package to Hawaii at an exceptionally low price. The special offer expires tomorrow, while the exam grade will not be available until the following day.
    b. Dependent variable: same as condition 1 and 2

Scenario 2 - Choice under Risk

1. Condition 1: Won 1st Game
    a. Independent Variable:
    Imagine that you have just played a game of chance that gave you a 50% chance to win $200 and a 50% chance to lose $100. The coin was tossed and you have won $200.
    You are now offered a second identical gamble:
    50% chance to win $200 and 50% to lose $100.
    b. Dependent variable:
    Would you:
    x. Accept the second gamble.
    y. Reject the second gamble.

2. Condition 2: Lost 1st Game
    a. Independent Variable:
    Imagine that you have just played a game of chance that gave you a 50% chance to win $200 and a 50% chance to lose $100. The coin was tossed and you have lost $100.

You are now offered a second identical gamble:
50% chance to win $200 and 50% to lose $100.

   b. Dependent variable: same as condition 1

3. Condition 3: Disjunctive (Game result unknown)
   a. Independent Variable:
      Imagine that you have just played a game of chance that gave you a 50% chance to win $200 and a 50% chance to lose $100. Imagine that the coin has already been tossed, but that you will not know whether you have won $200 or lost $100 until you make your decision concerning a second, identical gamble:
      50% chance to win $200 and 50% chance to lose $100.
   b. Dependent variable: same as condition 1 and 2

Effect size calculations of the original article

Chi Square Test, Calculation of Cramer's V of the original Study

Experiment 1: Paying to Know (Between Subject)

3 conditions: X^2 (4, N=200) = 17.18, p<0.002. V=0.207

Chi Square Test

Observed Values
Enter values in grey cells below:

		Conditions: Pass	Fail	Disjunctive	Row Totals
Options	Buy	36	38	21	95
	Not Buy	11	8	5	24
	Pay $5	20	21	40	81
	Column Totals	67	67	66	200 Grand Total

Expected Values

31.825	31.825	31.35
8.04	8.04	7.92
27.135	27.135	26.73

Test Results:

χ^2 =	17.1804
df =	4
p =	0.00178

Cramer's V

Calculation of Cramer's V

The total sample size is $n = 200$, and we have already computed the Chi-Square statistics, $\chi^2 = 17.18$. Let the number of rows be $r = 3$ and let $c = 3$ be the number of columns.

The Cramer's V statistic is computed as follows

$$V = \sqrt{\frac{\chi^2/n}{\min(c-1, r-1)}} = \sqrt{\frac{17.18/200}{\min(3-1, 3-1)}} = \sqrt{\frac{17.18/200}{2}} = 0.207$$

Pass & Disjunctive: X^2 (2, N=133) =12.86, p<0.002. V=0.311

Chi Square Test

Observed Values
Enter values in grey cells below:

	Condition: X	Y	Z	Row Totals
Option	36	11	20	67
	21	5	40	66
Column Totals	57	16	60	133 Grand Total

Expected Values

28.7143	8.06015	30.2256
28.2857	7.93985	29.7744

Test Results:

χ^2 =	12.8572
df =	2
p =	0.00161

Cramer's V

$$V = \sqrt{\frac{\chi^2/n}{\min(c-1, r-1)}} = \sqrt{\frac{12.857/133}{\min(3-1, 2-1)}} = \sqrt{\frac{12.857/133}{1}} = 0.311$$

Pass & Fail condition: X^2 (2, N=134) = 0.55, p > 0.5. V=0.064

Chi Square Test

Observed Values
Enter values in grey cells below:

Variable 1	Variable 2: X	Y	Row Totals
A	36	38	74
B	11	8	19
C	20	21	41
Column Totals	67	67	134 (Grand Total)

Expected Values

A	37	37
B	9.5	9.5
C	20.5	20.5

Test Results:
$\chi^2 =$ 0.55213
df = 2
$p =$ 0.75876

Cramer's V

$$V = \sqrt{\frac{\chi^2/n}{\min(c-1, r-1)}} = \sqrt{\frac{0.552/134}{\min(2-1, 3-1)}} = \sqrt{\frac{0.552/134}{1}} = 0.064$$

Fail & Disjunctive condition: X^2 (2, N=133) =11.50, p<0.004. V=0.294

Chi Square Test

Observed Values
Enter values in grey cells below:

Variable 1	Variable 2: X	Y	Row Totals
A	38	21	59
B	8	5	13
C	21	40	61
Column Totals	67	66	133 (Grand Total)

Expected Values

A	29.7218	29.2782
B	6.54887	6.45113
C	30.7293	30.2707

Test Results:
$\chi^2 =$ 11.5018
df = 2
$p =$ 0.00318

Cramer's V

$$V = \sqrt{\frac{\chi^2/n}{\min(c-1, r-1)}} = \sqrt{\frac{11.502/133}{\min(3-1, 2-1)}} = \sqrt{\frac{11.502/133}{1}} = 0.294$$

Chi Square Test, Calculation of Cohen's w and odds ratio of the original Study

Experiment 2: Choice under Risk (Within Subject)

3 conditions: X^2 (2, N=294) = 23.59, p<0.001. w=0.283,

Chi Square Test

Chi-Square Calculator

Success! The contingency table below provides the following information: the observed cell totals, (the expected cell totals) and [the chi-square statistic for each cell].

The chi-square statistic, *p*-value and statement of significance appear beneath the table. Blue means you're dealing with dependent variables; red, independent.

Results						
	accept	reject				*Row Totals*
won	68 (53.67) [3.83]	30 (44.33) [4.63]				98
lost	58 (53.67) [0.35]	40 (44.33) [0.42]				98
disj	35 (53.67) [6.49]	63 (44.33) [7.86]				98
Column Totals	161	133				294 (Grand Total)

The chi-square statistic is 23.5881. The *p*-value is < 0.00001. The result is significant at *p* < .05.

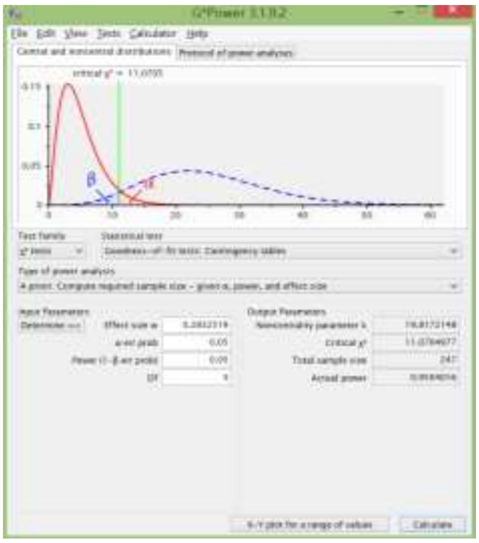
Cohen's w

```
>
> cohens_w(23.5881,294)
[1] 0.2832519
```

Power

χ² tests - Goodness-of-fit tests: Contingency tables
Analysis: A priori: Compute required sample size

Input:
Effect size w = 0.2832519
α err prob = 0.05
Power (1-β err prob) = 0.95
Df = 5
Output:
Noncentrality parameter λ = 19.8172148
Critical χ² = 11.0704977
Total sample size = 247
Actual power = 0.9504016



Won and Lost condition: X^2 (2, N=196) = 2.22, p=0.136. V=0.106

Chi Square Test

Observed Values
Enter values in grey cells below:

		Variable 2: X	Variable 2: Y	Row Totals
Variable 1	A	68	58	126
	B	30	40	70
Column Totals		98	98	196 (Grand Total)

Expected Values

63	63
35	35

Test Results:

χ^2 =	2.22222
df =	1
p =	0.13604

Cramer's V

$$V=\sqrt{\frac{\chi^2/n}{\min(c-1,r-1)}}=\sqrt{\frac{2.222/196}{\min(2-1,2-1)}}=\sqrt{\frac{2.222/196}{1}}=0.106$$

Won and Disjunctive condition: X^2 (2, N=196) = 22.28, p<0.001. V=0.235

Chi Square Test

Observed Values
Enter values in grey cells below:

		Variable 2: X	Variable 2: Y	Row Totals
Variable 1	A	68	35	103
	B	30	63	93
Column Totals		98	98	196 (Grand Total)

Expected Values

51.5	51.5
46.5	46.5

Test Results:

χ^2 =	22.2825
df =	1
p =	2.4E-06

Cramer's V

$$V=\sqrt{\frac{\chi^2/n}{\min(c-1,r-1)}}=\sqrt{\frac{10.824/196}{\min(2-1,2-1)}}=\sqrt{\frac{10.824/196}{1}}=0.235$$

Lost and Disjunctive condition: X^2 (2, N=196) = 10.82 , p=0.001. V=0.337

Chi Square Test

Observed Values
Enter values in grey cells below:

		Variable 2: X	Variable 2: Y	Row Totals
Variable 1	A	58	35	93
	B	40	63	103
Column Totals		98	98	196 (Grand Total)

Expected Values

46.5	46.5
51.5	51.5

Test Results:

χ^2 =	10.8241
df =	1
p =	0.001

Cramer's V

$$V=\sqrt{\frac{\chi^2/n}{\min(c-1,r-1)}}=\sqrt{\frac{22.282/196}{\min(2-1,2-1)}}=\sqrt{\frac{22.282/196}{1}}=0.337$$

Statistical Interpretation

Magnitude of Effect Size	Cramer's V/phi
Small	0.1
Medium	0.3
Large	0.5

Calculation tool
Jamovi
Chi-square
https://www.missouristate.edu/RStats/Tables-and-Calculators.htm
Cramer's V
https://mathcracker.com/cramers-v-calculator.php?)

Power analyses

First attempt (Initial calculation in pre-registration)

<u>Experiment 1: Paying to Know (Between subject)</u>

Chi-Square Calculator

Success! The contingency table below provides the following information: the observed cell totals, (the expected cell totals) and [the chi-square statistic for each cell].

The chi-square statistic, *p*-value and statement of significance appear beneath the table. Blue means you're dealing with dependent variables; red, independent.

Results						
	pass	fail	disjunctive			*Row Totals*
buy	36 (31.82) [0.55]	38 (31.82) [1.20]	21 (31.35) [3.42]			95
not buy	11 (8.04) [1.09]	8 (8.04) [0.00]	5 (7.92) [1.08]			24
pay	20 (27.14) [1.88]	21 (27.14) [1.39]	40 (26.73) [6.59]			81
Column Totals	67	67	66			200 (Grand Total)

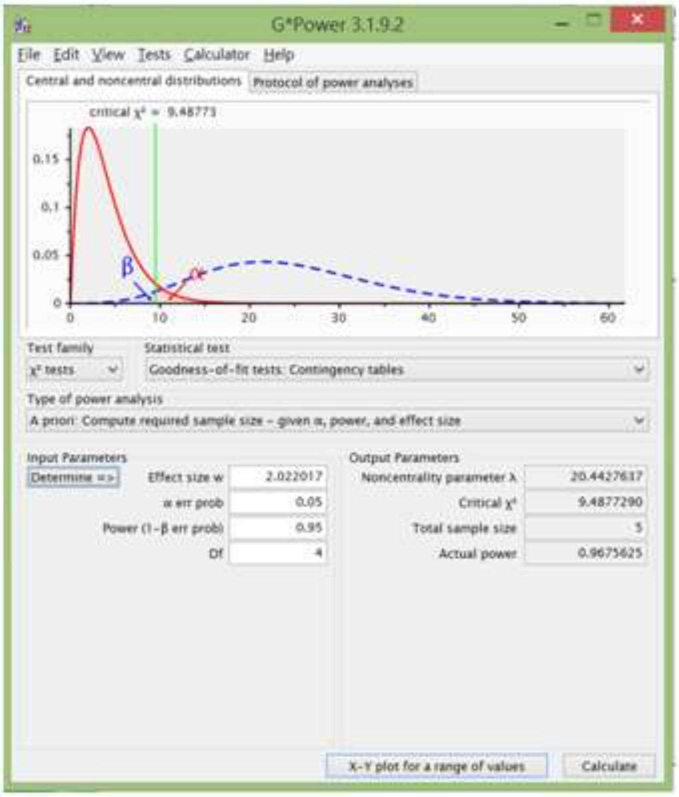
The chi-square statistic is 17.1804. The *p*-value is .001783. The result is significant at $p < .05$.

```
> pwr.chisq.test(w=ES.w2(PTK),df=(4),N=200)

     Chi squared power calculation

              w = 2.022017
              N = 200
             df = 4
      sig.level = 0.05
          power = 1

NOTE: N is the number of observations
```



Protocol of Power analysis

χ^2 tests - Goodness-of-fit tests: Contingency tables

Analysis: A priori: Compute required sample size

Input: Effect size w = 2.022017

α err prob = 0.05

Power (1-β err prob) = 0.95

Df = 4

Output: Noncentrality parameter λ = 20.4427637

Critical χ^2 = 9.4877290

Total sample size = 5

Actual power = 0.9675625

Experiment 2: Choice Under Risk (Within subject)

Chi-Square Calculator

Success! The contingency table below provides the following information: the observed cell totals, (the expected cell totals) and [the chi-square statistic for each cell].

The chi-square statistic, *p*-value and statement of significance appear beneath the table. Blue means you're dealing with dependent variables; red, independent.

Results						
	won	lost	disjunctive			*Row Totals*
accept	68 (53.67) [3.83]	58 (53.67) [0.35]	35 (53.67) [6.49]			161
reject	30 (44.33) [4.63]	40 (44.33) [0.42]	63 (44.33) [7.86]			133
Column Totals	98	98	98			294 (Grand Total)

Calculation suggested that the chi-square statistic is 23.5881. The *p*-value is < 0.00001. The result is significant at $p < .05$. The result indicated that there is a relationship between

> pwr.chisq.test(w=ES.w2(CUR),df=(2),N=98)

Chi squared power calculation

w = 2.019165

N = 98

df = 2

sig.level = 0.05

power = 1

NOTE: N is the number of observations

χ^2 tests - Goodness-of-fit tests: Contingency tables

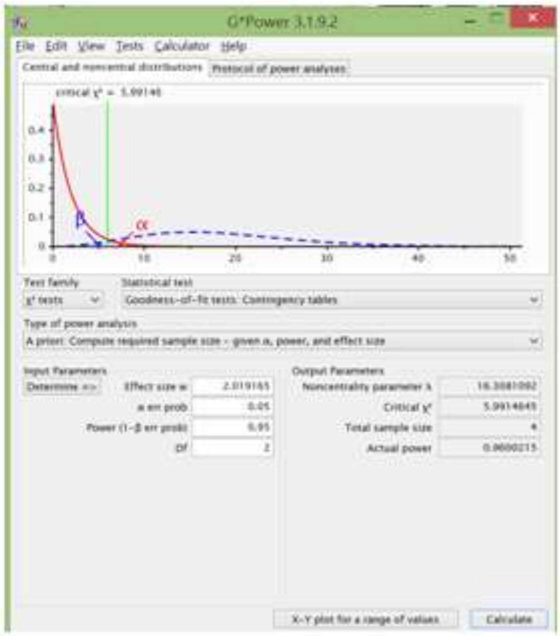
Analysis: A priori: Compute required sample size

Input:

Effect size w	=	2.019165
α err prob	=	0.05
Power (1-β err prob)	=	0.95
Df	=	2

Output:

Noncentrality parameter λ	=	16.3081092
Critical χ^2	=	5.9914645
Total sample size	=	4
Actual power	=	0.9600215



*Given the complexity of the design and the limitations of Cramer's V, a planned sample size of 2.5 of the original sample size was used with reference to Simonsohn's (2015) Small telescopes paper for the final pre-registration.

Second attempt (alternative approach to 2.5 times of the original sample)

** another approach using the Cohen's w and G*Power for Power analysis and planned sample size calculation was adopted with the aid of online tools and RStudio:

Chi-square calculator:

https://www.socscistatistics.com/tests/chisquare2/Default2.aspx

<u>Experiment 1: Paying to Know (Between Subject)</u>

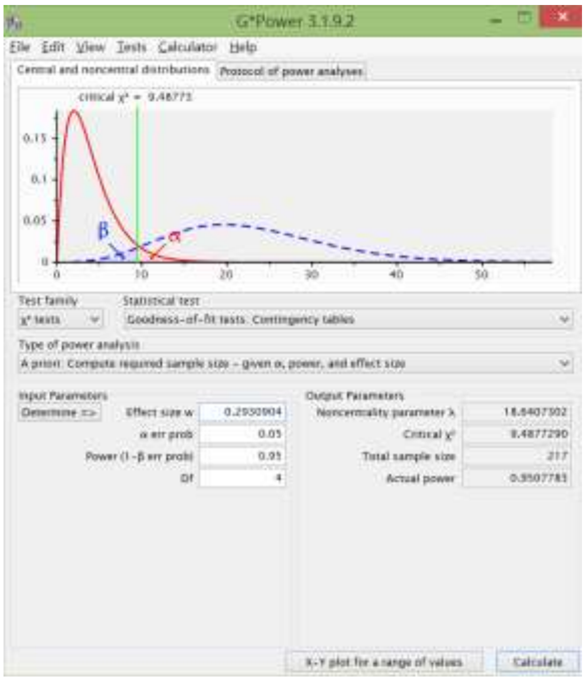
Chi-Square Calculator

Success! The contingency table below provides the following information: the observed cell totals, (the expected cell totals) and [the chi-square statistic for each cell].

The chi-square statistic, *p*-value and statement of significance appear beneath the table. Blue means you're dealing with dependent variables; red, independent.

Results						
	buy	not buy	pay $5			*Row Totals*
pass	36 (31.82) [0.55]	11 (8.04) [1.09]	20 (27.14) [1.88]			67
fail	38 (31.82) [1.20]	8 (8.04) [0.00]	21 (27.14) [1.39]			67
disjunctive	21 (31.35) [3.42]	5 (7.92) [1.08]	40 (26.73) [6.59]			66
Column Totals	95	24	81			200 (Grand Total)

The chi-square statistic is 17.1804. The *p*-value is .001783. The result is significant at $p < .05$.



χ^2 tests - Goodness-of-fit tests: Contingency tables

Analysis:	A priori: Compute required sample size		
Input:	Effect size w	=	0.2930904
	α err prob	=	0.05
	Power (1-β err prob)	=	0.95
	Df	=	4
Output:	Noncentrality parameter λ	=	18.6407302
	Critical χ^2	=	9.4877290
	Total sample size	=	217
	Actual power	=	0.9507783

The planned sample size of 217 was acquired.

Experiment 2: Choice under Risk (Within Subject)

Results						
	accept	reject				*Row Totals*
won	68 (53.67) [3.83]	30 (44.33) [4.63]				98
lost	58 (53.67) [0.35]	40 (44.33) [0.42]				98
disj	35 (53.67) [6.49]	63 (44.33) [7.86]				98
Column Totals	161	133				294 (Grand Total)

The chi-square statistic is 23.5881. The *p*-value is < 0.00001. The result is significant at *p* < .05.

```
> cohens_w(23.5881,294)
[1] 0.2832519
>
```

χ^2 **tests** - Goodness-of-fit tests: Contingency tables

Analysis: A priori: Compute required sample size

Input:

Effect size w = 0.2832519

α err prob = 0.05

Power (1-β err prob) = 0.95

Df = 4

Output:

Noncentrality parameter λ = 18.6137402

Critical χ^2 = 9.4877290

Total sample size = 232

Actual power = 0.9504755

The planned sample size of 232 was acquired.

References

Gallucci, M. (2018). GAMLj: General Analyses for the Linear Model in Jamovi. Retrieved from https://gamlj.github.io/

JAMOVI Project. (2019). jamovi. Retrieved from www.jamovi.org